NYU STERN

NEW YORK UNIVERSITY · LEONARD N. STERN SCHOOL OF BUSINESS

SALOMON CENTER FOR THE STUDY OF FINANCIAL INSTITUTIONS

The Seventh New York Fed/NYU Stern Conference on Financial Intermediation

A conference jointly sponsored by
the Federal Reserve Bank of New York,
the Salomon Center at NYU Stern,
and the Review of Financial Studies

May 4-5, 2012
NYU Stern School of Business

Program

Friday, May 4

8:00 Registration and Continental Breakfast

9:00 Welcome Remarks
Ingo Walter, Dean of Faculty, NYU Stern

MORNING SESSION
Chair: **Anthony Saunders**, New York University

9:05 "How did the crisis in international funding markets affect bank lending? Balance sheet evidence from the UK"
Shekhar Aiyar, International Monetary Fund
Discussant: **Nicola Cetorelli**, Federal Reserve Bank of New York

10:05 "Government intervention and information aggregation by prices"
Philip Bond, University of Minnesota
Itay Goldstein, University of Pennsylvania
Discussant: **Viral Acharya**, New York University

11:05 Coffee Break

11:30 "Is historical cost accounting a panacea? Market stress, incentive distortions, and gains trading
Andrew Ellul, Indiana University
Chtobhak Jotikasthira, University of North Carolina
Christian Lundblad, University of North Carolina
Yihui Wang, Chinese University of Hong Kong
Discussant: **Mark Carey,** Federal Reserve Board

12:30 Lunch
Introduction: **Matthew Richardson**, Director, Salomon Center
Keynote Speaker: **Susan Schmidt Bies**
Director, Bank of America
Former member of the Board of Governors of the Federal Reserve System

AFTERNOON SESSION
Chair: **Alexander Ljungqvist**, New York University

2:15 "Financial intermediary capital"
Adriano Rampini, Duke University
S. Viswanathan, Duke University
Discussant: **Martin Oehmke**, Columbia University

3:15 Coffee Break

3:30 "Financial integration, housing, and economic volatility"
Elena Loutskina, University of Virginia
Philip Strahan, Boston College
Discussant: **Victoria Ivashina**, Harvard University

4:30 "What fuels the boom drives the bust: Regulation and the mortgage crisis"
Jihad Dagher, International Monetary Fund
Ning Fu, Harvard University
Discussant: **Elena Loutskina**, University of Virginia

5:30 Reception

Saturday, May 5

8:00 Continental Breakfast

MORNING SESSION
Chair: **Hamid Mehran,** Federal Reserve Bank of New York

9:00 "Group lending or individual lending? Evidence from a randomized field experiment in Mongolia"
Orazio Attanasio, University College London
Britat Augsburg, Institute for Fiscal Studies
Ralph de Haas, EBRD
Emla Fitzsimons, Institute for Fiscal Studies
Heike Harmgart, EBRD
Discussant: **Gabriel Natividad,** New York University

10:00 "Understanding the incentives of commissions motivated agents: Theory and evidence from the Indian life insurance market"
Santosh Anagol, University of Pennsylvania
Shawn Cole, Harvard University
Shayak Sarkar, Harvard University
Discussant: **Greg Nini,** University of Pennsylvania

11:00 Coffee Break

11:15 "Interbank network and bank bailouts: Insurance mechanism for non-insured creditors"
Tim Eisert, Goethe University Frankfurt
Christian Eufinger, Goethe University Frankfurt
Discussant: **David Scharfstein**, Harvard University

12:15 Lunch and Adjourn

BANK OF ENGLAND

Working Paper No. 424

How did the crisis in international funding markets affect bank lending? Balance sheet evidence from the UK

Shekhar Aiyar[(1)]

Abstract

Evidence abounds on the propagation of financial stresses originating in the US mortgage market to banking systems worldwide through international funding markets. But the transmission of this external funding shock to the real economy via bank lending is surprisingly under-examined, given the central importance ascribed to this channel of contagion by policymakers. This paper provides evidence of this transmission for the UK-resident banking system, the largest in the world by asset size. It uses a novel dataset, created from detailed balance sheet data reported by resident banks quarterly to the Bank of England. It finds that the shock to foreign funding during the financial crisis caused a substantial pullback in domestic lending. A range of instrumental variables are used to correct for endogeneity and omitted variable bias, and the results are robust to various sensitivity tests. Resident subsidiaries and branches of foreign-owned banks reduced lending by a larger amount than domestically-owned banks, while the latter calibrated the reduction in domestic lending more closely to the size of the funding shock.

Key words: Liquidity shock, financial crisis, transmission mechanism, bank lending, instrumental variables.

***JEL* classification**: G01, G2, E3, E5.

---

(1) International Monetary Fund. Email: saiyar@imf.org

This article represents the views of the author and should not be thought to represent those of the Bank of England or the International Monetary Fund. I am grateful to Andrew Rose, Peter Garber, Greg Thwaites, Tomasz Wieladek, Martin Brooke, Glenn Hoggarth, Lavan Mahadeva, Erik Berglof, Jeromin Zettlemeyer, Franziska Ohnsorge, Mardi Dungey, Michael King and seminar participants at the Bank of England, EBRD, CIGI and Wilfred Laurier University for useful comments. All errors and omissions remain my own.

The Bank of England's working paper series is externally refereed.
Information on the Bank's working paper series can be found at
www.bankofengland.co.uk/publications/workingpapers/index.html

Publications Group, Bank of England, Threadneedle Street, London, EC2R 8AH
Telephone +44 (0)20 7601 4030, Fax +44 (0)20 7601 3298, email mapublications@bankofengland.co.uk.

ISSN 1749-9135 (on-line)

1 Introduction

It is widely accepted that globalized banks were central to the process of transmitting the financial crisis from country to country, and triggering the Great Recession. Contemporary banking systems, especially in advanced economies, are characterised by their web of international linkages, with large claims on and liabilities to non-resident entities. Allen et al (2011) reflect the consensus view when they state that "*Understanding the role of banks in cross-border finance has become an urgent priority....[they] played a leading role in the dynamics of the global crisis of 2007-2009*".

The importance ascribed to globalized banks arises from a standard, two-part narrative that runs roughly as follows. First, stress in the US banking system (and others directly exposed to US mortgages and structured products) spread to banks worldwide through funding markets, both secured and unsecured. Second, this external funding shock to the banking systems of various countries was transmitted domestically through a reduction in credit supply. But while there is a substantial empirical literature documenting the first step above, direct evidence on the second step is relatively slim. This study contributes towards filling that gap.

The literature on the impact of non-monetary shocks on bank lending has a long pedigree. Bernanke (1983) provides evidence that the bank runs and defaults that occurred during the Great Depression caused a reduction in loan supply, and Bernanke and Blinder (1998) model the impact of bank lending on the real economy. A number of papers provide evidence on the real impact of external shocks to bank liquidity. Peek and Rosengren (1997) show that a shock to Japanese banks' liquidity (arising from falling Japanese equity prices) led to a reduction in their lending into the US economy. Khwaja and Mian (2008) document a fall in loans extended by Pakistani banks, in response to an external funding shock (the imposition of capital controls in the wake of the country's 1998 nuclear tests). Schnabl (2011) finds that the liquidity shock to global banks arising from the Russian default in 1998 led to a pullback in lending to Peruvian banks, and that Peruvian banks responded by reducing domestic credit.

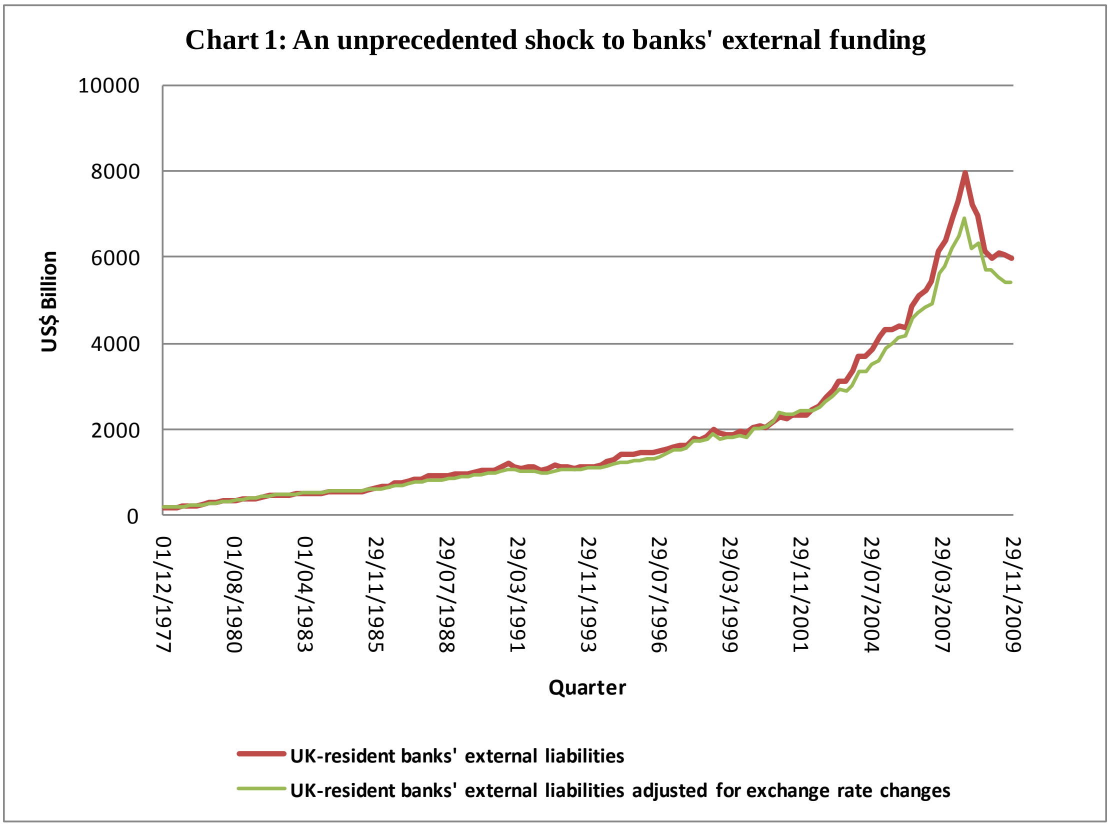
That literature certainly suggests that a cutback in credit supply following a shock to external sources of bank funding is likely to have been an important channel of contagion in the recent crisis. This view is mirrored in the almost universal policy consensus that the Great Recession was a bank-led recession, i.e. that the deterioration in the real economy was initiated by a tightening of international credit conditions rather than vice-versa. In a speech given in April 2010, Jean-Claude Trichet summarized the crisis as follows: *"Given heightened concerns about counterparty risk – which intensified dramatically after the failure of Lehman – cash-rich banks proved unwilling to lend to banks needing liquidity. As a result, the global money market came close to a total freeze. The ensuing decline in banks' ability to raise funds led to a tightening of credit conditions facing enterprises and households."*[1] Similar diagnoses may be found on the lips of other central bankers and policy makers. Successive *World Economic Outlooks* (WEOs) and *Global Financial Stability Reports* (GFSRs) from the IMF have placed the "global credit crunch" at the heart of the recession. In the UK, several recent issues of the Bank of England's *Financial Stability Report* have emphasised the impairment of bank credit arising from the liquidity shock.

This paper investigates how the shock to banks' international funding impacted bank credit supply in a large, non-US, advanced economy during the crisis, thereby providing direct evidence of the transmission channel discussed above. The UK economy provides an ideal testing ground. As a global financial centre, it hosts a large and heterogeneous set of banks, some of which are domestically-owned, but many of which are branches and subsidiaries of banks headquartered in a range of foreign countries. Many of these resident banks have substantial liabilities to non-residents, and are therefore particularly subject to contagion from abroad. And indeed, the shock to external funding that occurred during this crisis was not just large but unprecedented. Chart 1 shows time series from the Bank of International Settlements (BIS) on the aggregate external liabilities of all UK-resident banks, both on an exchange rate-adjusted basis and an unadjusted basis.[2] On an adjusted basis, these

[1] Trichet (2010)

[2] The series showing the stock of exchange rate-adjusted external liabilities is generated by adding quarterly data on exchange rate-adjusted flows to the initial stock.

liabilities fell by 22 percent from their peak in end-March 2008 to end-October 2009 , when they started stabilizing again. By way of comparison, the previous largest 6-quarter fall in external liabilities was only 9 percent, during the ERM crisis in the early 1990s.



From a balance sheet perspective, a bank can react to a shock to external liabilities in any of three ways, or some combination thereof: (i) it can increase its domestic liabilities, that is, borrow more from resident entities, (ii) it can reduce its foreign assets, that is, lend less to non-residents, or (iii) it can reduce its domestic claims, that is, lend less to residents. The focus of investigation here is whether and to what extent banks reacted using option (iii), thereby transmitting the financial contagion to the real domestic economy. A novel dataset is employed, created from the confidential regulatory returns that every UK-resident bank must file quarterly with the Bank of England. These reports contain detailed balance sheet data, providing considerable bank-by-bank heterogeneity in both the external liquidity shock and domestic lending. Exploiting this heterogeneity enables identification of an effect which would usually be difficult to estimate.

The aim is to estimate the impact of the change in banks' external liabilities during the crisis on the change in their domestic lending. OLS is potentially subject to endogeneity and omitted variable bias. Identification is therefore sought by instrumenting the change in banks' external liabilities over the crisis period using three variables. These are: (i) a measure of reliance on wholesale funding, viz. the share of repos in total external liabilities of a bank at the beginning of the crisis; (ii) the share of external liabilities owed to affiliates (as opposed to unaffiliated entities) at the beginning of the crisis; and (iii) a measure of banking system stress in the country in which the bank is headquartered, using the heterogeneity of LIBOR-OIS spreads in different regions of the world. I argue that these instruments are intuitively plausible: all three should be indicative of the size of the funding shock—as attested by a sizeable literature—while not exercising any independent impact on the response variable. Post-estimation tests offer strong support for the validity of the instruments.

The paper is closely related to Cetorelli and Goldberg (2011), who show how the liquidity shock to the banking systems of advanced countries was transmitted to emerging economies via a reduction in bank credit. But of course, contagion was not restricted to transmission from advanced countries to emerging economies. Accordingly, this paper investigates transmission *to* the real economy of a single advanced country arising from a shock to *any* external source of bank liquidity. Because it uses individual bank data, it is able to exploit heterogeneity across banks, rather than relying on cross-country differences.

The study adds to a broader literature on the real impact of the liquidity crisis. Ivashina and Scharfstein (2009) document the fall in syndicated bank lending in the USA during the crisis, providing evidence that this varied according to a bank's access to stable deposit funding and according to exposure to drawdowns on existing lines of credit. Other studies attempt to identify the impact of the funding shock on particular facets of bank lending, such as trade finance (Amiti and Weinstein (2009), Chor and Manova (2009)). A different approach involves the use of survey data: for example Campello, Graham and Harvey (2010) survey CFOs worldwide to ascertain that credit constrained firms planned deeper cuts in employment and investment, drew down on existing credit lines more and sold more assets to fund operations.

The work here takes as an input the shock to banks' international sources of liquidity during the financial crisis, a topic on which there is by now a voluminous empirical literature.[3] Gorton and Metrick (2010) trace the genesis of a "run on repo", i.e. a systemic bank run which occurred not through the traditional channel of depositors withdrawing their funds, but through the withdrawal of repurchase agreements in the vast and global repo market. With minor variations in timing, the pattern was repeated in the inter-bank market for unsecured funding (Acharya and Merrouche (2010)).[4] Short-term funding in US dollars came under particular stress, as documented by McGuire and von Peter (2010) and Coffey, Hrung and Sarkar (2009).

To preview the main results of the paper, I find that a shock to banks' external funding was associated with a substantial contraction in domestic lending. This impact is robust across all deciles of the conditional distribution of the response variable. Foreign subsidiaries and branches reduced lending by a larger amount on average than domestically-owned banks, while the latter calibrated the reduction in domestic lending more closely to the size of the funding shock. There is little evidence that foreign assets buffered domestic lending against shocks to foreign liabilities. I also explore the transmission of the external shock to different sub-components of domestic lending. With the caveat that these sub-samples of the data are smaller and noisier, I find evidence that the shock caused a significant cutback in lending to businesses, to other banks, and to other financial institutions. But I find no evidence for an impact on household lending. This could be because the financial crisis led to the unravelling of the securitisation model of household mortgage lending and caused banks to take mortgages back onto their balance sheets, a development which would tend to increase reported bank lending to households.

The remainder of the paper proceeds as follows. The next section describes the data and estimation strategy. Section 3 provides the main empirical results and section 4 presents some additional results. Section 5 disaggregates domestic lending by sector. Section 6 concludes.

---

[3] Only a small selection of the literature is described here. Other papers include Eichengreen, Mody, Nedeljkovic and Sarno (2009). A rapidly growing theory literature includes Acharya, Gale and Yorulmazer (2009), Brunnermeir and Pederson (2009), Geanakoplos (2009), Dang, Gorton and Holmstrom (2010) and Pagano and Volpin (2009).

[4] Runs also occurred in other funding markets, such as asset backed commercial paper and structured investment vehicles (Covitz, Lang and Suarez (2009), Carey, Correa and Cotter (2009)).

2 Data and estimation strategy

3.1 Data

The UK's resident banking sector comprises the domestically-incorporated units of UK-owned banks, as well as the subsidiaries and branches of banks headquartered in several other countries.[5] It is the world's largest banking sector by asset size. At end-2009, there were over 300 banks resident in the UK, with total assets amounting to £ 7.6 trillion, or over 500% of GDP.[6] While UK-owned banks are on average larger than foreign branches and subsidiaries, the latter are more numerous, so that the assets of foreign-owned and UK-owned banks are about equal (at 50.5 % and 49.5 % of total assets respectively). Of the foreign-owned banks, European banks have the largest presence, accounting for 27.2 % of total assets, followed by US banks (7.9 %) and Japanese banks (2.4%). There is considerable but not overwhelming concentration in assets; thus the top 10 banks account for about 59.8 % of all banking assets.[7]

As part of the UK's regulatory regime, all resident banks must report detailed balance sheet data to the Bank of England on a quarterly basis. Data are reported on a locational (unconsolidated) basis. Thus the liabilities and assets reported by the London subsidiary of, say, a bank headquartered in New York, pertain only to the balance sheet of the subsidiary, not the balance sheet of the banking group.

The main reporting vehicle for balance sheet information is the BT form, which disaggregates banks' liabilities into 11 broad categories (such as sight deposits, time deposits, etc.) and assets into 13 categories (such as cash, bills and commercial paper,

[5] A "foreign subsidiary" is defined for regulatory purposes as a UK-based company in which a foreign bank holds more than 50% of the nominal value of the share capital, or in which a foreign bank, while holding less than 50% of the share capital, nevertheless controls the composition of the board of directors. A "foreign branch" is any permanent establishment (as defined for UK tax purposes) other than a foreign subsidiary, which has and habitually exercises the authority to negotiate and conclude contracts on behalf of its foreign owner. Subsidiaries are subject to regulation—for example on minimum capital requirements—by the Financial Services Authority (FSA), while branches are not. See Aiyar, Calomiris and Wieladek (2011) for further discussion of the UK banking industry and regulatory differences between institutions.

[6] By way of comparison, US-resident banks at end-2009 had assets of US$ 11.67 trillion, or £ 7.19 trillion.

[7] This is in contrast to the much greater concentration in the assets of UK-*owned* banks on a *consolidated* (banking group) basis: the top 10 UK-owned banks account for over 95% of the consolidated assets of UK-owned banks.

market loans, etc.).[8] Each category is split into several sub-categories, some of which contain information on counterparties. The BE form further disaggregates line items from the BT form, focusing particularly on providing more granular counterparty data. The CL and CC forms are used to report on, respectively, banks' external liabilities and assets, i.e. their funding from and their claims on non-residents.

Using data reported on the BT and BE forms, I construct for each UK-resident bank a time-series for claims on households, on businesses, on other banks and on other financial institutions (OFIs). The sum of claims on these four sectors is defined as domestic lending, which is analyzed in conjunction with data on external liabilities from the CL form. Bank mergers are dealt with by creating a synthetic merged series for the entire period.[9] Banks which started or ceased operations during the period studied, or which reported no external liabilities, or which stopped reporting external liabilities during the period studied, are omitted from the sample.[10] These adjustments yield a sample of 141 banks, of which 17 are UK-owned, 32 are foreign subsidiaries, and 92 are foreign branches. These 141 banks accounted for 92.5 % of the assets of all UK-resident banks at the beginning of the sample period.

Table A below shows some summary statistics for the sample. I focus on domestic lending and external liabilities, the two main variables of interest. Since there are considerable differences by bank type—whether a bank is UK-owned, a subsidiary or a branch—reflecting differences in business models, the summary statistics are disaggregated accordingly.[11]The stock of domestic lending and external liabilities is measured at the beginning of what is called the "shock period": the period between end-Q1 2008 and end-Q3 2009 during which external liabilities collapsed so dramatically (see Chart 1 above). Changes in the variables of interest are measured as

---

[8] All regulatory forms used in this study can be viewed at http://www.bankofengland.co.uk/statistics/reporters/defs/defs.htm.

[9] As a robustness check, the main regressions in this paper are repeated using a data sample in which merging banks are *not* combined into a single synthetic series. The results are qualitatively unchanged.

[10] Banks are required to report external liabilities using the CL form only if such liabilities exceed £300 million, so a bank could cease to report external liabilities within the period of study if such liabilities fell below this threshold.

[11] Apart from the differences between locally-owned banks, subsidiaries and branches documented here, another significant feature of the UK banking industry is the high degree of concentration in lending, especially to the household sector. This is examined in Section 6.

changes over the shock period, and adjusted for exchange rate movements using data on currency composition.

Table A: Summary statistics

	Stock 1/			% change		
	Mean	Median	S.D.	Mean	Median	S.D.
		£ millions				
External liabilities						
All banks 2/	23,593	3,245	65,332	-16.1	-15.7	25.9
UK-owned banks	62,436	3,120	131,069	-13.3	-11.4	27.2
Foreign subsidiaries	6,712	1,438	12,753	-20.3	-20.3	27.9
Foreign branches	22,287	5,082	55,740	-15.1	-16.2	25.0
Domestic lending						
All banks	20,434	1,310	69,160	-15.4	-12.6	33.9
UK-owned banks	93,912	6,647	169,303	8.6	10.5	26.0
Foreign subsidiaries	15,515	1,264	41,153	-19.9	-19.6	27.7
Foreign branches	8,568	1,106	24,134	-18.2	-18.3	35.6
		% of total assets				
External liabilities						
All banks	62.7	67.2	24.3			
UK-owned banks	40.8	37.3	29.3			
Foreign subsidiaries	51.4	55.0	25.1			
Foreign branches	70.6	72.7	18.6			
Domestic lending						
All banks	33.6	29.4	23.6			
UK-owned banks	58.1	57.6	26.5			
Foreign subsidiaries	46.6	41.5	20.5			
Foreign branches	24.5	19.8	18.3			

1/ Measured at end-March 2008
2/ The sample comprises 141 UK-resident banks, of which 17 are UK-owned, 32 are foreign subsidiaries, and 92 are foreign branches.

At the beginning of the shock period, UK-resident banks on average had large external liabilities as a share of total liabilities. The ratio was highest for foreign branches, followed by foreign subsidiaries, but even the UK-owned banks sourced more than 40% of their funding from abroad. This pattern was inverted for domestic lending, with UK-owned banks having the largest domestic lending as a share of total assets, followed by subsidiaries and then branches. But even the foreign branches held a substantial fraction of their total assets—about a quarter—in domestic claims.

The shock to external liabilities was very large for all bank types. But it was greatest for foreign subsidiaries, followed by foreign branches and then UK-owned banks. The change in domestic lending was correspondingly large for subsidiaries and branches. UK-owned banks, in contrast, actually expanded their domestic loan book on average over the shock period (but with much variation within the group). These differences in initial conditions, and in the magnitude of the shock, suggest that the response to the shock may also have differed by bank type, an issue which is pursued in section 5.1.

3.2 Estimation

The aim is to examine the impact of a change in banks' external liabilities on its domestic lending over the shock period. Since this is primarily an event study rather than an effort to identify long-run relationships that hold in normal times, the approach employed here relies on cross-sectional heterogeneity in differenced variables.[12] Focusing on cross-sectional heterogeneity over a well-specified shock period has two attractive features in this context. First, it enables the study to abstract from the questions of appropriate lag structure that would arise in a panel context. Second, collapsing the data circumvents the bias introduced by serial correlation in the independent variable (Betrand, Dufflo and Mullainathan (2004)).

The following baseline specification is used.

$$\Delta DL_i = \alpha + \beta_1 \Delta XL_i + \beta_2 DEMAND_i + \gamma' Z_i + \varepsilon_i \qquad (1)$$

where i indexes banks;

ΔDL denotes the change in (log) domestic lending over the shock period;

ΔXL denotes the change in (log) external liabilities over the shock period;

$DEMAND_i$ denotes a bank-specific demand shock; and

Z is a vector of controls (with a corresponding vector of parameters γ).

[12] The empirical design is thus similar to Schnabl's (2011) event study of the liquidity shock caused by the 1998 Russian default, and its impact on bank lending in Peru.

Equation (1) attempts to control for bank-specific demand shocks through the term $DEMAND_i$. This is constructed as follows:

$$DEMAND_i = \sum_{j \in J} s_{ij} \Delta TBL_{ij}$$

where j indexes sector and $j \in J =$ {Households, Businesses, Other Banks, OFIs};

s_{ij} denotes bank i's claims on sector j as a ratio of its total domestic claims; and

ΔTBL_{ij} denotes the change in lending by all banks *except* bank i to sector j.

DEMAND thus seeks identification of the impact of demand by exploiting the heterogeneity of sectoral exposures across banks. For each bank, it uses the sectoral exposure pattern of that bank to weight lending growth by all *other* banks across sectors. All other things equal, banks with large exposure to a sector which experiences a relatively large fall in demand will see domestic lending fall by more than banks with small exposure to that sector: the coefficient on DEMAND should pick up this effect. To some extent this proxy for demand conditions may also pick up supply-side effects (as would any other proxy for demand, such as value-added in each sector, which may determine demand for bank loans but would also reflect loan availability). But it will only pick up *aggregate* supply side effects that affect lending by all banks, not supply side effects which are specific to any particular bank. The bank-specific heterogeneity in the variable arises from differential exposures across sectors.

Given the origins of the financial crisis in the US mortgage market, it seems plausible that ΔXL is exogenous. But this needs to be established rather than assumed, so that estimating equation (1) using OLS is potentially subject to the standard problems of omitted variable bias and endogeneity. One or more non-observables might affect both the response variable (the change in domestic lending) and the explanatory variable of interest (the change in external liabilities). Moreover, given the imperfections of the demand control, a relationship between the response and conditioning variable could occur, say, because weak demand generates a fall in the need for external funding.

These issues are addressed by instrumenting the conditioning variable, the change in external liabilities over the shock period. Three instruments are used.

The first instrument is the share of repos—repurchase agreements—in a bank's total external liabilities, immediately prior to the shock. This is a measure of ex-ante reliance on wholesale external funding. As described in the literature review, there is ample evidence showing that the funding shock was transmitted through the repo market, with the haircut on repos increasing to unprecedented levels in the aftermath of the Lehman collapse. Gorton and Metrick (2010) argue that the run on repo was the chief distinguishing feature of this financial crisis. Raddatsz (2010) presents cross-country evidence that banks with more reliance on wholesale funding came under greater stress—as measured by returns—following Lehman. So it is plausible that this instrument should predict the size of the funding shock in the subsequent period. Both the stock nature of the instrument and its time of measurement would suggest that it should not itself be affected by the subsequent change in banks' domestic lending. Moreover, it seems unlikely that it would impact a future change in domestic lending except through the funding shock.

The second instrument is the ex-ante share of external liabilities owed to foreign affiliates, i.e. "within firm" borrowing as opposed to borrowing from unaffiliated firms. There is substantial evidence that globalised banks with foreign affiliates activate internal capital markets in the face of liquidity shocks. A series of papers demonstrate that this smoothing of liquidity operates in both directions. Thus Cetorelli and Goldberg (2009) show that large US banks absorb liquidity from foreign affiliates in the face of domestic shocks, while de Haas and van Lelyveld (2010) show that in a financial crisis, foreign subsidiaries rely on liquidity support from parents to smooth credit supply.[13] Therefore it is likely that banks with a larger share of exposure to foreign affiliates enjoy relatively greater insulation from external liquidity shocks. As with the repo instrument, the share of liabilities to foreign affiliates is measured immediately prior to the shock period.

The third instrument is a measure of banking system stress during the shock in the region in which a bank is headquartered. LIBOR-OIS spreads (or local equivalents)

[13] Further evidence on internal capital markets is contained in Campello (2002) and Ashcraft (2008).

are used to gauge the level of banking system stress.[14] All countries which own sample banks are grouped into one of nine regions: UK, USA, Eurozone, Switzerland, Australia, Canada, Japan, non-Japan Asia and Other. For each rion, a variable is constructed containing the difference between the average LIBOR-OIS spread during the shock period and the average during the previous 6-quarter period.[15] As Charts 2 and 3 below show, while LIBOR-OIS spreads shot up in all regions during the shock, there was considerable heterogeneity in the extent of this upward movement, with Australian, Canadian and Asian banking systems registering a much smaller mean increase than major Western banking systems.

[14] An overnight index swap (OIS) is an interest rate swap in which the floating leg is tied to an index of overnight rates. The two parties agree to exchange, on a given notional amount, the difference between interest accrued on the fixed and floating legs. The fixed rate is a proxy for market expectations of future overnight rates, with minimal credit risk (because of the short maturity of the claim). Therefore the spread against LIBOR provides a measure of credit risk in the interbank market.

[15] Wherever possible, a regional equivalent is used in place of the LIBOR. Thus the EURIBOR is used for the Eurozone, the TIBOR for Japan, the SIBOR for Singapore, the HIBOR for Hong Kong, the CDOR for Canada and the Bank Bill Swap Reference Rate (BBSW) for Australia, with spreads taken over the corresponding overnight index swap (OIS). For the region non-Japan Asia, an average of the SIBOR-OIS and HIBOR-OIS spread is used, while for the residual region Other, an unweighted average of the spreads for all other regions is used.

Chart 2: LIBOR-OIS spreads

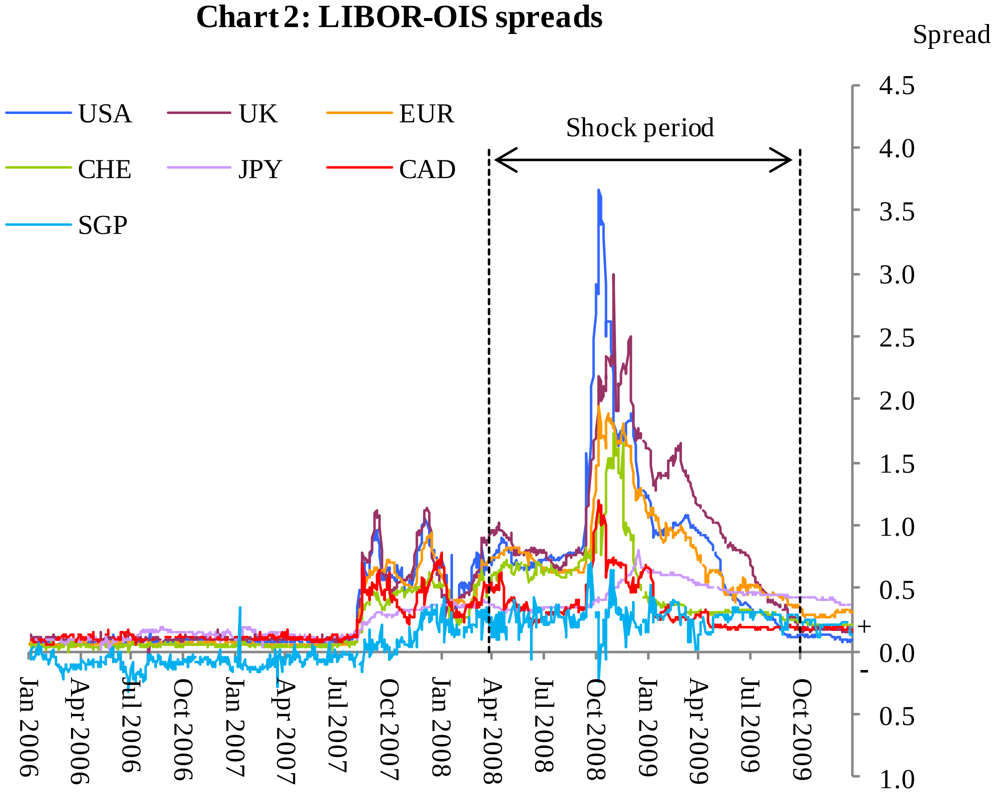
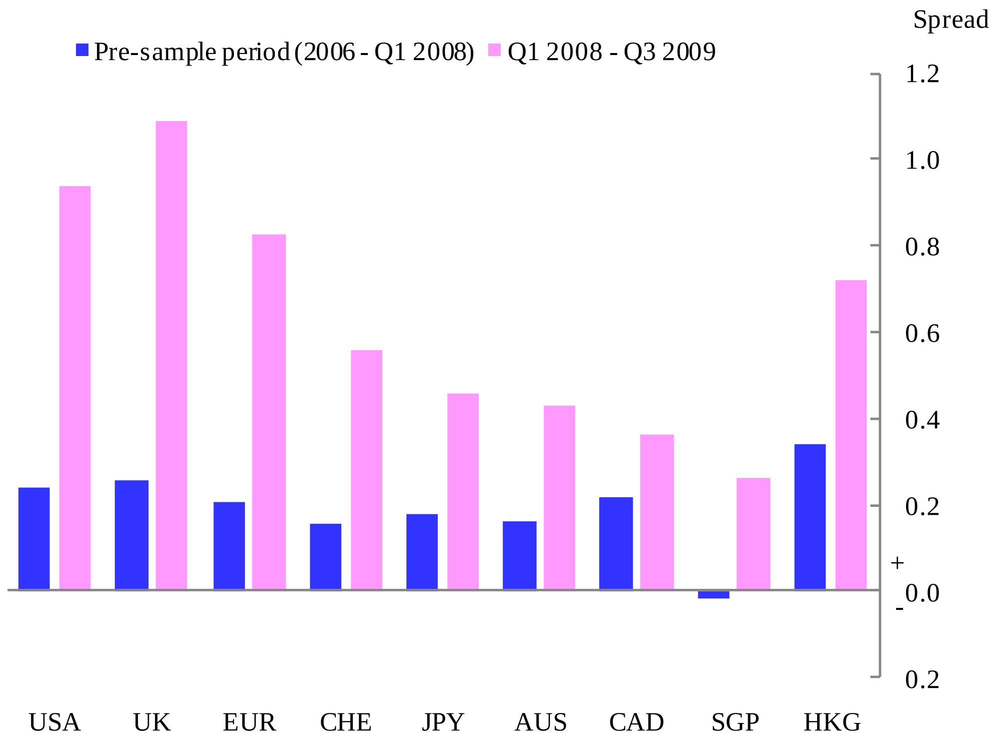


Chart 3: Mean LIBOR-OIS spreads



The LIBOR-OIS spread measures counterparty risk among participating banks.[16] The heterogeneity of counterparty risk among different banking systems during the crisis is well documented; see, for example, Genberg, Hui, Wong and Chung (2009) and Baba and Packer (2009). The divergence by region in the mean increase of the LIBOR-OIS spread provides a measure of this heterogeneity. Other things equal, a greater increase in counterparty risk in a particular banking system should be associated with a greater withdrawal of interbank liquidity.

4 Main Results

Table B presents the results of 2SLS estimation using the instruments described above.

[16] See Taylor and Williams (2008) for evidence that the LIBOR-OIS spread indeed provides a measure of counterparty risk among banks. In particular, they refute the hypothesis that the spread also picks-up liquidity constraints.

Table B: Impact of change in external liabilities on change in domestic lending [(a)]

	1	2	3	4
Dependent variable: ΔDL	**2SLS**	**2SLS**	**2SLS**	**2SLS**
ΔXL	.55**	0.59**	.65**	.60**
	0.27	0.25	0.28	0.28
DEMAND		.035***		.032***
		0.009		0.01
Size controls	No	No	Yes	Yes
N	141	141	141	141
Underidentification (H0: Not identified)				
K-P Wald rank LM-statistic	10.3	11.83	9.3	10.12
p-value	0.012	0.008	0.02	0.01
Overidentifying restrictions (H0: Instruments uncorrelated with error process)				
Sargan chi-squared statistic	0.35	0.17	0.12	0.071
p-value	0.84	0.92	0.94	0.96
Weak instruments (H0: Instruments are weak)				
K-P rank Wald F-statistic	10.23	12.46	9.74	10.25
10% critical value (Stock and Yogo)	9.1	9.1	9.1	9.1

(a) Heteroskedasticity-robust standard errors reported below coefficients. *, ** and *** denote 1%, 5% and 10% levels of significance respectively. These conventions apply to all following tables of regression results. Size controls include total bank assets prior to the shock period, and total external liabilities prior to the shock period.

Column 2 estimates equation (1). A fall in external liabilities of 1 percent leads to a reduction in domestic lending of about 0.6 percent, a substantial impact. Demand shocks, proxied by bank-specific sectoral exposures, exert a significant independent effect on domestic lending, with the expected sign. If the instruments used are valid, including or excluding the demand shock variable should have little impact on the co-efficient of interest. This is confirmed by column 1, where DEMAND is omitted from the specification; the co-efficient on ΔXL remains significant and of a very similar magnitude.

Columns 3 and 4 introduce two controls relating to the size of the bank's balance sheet and external operations. The first is total assets immediately prior to the crisis, and the second is total external liabilities immediately prior to the crisis. Both variables are significant with the expected signs (positive and negative, respectively). They are retained as controls for subsequent regressions tabulated in this paper, but not individually reported, since they make no significant difference to the estimate of the parameters of interest (as can be seen by comparing columns 3 and 4 with columns 1 and 2).

A comprehensive set of post-estimation tests of instrument validity is reported for each regression. The Kleinbergen-Paap rank LM-statistic tests for identification (Kleinbergen and Paap (2006)): the null hypothesis that the instruments are uncorrelated with the endogenous regressor is strongly rejected. Because three instruments are used for a single endogenous regressor, it is possible to conduct Sargan-Hansen tests of overidentifying restrictions. Under all specifications above, the null hypothesis that the exclusion restriction is valid—i.e. that the instruments are uncorrelated with the error term of the structural equation (1)—cannot be rejected. Moreover, p-values indicate that the Sargan statistic lies far to the left of the rejection zone. Finally, the Kleinberg-Paap rank Wald F-statistic indicates that the instruments used are sufficiently strong.[17]

On the basis of strong support from post-estimation tests and the intuitive appeal of the instruments used, I conclude that the impact of the external funding shock on banks' domestic lending is well identified and substantial. This is the paper's central result.

[17] The Cragg-Donald Wald statistic is more conventionally used to test for weakness of instruments, but is invalid under heteroskedasticity-robust standard errors. Critical values are from Stock and Yogo (2005). As a further robustness check, I estimate, but do not report, the regressions in Table B using limited information maximum likelihood (LIML). Again, the validity of the instruments is strongly supported.

Table C: 2SLS and OLS

Dependent variable: ΔDL	**1** **2SLS**	**2** **OLS**
ΔXL	.60**	.51***
	0.28	0.09
DEMAND	.032***	.034***
	0.01	0.01
Size controls	Yes	Yes
N	141	141
R-squared		0.27
Exogeneity of explanatory variable (H0: Variable is exogenous)		
Difference-in-Sargan statistic	0.14	
p-value	0.71	

It is now possible to re-examine whether the external funding shock in equation (1) was indeed exogenous, by comparing an OLS estimate with the 2SLS estimate above.

A comparison of columns 1 and 2 in Table C reveals no significant difference between the OLS estimates and the instrumental variables estimates. *Provided* that the instruments used are valid, this suggests that the funding shock was indeed exogenous. A formal test of the exogeneity of ΔXL is provided by the Difference-in-Sargan statistic. This is constructed as the difference of two Sargan-Hansen statistics, one in which the suspect regressor is treated as endogenous, and one in which the suspect regressor is treated as exogenous. Under the null hypothesis that the regressor is actually exogenous, the statistic is distributed as chi-squared with one degree of freedom.[18]The null cannot be rejected at conventional levels of significance, and the p-value indicates that the statistic lies far to the left of the rejection zone.

Given the exogeneity of ΔXL, OLS is preferred to the 2SLS estimator since it is more efficient. Accordingly, OLS is employed for the remainder of this paper. Before exploring various interactions with the funding shock, I check that the estimated relationship is robust to outliers, and whether the relationship is driven by particular

[18] The test is a heteroskedasticity-robust variant of a Hausman test, to which it is numerically equivalent under homoskedastic errors.

sub-samples of the data. This is an important concern in an economy in which there is much concentration of lending among certain banks, a point that is elaborated in Section 6, where domestic lending is disaggregated on a sectoral basis.

Table D: Median impact on change in domestic lending

	1	2
Dependant variable: ΔDL	**OLS**	**Median Regression**
ΔXL	.51***	.55***
	0.09	0.1
DEMAND	.034***	.031***
	0.01	0.01
Size controls	Yes	Yes
N	141	141
R-squared	0.27	0.21

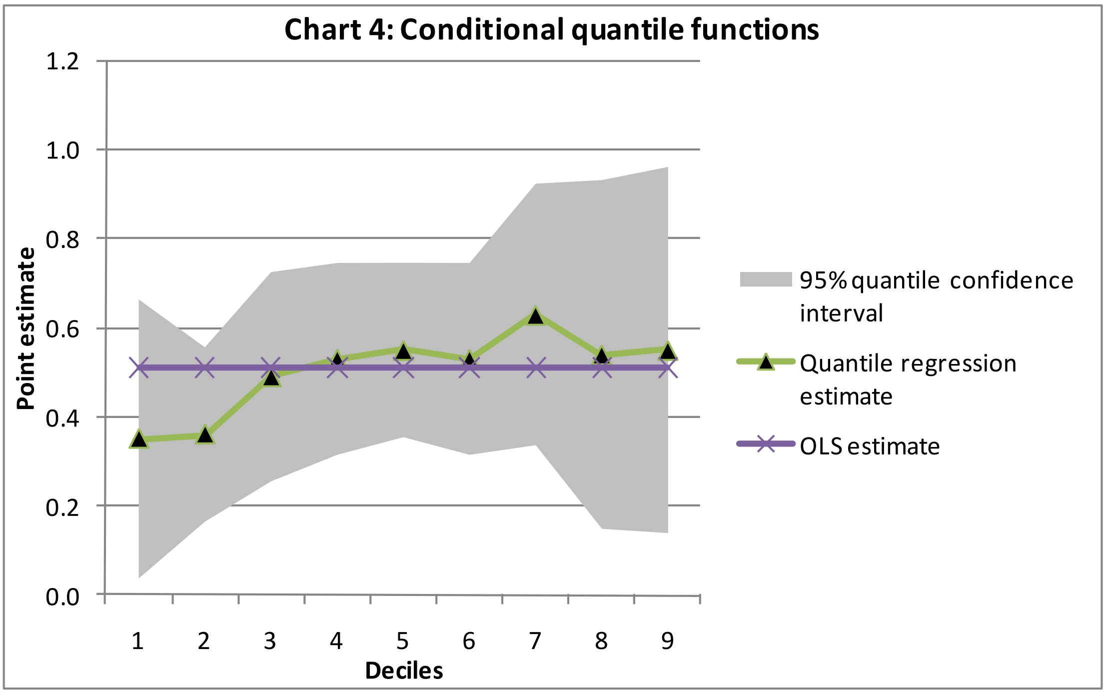


Chart 4: Conditional quantile functions

Table D compares the OLS specification against a median regression. Since the latter attaches less weight to outliers, the close correspondence between the two estimates is reassuring. Chart 4 shows point estimates from a family of quantile regressions. Although there is some variation in these estimates across different deciles of the conditional distribution of the response variable, all estimates are significant. Moreover the 95 percent confidence interval for each decile encompasses the OLS estimate. This assuages concerns about influential observations or sub-samples driving the results.

5 Some further empirical investigations

5.1 What role is played by institutional structure?

The summary statistics presented earlier showed that foreign-owned banks sourced a greater proportion (in the case of foreign branches, a far greater proportion) of their funding from abroad than domestically-owned banks. And the shock to their foreign funding was proportionately larger than for UK-owned banks. These heterogeneous initial conditions suggest that the response to the shock may differ by bank type. In addition, there are numerous theoretical reasons why the credit supply response of domestically-owned banks faced with a financial crisis or economic downturn may differ from the response of a foreign-owned bank (see de Haas and Lelyveld (2006) for a summary). Most importantly, foreign-owned banks may not consider lending in the host country to be a core business activity to the same extent as credit extension in their home country. This may induce them to extend credit on a "transaction-by-transaction basis" in the host country, implying a more volatile pattern of lending relative to a "through-the-cycle" model. Moreover, the differences in institutional structure between subsidiaries and branches—independent capitalization, location of regulator, legal relationship with the parent bank, etc.—might indicate differential responses to a crisis.[19] And, as noted earlier, in the UK branches rely on external funding to a greater extent than subsidiaries and lend less domestically.

Column 1 of Table E includes a dummy signifying UK-ownership (UOB), as well as a term that interacts UK ownership with the change in external liabilities. Being a domestically owned bank had a large and significant positive impact on domestic lending during the crisis. On the other hand, the positive interaction term suggests a sharper pullback in domestic lending in response to a given shock to external liabilities for a domestically-owned bank.

[19] Cerutti et al (2007) provide useful stylized facts about the characteristics of subsidiaries and branches, together with an analysis of organizational choice.

Table E: The impact of bank type

Dependant variable: ΔDL	**1** **OLS**	**2** **OLS**
ΔXL	.45***	.83***
	0.10	0.12
DEMAND	.032***	.033***
	0.01	0.01
UOB	25.98***	
	6.3	
SUB		-26.8***
		6.95
BRN		-26.1***
		6.92
UOB*ΔXL	.38**	
	0.17	
SUB*ΔXL		-.52***
		0.17
BRN*ΔXL		-.32*
		0.19
Constant	3.02	29.6***
	4.56	6.66
Size controls	Yes	Yes
N	141	141
R-squared	0.31	0.32

This result suggests a "head for the exits" impact—a disorderly rush to deleverage—of the financial crisis on foreign-owned banks. That is, foreign-owned banks reduced domestic lending by a large amount irrespective of the size of the actual shock they faced to external liabilities. In contrast, domestically-owned banks calibrated the change in their domestic lending more closely to the size of the external funding shock.

Column 2 replaces the UK-ownership dummy with two dummies signifying whether a bank is a foreign subsidiary (SUB) or a foreign branch (BRN), together with corresponding interaction terms. This corroborates the "head for the exits" phenomenon for both subsidiaries and branches. No evidence is found of substantial differences in response between subsidiaries and branches. It seems that—however differently they may respond to lesser liquidity shocks or economic downturns—their response was very similar in a financial crisis of this magnitude.

5.2 Does FX-denominated domestic lending respond differently?

If foreign liabilities are incurred primarily to support domestic lending in foreign exchange (FX), then we might expect an external funding shock to disproportionately impact FX-denominated domestic lending. Consistent with this hypothesis, column 1 of Table F provides some (weak) evidence of a smaller intercept term for FX-denominated lending. But the effect disappears once the UK-ownership dummy is introduced. Branches and subsidiaries are more likely to lend in foreign exchange, but the differential impact on domestic lending comes from their institutional structure rather than from the currency denomination of their loans.

Table F: Lending in FX

Dependant variable: ΔDL	**1** **OLS**	**2** **OLS**
ΔXL	.54***	.45**
	0.15	0.19
DEMAND	.024**	.024**
	0.011	0.012
Fraction of DL in FX (t=0)	-21.5*	-16.79
	12.77	11.59
(Fraction of DL in FX)*ΔXL	-0.8	-0.01
	0.32	0.31
UOB		23.46***
		7.78
UOB*ΔXL		.41**
		0.22
Constant	12.19*	7.68
	4.84	5.20
Size controls	Yes	Yes
N	141	141
R-squared	0.29	0.32

5.3 Do foreign assets buffer the lending response?

To what extent do foreign assets provide a buffer against a shock to external liabilities? In the extreme case, if foreign liabilities were incurred only to fund foreign assets and if these assets could be easily liquidated in the face of a funding shock, foreign assets could, in principle, completely insulate the domestic economy from the shock. This is clearly not the case: as demonstrated by the regressions presented so far, the funding shock to banks was transmitted to domestic lending. But is the strength of the transmission related to the size of a bank's portfolio of foreign assets?

Table G: Are foreign assets a significant buffer?

	1	**2**	**3**
Dependant variable: ΔDL	**OLS**	**OLS**	**OLS**
ΔXL	.56***	.49***	.39***
	0.15	0.10	0.12
DEMAND	.033***	.033***	.031***
	0.01	0.01	0.01
Foreign assets / Total assets (t=0)	-14.72		
	10.74		
(Foreign assets / Total assets)*ΔXL	-0.11		
	0.32		
Foreign assets / Foreign liabilities (t=0)		-6.56**	-4.54
		3.21	3.2
(Foreign assets / Foreign liabilities)*ΔXL		-0.003	0.04
		0.06	0.06
UOB			24.18***
			6.44
UOB*ΔXL			.41**
			0.17
Constant	13.46**	11.98**	6.72
	6.59	5.35	5.66
Size controls	Yes	Yes	Yes
N	141	141	141
R-squared	0.28	0.29	0.33

Column 1 of Table G introduces the ex-ante ratio of foreign assets to total assets as a regressor, together with an interaction term. No evidence of a buffering role is found by this measure. But this measure is probably less relevant than the one introduced in column 2: the ratio of foreign assets to foreign liabilities (FAFL).

Here, too, the interaction term is insignificant. This may seem surprising, since the buffer effect hypothesized above should drive a significant negative co-efficient on this term. The explanation probably lies in the countervailing impact of what might be called a core business effect. Consider banks whose core business is domestic lending. Other things equal, they will have a small ratio of foreign assets to foreign liabilities. Faced by an external funding shock, these banks will try to cut back first on foreign lending to save core business. This effect would tend to drive a positive interaction term. The fact that the interaction term is found here to be close to zero could indicate that these effects are cancelling each other out.

Column 2 does indicate a lower intercept for banks with large foreign assets relative to foreign liabilities. But this looks very much like the "head for the exits" phenomenon identified for branches and subsidiaries. And indeed, column 3 shows that when a UK-ownership dummy is included in the regression, the co-efficient on FAFL ceases to be significant. Branches and subsidiaries are simply more likely to have a large ratio of foreign assets to foreign liabilities than UK-owned banks.

6 Sectoral components of domestic lending

In this section I decompose domestic lending into its constituent parts—lending to households, lending to businesses, lending to other banks and lending to other financial institutions—and examine separately the impact of the external funding shock on each of these. The evidence presented here is subject to several important caveats. First, the sample of banks which lends to each particular sector is smaller than the full set of banks. Second, and more important, the samples are noisier, because of the concentration of lending in each sector. Finally, I cannot control for demand using the heterogeneity of sectoral exposures across banks as before, since the regressions are now sector-specific.

Table H below illustrates the high degree of concentration in bank lending by sector. The bottom line of the first panel shows the number of banks, in each sector, which lend to that sector. The second panel restricts the sample to those banks with claims on a particular sector of more than £100 million (measured at the beginning of the shock period). The third panel further restricts the sample to those banks with claims of more than £500 million, and the fourth panel to banks with claims of over £ 1 billion. It is evident that while there is concentration in each sector, the degree of concentration is by far the highest in the household sector. Banks with individual claims of more than £ 1 billion account for over 99 percent of total claims on the household sector (compared with a ratio of 96 to 99 percent for the other sectors). Moreover, there are only 15 such banks in the household sector (compared with 45 to 50 banks in the other three sectors).

The first panel therefore contains a large proportion of banks which lend relatively trivial amounts (and are therefore subject to large percentage changes in lending). This introduces a lot of noise into the sample, and the regression results are correspondingly weak. The second, third and fourth panels—in which the sample is restricted by increasing levels of minimum sectoral claims—are more interesting. They show that the shock to external funding had a substantial impact on lending to businesses, to other banks, and to other financial institutions. Moreover, the third and fourth panels seem to indicate that the transmission was strongest for lending to OFIs, followed by lending to other banks, and then by lending to businesses. I find no evidence for an impact on household lending.

Why is there no statistically significant relationship between the shock to external liabilities and the change in household lending? One obvious explanation is that, because of the high degree of concentration, the sample size in the second, third and fourth panels is too small for reliable statistical inference. But there is probably a more fundamental factor at work. To the extent that the securitisation model of household mortgage lending was unwinding during the shock period—with securitized assets held off balance sheet in special purpose vehicles (SPVs) coming back onto banks' balance sheets— this would appear in the data as an increase in lending to the household sector, offsetting the impact of other falls in lending to the sector.[20] Moreover, to the extent that the SPVs are domestic, and financed their purchase of the mortgages through a loan from the originating bank, the unwinding of securitisation would also be manifest in the data as a decrease in lending to OFIs, potentially exaggerating the relationship between the change in external liabilities and the change in domestic lending for the OFI sector.

Another possible explanation for the lack of a statistical impact on household lending could be pressure exerted by the government on banks to keep up lending to households and businesses. This pressure may have been especially acute on banks that were recapitalized by the Treasury or accessed special liquidity facilities (see HM Treasury (2008)). However, this explanation is somewhat less promising, because the government was keen to see lending maintained to both households and businesses, and a statistical impact is found for lending to businesses.

Table H: Sectoral regressions

		Dependent variable: change in lending to sector			
		Households	**Businesses**	**Other Banks**	**OFIs**
		1	**2**	**3**	**4**
	Full sample				
	ΔXL	-66.31	-578	1.13*	.50*
		58.36	584	0.59	0.29
	% of total lending	100	100	100	100
	N	122	134	139	130
	Sectoral lending > £100 m				
	ΔXL	-0.28	.53***	.38*	.69***
		0.39	0.17	0.28	0.24
	% of total lending	99.8	99.8	99.8	99.9
	N	27	91	105	73
Sample	**Sectoral lending > £500 m**				
	ΔXL	0.08	.41***	.50*	.92***
		0.29	0.15	0.29	0.33
	% of total lending	99.6	98.4	99.2	98.8
	N	19	60	70	47
	Sectoral lending > £1000 m				
	ΔXL	0.33	.39**	.79***	1.03***
		0.21	0.18	0.27	0.33
	% of total lending	99.1	96.5	98.6	96.9
	N	15	47	48	40

[20] Unfortunately the balance sheet data used in this study do not include information on mortgage securitisations, and there does not exist, to the best of my knowledge, any alternative data source with bank-specific information on mortgage securitisations on an unconsolidated basis.

In view of the high concentration of bank lending in particular sectors, and the sensitivity of estimates to different sample restrictions, I also examine a family of conditional quantile regressions separately for the business sector, for other banks, and for OFIs. Unlike the quantile regressions for domestic lending as a whole, here there is considerable variation across deciles. Moreover, a number of decile point estimates are insignificant.

Table I: Quantile regressions on components of domestic lending
Dependent variable: change in lending to sector (a)

	Deciles of conditional distribution								
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Businesses									
ΔXL	0.02	.51*	.65**	0.38	.46**	.48***	.53***	.48**	.61*
s.e.	0.51	0.29	0.32	0.26	0.18	0.11	0.13	0.22	0.38
Other Banks									
ΔXL	0.11	0.29	.43**	.45**	.39***	.52**	.50*	0.29	-0.42
s.e.	0.22	0.23	0.18	0.18	0.15	0.21	0.30	0.59	1.1
OFIs									
ΔXL	-0.02	.59**	.64*	.77***	1.03**	1.05**	1.14***	1.19***	1.15
s.e.	0.39	0.25	0.34	0.29	0.41	0.42	0.39	0.31	0.79

(a) All regressions exclude banks with sectoral claims of less than £100 million prior to the shock.

Overall, the evidence seems to indicate a substantial impact of the external funding shock on lending to businesses, other banks and OFIs. But the evidence is weaker than for domestic lending taken as a whole, and point estimates are subject to considerable uncertainty.

7 Conclusion

It is by now widely held that a primary international transmission channel of the financial crisis was through a retrenchment of credit by globalised banks facing a funding shock. But the literature on this bank lending channel is surprisingly sketchy. The UK provides a good testing ground for this channel, because of the size and importance of its resident banking sector. The large number of banks operating in the UK and their heterogeneity provide an ideal sample for statistical inference.

This paper has used detailed regulatory bank returns to identify a substantial impact of the external funding shock on the provision of domestic bank credit. This includes not only direct credit provided to the real economy, but also lending to other banks and OFIs, which would be expected to have further knock-on effects on credit provision to the real economy. Quantile regressions suggest that the impact identified is robust to outliers in the data.

I find evidence of a "head for the exits" phenomenon among foreign-owned banks—both branches and subsidiaries—relative to UK-owned banks. That is, the typical branch or subsidiary cut back on domestic lending to a much larger extent than the typical UK-owned bank, irrespective of the size of the shock to external funding. UK-owned banks, on the other hand, calibrated the credit pullback more closely to the size of the funding shock. This is consistent with UK-owned banks regarding lending within the UK as a core business activity to a greater extent than branches and subsidiaries, and with banks acting to preserve core business. To the extent that we can use these results to think about the experience of other countries, this differential response by bank type is relevant to the transmission of the global funding shock to bank lending in countries with smaller banking sectors, and, in particular, a smaller presence of foreign-owned banks. They suggest that while all advanced countries with globalised banks should have seen some transmission to their real economies through the bank lending channel, the impact would be increasing in the share of foreign-owned banks.

There is some evidence that FX-denominated lending was cut back more than sterling lending, but this is probably because foreign-owned banks are more likely to lend in foreign exchange. There is little evidence that foreign assets acted as a significant buffer to protect domestic lending against the external funding shock. Any buffering role was overwhelmed by the core business effect, by which foreign-owned banks—which tend to have a relatively large foreign assets-to-foreign liabilities ratio—pulled back domestic credit more sharply than UK-owned banks.

The evidence of the impact of the funding shock on lending by sector relies on smaller and noisier samples. Nonetheless, a substantial impact is found on lending to businesses, to other banks, and to OFIs. I find no evidence of an impact on lending to households, perhaps because of the contemporaneous unwinding of the securitisation model of mortgage lending.

Overall, the results lend considerable support to the standard narrative of the global financial crisis and the Great Recession. First, stresses spread from banking systems with direct exposure to US "toxic assets" to secured and unsecured funding markets. This caused a large funding shock to banking systems in various countries, irrespective of direct exposure to US assets, as amply documented in the literature. Second, banks responded to this shock to the liabilities side of their balance sheet by retrenching domestic assets, i.e. reducing lending to resident entities. Thus financial contagion was transmitted to the real economy.

References

Acharya, V, Gale, D and Yorulmazer, T (2009), 'Rollover risk and market freezes', *CEPR Discussion Papers 7122.*

Acharya, V and Merrouche, O (2010), 'Precautionary hoarding of liquidity and inter-bank markets:evidence from the sub-prime crisis', *NYU Working Paper No. FIN-09-018.*

Aiyar, S, Calomiris, C and Wieladek, T (2011), 'Does macro-pru leak? Evidence from a UK natural experiment', *Bank of England Working Paper,* forthcoming.

Allen, F, Beck, T, Carletti, E, Lane, P, Schoenmaker, D and Wagner, W (2011), *Cross-Border Banking in Europe: Implications for Financial Stability and Macroeconomic Policies,* Centre for Economic Policy Research.

Amiti, M and Weinstein, D (2009), 'Exports and Financial Shocks', *NBER Working Paper 15556.*

Ashcraft, A (2008), "Are Bank Holding Companies a Source of Strength to Their Banking Subsidiaries?", *Journal of Money, Credit, and Banking*, vol. 40, no. 2-3.

Baba, N and Packer, F (2009), 'From turmoil to crisis: dislocations in the FX swap market before and after the failure of Lehman Brothers', *BIS Working Paper No. 285.*

Bernanke, B (1983), 'Non-monetary effects of the Financial Crisis in the Propogation of the Great Depression', *American Economic Review,* Vol. 73, No. 3.

Bernanke, B and Blinder, A (1998), 'Credit, Money and Aggregate Demand', *American Economic Review*, Vol. 78, No. 2.

Brunnermeir, M and Pederson, L (2009), 'Market liquidity and funding liquidity', *Review of Financial Studies,* Vol. 22, No. 6.

Campello, M, Graham, J and Harvey, C (2010), 'The Real Effects of Financial Constraints: Evidence from a Financial Crisis', *Journal of Financial Economics*, Forthcoming.

Campello, M (2002), "Internal Capital Markets in Financial Conglomerates: Evidence from Small Bank Responses to Monetary Policy", *Journal of Finance*, vol. 57, no. 6.

Carey, M, Correa, R and Kotter, J (2009), "Revenge of the Steamroller: ABCP as a Window on Risk Choices," Board of Governors of the Federal Reserve System, Working Paper.

Cerutti, E, Dell'Ariccia, G, and Martinez Peria, M (2007), 'How banks go abroad: Branches or subsidiaries?', *Journal of Banking & Finance*, Vol. 31(6).

Cetorelli, N and Goldberg, L (2011), 'Global Banks and International Shock Transmission: Evidence from the Crisis', *IMF Economic Review*, Vol. 59 (1).

Cetorelli, N and Goldberg, L (2008), "Banking Globalization and Monetary Transmission", *NBER Working Paper no. 14101*.

Chor, D and Manova, K (2009), "Off the Cliff and Back? Credit Conditions and International trade During the Global Financial Crisis," *NBER Working Paper 16174.*

Coffey, N, Hrung, W, Nguyen, H, Sarkar, A (2009), 'The global financial crisis and offshore dollar markets', *Current Issues in Economics and Finance*, Vol. 15, No. 6.

Covitz, D, Liang, N and Suarez, G (2009), "The Evolution of a Financial Crisis: Runs in the Asset-Backed Commercial Paper Market," Board of Governors of the Federal Reserve System, Working Paper No. 2009-36.

Dang, T, Gorton, G and Holmström, B (2010a), "Financial Crises and the Optimality of Debt for Liquidity Provision," working paper.

De Haas, R and van Lelyveld, I (2010), 'Internal Capital Markets and Lending by multinational bank subsidiaries', *Journal of Financial Intermediation*, Vol. 19, Issue 1.

De Haas, R and van Lelyveld, I (2006), 'Foreign banks and credit stability in Central and Eastern Europe: a panel data analysis', *Journal of Banking and Finance*, Vol. 30 (7).

Eichengreen, B, Mody, A, Nedeljkovic, M and Sarno, L (2009), 'How the Subprime Crisis Went Global: Evidence from Bank Credit Default Swap Spreads', *NBER Working Paper No. 1490.4*

Geanakoplos, J (2009), 'The leverage cycle', *Cowles Foundation Discussion Paper No. 1715.*

Greenberg, H, Hui, C, Wong, A and Chung, T (2009), 'The link between FX swaps and currency strength during the credit crisis of 2007-2008', *Hong Kong Monetary Authority Research Note 01/2009.*

HM Treasury (2008), 'Statement on banks and building societies', Press Release 126/08 (http://webarchive.nationalarchives.gov.uk/20100407010852/http://www.hm-treasury.gov.uk/press_126_08.htm).

Ivashina, V, and Scharfstein, D (2010), 'Bank lending during the financial crisis', *Journal of Financial Economics* 97, no. 3 (September 2010): 319-338.

Khwaja, A and Mian, A (2008), 'Tracing the effect of bank liquidity shocks:evidence from an emerging market', *American Economic Review*, Vol. 98(4).

Kleibergen, F and Paap, R (2006), 'Generalized Reduced Rank Tests Using the Singular Value Decomposition', *Journal of Econometrics*, Vol. 133.

McGuire, P and von Peter, G (2009), 'The US dollar shortage in global banking', *BIS Quarterly Review*, March 2009.

Pagano, M and Volpin, P (2009), 'Credit ratings failures and policy options', CEPR Discussion Paper 7556.

Peek, J and Rosengren, E (1997), 'The International Transmission of Financial Shocks: The Case of Japan', *American Economic Review*, Vol. 87(4).

Raddatz, C (2010), "When the Rivers Run Dry: Liquidity and the Use of Wholesale Funds in the Transmission of the US Subprime Crisis", Policy Research Working Paper 5203, The World Bank.

Schnabl, P (2011), 'The International Transmission of Bank Liquidity Shocks: Evidence from an Emerging Market', *Journal of Finance*, forthcoming.

Stock, J and Yogo, M (2005), 'Testing for Weak Instruments in Linear IV Regression', in D.W.K. Andrews and J.H. Stock, eds. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge: Cambridge University Press.

Taylor, J and Williams, J (2008), 'A black swan in the money market', NBER Working Paper No. 13943.

Trichet, J-C (2010), 'What can central banks do in a financial crisis?', Susan Bies Lecture, Kellog Distinguished Lecture Series, Evanston, Illinois.

Government intervention and information aggregation by prices[1]

Philip Bond[2] Itay Goldstein[3]

June 8, 2011

[1]We thank Michael Fishman, Qi Liu, Adriano Rampini, Jean-Charles Rochet, Duane Seppi, and seminar audiences at the Bank of Israel, Dartmouth College, the Federal Reserve Banks of Chicago and Minneapolis, the International Monetary Fund, New York University, the University of Maryland, the University of Waterloo, the 2009 Financial Crisis Workshop at Wharton, the 2010 American Economic Association meetings, the 2010 FIRS Conference, the 2010 NBER Summer Institute on Capital Markets and the Economy, the 2010 Chicago-Minnesota Accounting Theory Conference, the third Theory Workshop on Corporate Finance and Financial Markets at Boston University, and the 2011 American Finance Association meetings for helpful comments. Bond thanks the Cynthia and Bennett Golub Endowed Faculty Scholar Award Fund for financial support. All errors are our own.

[2]University of Minnesota.

[3]University of Pennsylvania.

Abstract

Market prices are thought to contain a lot of useful information. Hence, government regulators (and other economic agents) are often urged to use market prices to guide decisions. An important issue to consider is the endogeneity of market prices and how they are affected by the prospect of government intervention. We show that if the government learns from the price when taking a corrective action, it might reduce the incentives of speculators to trade on their information, and hence reduce price informativeness. We show that transparency may reduce trading incentives and price informativeness further. Diametrically opposite implications hold for the alternative case in which the government's action amplifies the effect of underlying fundamentals. We derive implications for the optimal use of market information and for the government's incentives to produce its own information.

1 Introduction

Market prices of financial securities contain a great deal of information. As such, they can provide valuable guidance for government decisions. Consistent with this, existing research establishes that government actions do indeed reflect market prices.[1] Moreover, numerous policy proposals call for governments to make even more use of market prices, particularly in the realm of bank supervision.[2] Such policy proposals are increasingly prominent in the wake of the recent economic crisis and the perceived failure of financial regulation prior to it.[3]

An important issue to be considered when discussing the use of market prices in government policy is that prices are endogenous and their information content might be affected by government policy. In the recent crisis, government actions were not only perceived to be reactions to market prices, but expectations about them were often a major driver of changes in asset prices. For example, market activity in the weeks leading up to the eventual announcement of government support for Fannie Mae and Freddie Mac, for Citigroup, and for General Motors was largely driven by speculation about the government's behavior. Hence, government actions affect prices, and consequently may also affect the ability of the government to learn from prices. This may affect the desirability of market-based intervention.

To study these effects, we consider the process by which information gets aggregated into the price. Our paper analyzes the effect of market-based government policy on the trading incentives of speculators and hence on the ability of the financial market to aggregate speculators' dispersed information. We derive positive implications for the behavior of prices and government actions when the government learns from prices, and also implications for

[1] See Feldman and Schmidt (2003), Krainer and Lopez (2004), Piazzesi (2005), and Furlong and Williams (2006).

[2] See, e.g., Evanoff and Wall (2004) and Herring (2004).

[3] For example, Hart and Zingales (2009) propose a mechanism, by which the government will perform a stress test on banks whose market price deteriorates below a certain level, in order to evaluate whether there is a need for intervention. Other recent proposals say that banks should issue contingent capital (i.e., debt that converts to equity) with market-based conversion triggers (see Flannery (2009), McDonald (2010)).

how the government should best make use of market information. We distinguish between *corrective* government policy – i.e., one that aims to help firms in trouble, for example, by bailing out struggling banks – and *amplifying* government policy – for example, shutting down bad banks – and show that they generate very different implications.

While we focus here on government intervention, some of our results apply more generally to other contexts in which individuals incorporate information from market prices into their decisions. Other possible applications include the actions of boards of directors; of providers of capital; and of managers themselves.

The canonical model of information aggregation was developed by Grossman and Stiglitz (1980), Hellwig (1980), and Admati (1985). Speculators possess information about the fundamentals of an asset and trade on it in a market that is subject to noise/liquidity shocks. In the existing literature, the cash flows produced by the asset are exogenous. However, if the government (or some other decision-maker) responds to prices, the cash flows are instead endogenous. We extend the canonical model of information aggregation to account for this. In addition to the market price, the government observes a private signal of its own. The government's intervention can be either corrective or amplifying, depending on its objective. The informativeness of the price in this model is determined by the trading incentives of speculators, i.e., the aggressiveness with which they trade on their information.

A key determinant of speculators' trading behavior is the uncertainty to which they are exposed. Being risk averse, they trade less when the risk is higher. In the face of such uncertainty, speculators benefit when the government takes a corrective action based on information not contained in the price, but correlated with the fundamental. Consequently, speculators can trade more aggressively on their information, and the equilibrium price is more informative. However, if the government increases its reliance on market prices as a source of information, this benefit is lost, and speculators trade less aggressively resulting in a less informative price. Hence, the government's use of market prices in its decision on a corrective action reduces the informational content of prices.

This result has a couple of implications. First, even though ex post the government wants to apply Bayes rule to extract information from market prices, from an ex ante perspective the government could do better: we show that, for a moderate corrective action, a government would always want to commit to refrain, to some extent, from fully using market prices ex post. This increases the informativeness of the price and enables the government to make better decisions. Such commitment could be achieved, for example, by having an overconfident policymaker who thinks his information is more precise than it really is. Second, our model implies that the government's own information has more value than its direct effect on the efficiency of the government's decision. When the government has more precise information, it relies less on the market price, and this makes the market price more informative. Hence there are complementarities between the government's own information and the market's information, and so it is not advisable for the government to rely completely on market information.

Our paper also delivers implications about transparency. Governments are often criticized for not conveying their information or policy goals. The question is whether such disclosure is desirable when the government tries to learn from the market. In the case of corrective actions, we show that the type of transparency that is considered matters a lot. Disclosing the government's information about the fundamentals reduces trading incentives and price informativeness, while disclosing the government's policy goal increases them. The key distinction is that in the first case the government reveals information about the fundamental which is the object that speculators are informed about, and so this decreases their informational advantage and their trading incentive, while in the second case the dominant effect is the decrease in uncertainty that pushes speculators to trade more aggressively. This distinction is new to the literature on transparency.[4]

[4]There are recent papers showing that transparency might be welfare reducing, e.g., Morris and Shin (2002) and Angeletos and Pavan (2007). In these papers, the source for the result is the existence of coordination motives across economic agents. In contrast, such coordination motives do not exist in our model, where, conditional on the price (which is observed to all), speculators do not care about what other speculators do. Importantly, the above-mentioned papers do not explore the implications of transparency about different types of information, as we do here.

Importantly, we show that the implications change drastically once we consider the case of amplifying actions. The key effect here is opposite since the government's action based on its information amplifies the uncertainty that speculators are exposed to and decreases their trading incentives. Hence, when the government relies more on market prices it leads to an increase in trading incentives and price informativeness. In contrast to the case of corrective actions, this implies that there is a rationale for the government to commit to market-based rules rather than act on the basis of its own information.[5] Overall, inferring information from the price is harder in the case of an amplifying action than in the case of a moderate corrective action.

We conclude the analysis by considering the case where the government learns from multiple securities, and show that this is not an easy solution to the problem of inferring information from prices, and in some cases adding more securities actually reduces overall informativeness. Hence, the problems we identify in this paper are not a result of market incompleteness.

Our paper adds to a growing literature on the informational feedback from asset prices to real decisions; see, for example, Fishman and Hagerty (1992), Leland (1992), Khanna, Slezak, and Bradley (1994), Boot and Thakor (1997), Dow and Gorton (1997), Subrahmanyam and Titman (1999), Fulghieri and Lukin (2001), Foucault and Gehrig (2008), and Bond and Eraslan (2010). In particular, it complements papers such as Bernanke and Woodford (1997), Goldstein and Guembel (2008), Bond, Goldstein and Prescott (2010), Dow, Goldstein and Guembel (2010), and Lehar, Seppi and Strobl (2010), which analyze distinct mechanisms via which the use of price information in real decisions might reduce the informational content of the price.

Relative to these papers, our focus is on the efficiency of aggregation of dispersed information by market prices. This topic, which has long been central in economics and finance (e.g., Hellwig (1980)), has not been analyzed in any of the related papers. For example, in

[5] Other papers emphasize the commitment aspect associated with market-based rules. See Faure-Grimaud (2002), Rochet (2004), and Lehar, Seppi, and Strobl (2010).

Bond, Goldstein and Prescott (2010), the price of any traded asset after a realization of some underlying state variable θ is assumed to equal the expected payoff of the asset conditional on θ. In other words, even if information about the state variable θ is dispersed across many investors, the price is assumed to fully and efficiently aggregate this dispersed information. In this paper we are particularly interested in what is going on inside the black box, i.e., in how information gets aggregated given the expected government intervention. Moreover, the model we develop enables very tractable analysis of information aggregation by prices in the presence of informational feedback to real decisions.

The remainder of the paper is organized as follows. Section 2 describes the model. The analysis and solution of the model are contained in Section 3. In Section 4, we provide our main results about the effect of the government's use of the price on price informativeness and the implications for the best use of market information. In Section 5, we analyze the implications of our model for whether the government benefits from transparency. Section 6 provides an extension of the model to consider multiple securities. Section 7 concludes. All proofs are relegated to an appendix.

2 The model

The canonical model of information aggregation was developed by Grossman and Stiglitz (1980), Hellwig (1980), and Admati (1985). Speculators possess information about the fundamentals of an asset and trade on it in a market that is subject to noise/liquidity shocks. In the existing literature, the cash flows produced by the asset are exogenous. However, if the government (or some other decision-maker) responds to prices, the cash flows are instead endogenous. We extend the canonical model of information aggregation to account for this. In addition to the market price, the government observes a private signal of its own. The government's intervention can be either corrective or amplifying, depending on its objective. The informativeness of the price in this model is determined by the trading incentives of speculators, i.e., the aggressiveness with which they trade on their information.

We focus on one firm (a financial institution, for example), whose stock is traded in the financial market. In $t = 0$, speculators obtain signals about the cash flow that will be generated from the firm's operations, and trade on them. In $t = 1$, the government, who learns information about the expected cash flow from the price of the stock, makes a decision about its intervention. In $t = 2$, cash flows are realized and speculators get paid.

2.1 Cash flows and government intervention

Absent government intervention, the firm generates a cash flow of θ. We refer to θ as the fundamental of the firm. It is distributed normally with mean $\overline{\theta}$ and standard deviation σ_θ. We denote the precision of prior information by $\tau_\theta \equiv \frac{1}{\sigma_\theta^2}$.

The government has the ability to affect firm cash flows. For example, the government may directly transfer cash to or from a firm; may provide liquidity support in the form of loans at below-market rates; or may directly intervene in the firm's management. Regardless of the type of intervention, we denote by T the change in the firm's cash flows.

In deciding on T, the government has to weigh the benefit against the cost. We assume that the government's benefit is weakly concave in the change in cash flow T. That is, the incremental benefit from supporting firms diminishes as the support gets larger. In addition, and crucially, the government's benefit also depends on the firm's fundamentals θ. For example, the government may have a preference to help firms with poor fundamentals if doing so reduces socially inefficient liquidation of assets; or it may prefer to help firms with strong fundamentals if firms with better fundamentals contribute more to social welfare.

For tractability, we assume that the government's benefit function takes the form

$$V(T, \theta) \equiv a_2 T^2 + a_1 T + a_T T\theta + v(\theta), \tag{1}$$

where a_2, a_1 and a_T are constants, and v is a function. As noted, $a_2 \leq 0$, while the sign of a_T reflects whether the government prefers to help firms with strong ($a_T > 0$) or weak ($a_T < 0$) fundamentals.

The cost to the government of changing cash flows by T is given by $\gamma(T)$, which is a weakly convex function of T. Again, for tractability we assume that γ is quadratic.

Assuming that V and/or γ are non-linear in T, and equating marginal benefit to marginal cost, we can write the change in cash flow T that a fully-informed government would like to implement as $\lambda\left(\hat{\theta} - \theta\right)$, where λ and $\hat{\theta}$ are constants.[6] In particular, note that λ is positive if the government cares more about helping weak firms ($a_T < 0$) and negative if the government cares about helping strong firms ($a_T > 0$). We refer to the two cases as *corrective* and *amplifying* actions, respectively. In the first case, the government transfers cash to firms with fundamentals below a threshold at the expense of firms with fundamentals above a threshold. In the second case, it does the opposite. In the context of intervention in the banking sector, corrective actions often come in the form of bailing out weak banks (while potentially taxing strong banks), whereas amplifying actions can come in the form of shutting down weak banks (while potentially easing constraints on strong banks).

When the government is not fully-informed, it must base its intervention on its beliefs about the fundamental θ. In this case, it intervenes according to the following rule:

$$T \equiv \lambda\left(\hat{\theta} - E\left[\theta|I_G\right]\right), \tag{2}$$

where $E\left[\theta|I_G\right]$ is the expected cash flow of the firm given the information available to the government I_G. We will elaborate below on the sources of government information. Note that the key benefit of the simple functional forms we have adopted for V and γ is that the policy rule is linear in θ. This helps us maintain the linear solution that is heavily used in the literature on information aggregation, and thus is important for the tractability of the model.

[6] Explicitly, equating marginal cost to marginal benefit gives $2a_2T + a_1 + a_T\theta = \gamma''(0)T + \gamma'(0)$, and hence $T = \frac{a_1 - \gamma'(0) + a_T\theta}{\gamma''(0) - 2a_2}$. So, $\lambda = \frac{-a_T}{\gamma''(0) - 2a_2}$ and $\hat{\theta} = \frac{a_1 - \gamma'(0)}{-a_T}$.

2.2 Information and trading

There is a continuum $[0,1]$ of speculators in the financial market with constant absolute risk aversion (CARA) utility, $u(c) = -e^{-\alpha c}$, where c denotes consumption and α is the absolute risk aversion coefficient. Each speculator i receives a noisy signal about the fundamental:

$$s_i = \theta + \varepsilon_i, \tag{3}$$

where the noise term ε_i is independently and identically distributed across speculators. It is drawn from a normal distribution with mean 0 and standard deviation σ_ε. We use $\tau_\varepsilon \equiv \frac{1}{\sigma_\varepsilon^2}$ to denote the precision of speculators' signals.

Each speculator chooses a quantity x_i to trade to maximize his expected utility given his private signal s_i and the price P that is set in the market for the firm's stock:

$$x_i(s_i, P) = \arg\max_{\tilde{x}} E\left[-e^{-\alpha\tilde{x}(\theta+T-P)}|s_i, P\right]. \tag{4}$$

Here, trading a quantity x_i, the speculator will have an overall wealth of $x_i \cdot (\theta + T - P)$, where $\theta + T$ is the cash flow from the security after intervention, and P is the price paid for it. The speculator's information consists of his private signal s_i and the market price P.

In addition to the informed trading by speculators, there is a noisy supply shock, $-Z$, which is distributed normally with mean 0 and standard deviation σ_z. We again use the notation $\tau_z \equiv \frac{1}{\sigma_z^2}$. In equilibrium, the market clears and so:

$$\int x_i(s_i, P)\, di = -Z. \tag{5}$$

The government's information I_G consists of two components. First, the government observes the price P, which provides a noisy signal of the fundamental θ. Second, the government observes a private signal s_G of the fundamental:

$$s_G = \theta + \varepsilon_G, \tag{6}$$

where the noise term ε_G is drawn from a normal distribution with mean 0 and standard deviation σ_G. We use $\tau_G \equiv \frac{1}{\sigma_G^2}$ to denote the precision of the government's signal. The government then sets T based on the rule in (2) using its two pieces of information P and s_G.

3 Analysis

An equilibrium consists of a mapping from signal realizations and the supply shock Z to price P, and individual demands $x_i(s_i, P)$, such that individual speculators' demands maximize utility given s_i and P (according to (4)) and such that the market clearing condition (5) holds. In addition, here the government's choice of T maximizes its objective, given its signal s_G and the price P, as in (2).

As is standard in almost all the literature, we focus on linear equilibria in which the price P is a linear function of the average signal realization—which equals the fundamental θ—and the supply shock $-Z$.[7]

Proposition 1 below formally establishes the existence of a linear equilibrium. Before stating the proposition, we now provide a less formal derivation focusing on the informativeness of the equilibrium price.

In a linear equilibrium, the price can be written as

$$P = p_0 + p_Z(\rho\theta + Z), \tag{7}$$

for some parameters p_0, p_Z and ρ. In particular, ρ measures the informativeness of the

[7] In a linear equilibrium, an individual speculator's demand is linear in his own signal, as we show below. Consequently, total speculator demand is a linear function of the average signal realization, which equals θ.

price, since the informational content of the price is the same as the linear transformation $\frac{1}{\rho p_Z}(P - p_0) = \theta + \rho^{-1}Z$. This transformation is an unbiased estimate of the fundamental with precision $\rho^2\tau_Z$, where as one would expect, precision increases in price informativeness ρ. Intuitively, the price of the security is affected by both changes in the fundamental θ and changes in the noise variable Z. The informativeness of the price about the fundamental can be summarized by the ratio between the effect of the fundamental on the price and the effect of noise on the price.

It is worth highlighting that our measure of informativeness relates to the fundamental θ, and not the cash flow $\theta + T$ (as would be the case in measures of price efficiency). This is because the government is attempting to learn the fundamental θ from the price, and so the informativeness about θ is the relevant object for the government's ability to take an appropriate action attempting to maximize its objective.

Given normality of the fundamental θ and the supply shock $-Z$, the price P is itself normal. Consequently, given normality of the error term ε_G, the government's posterior of the fundamental θ is normal. Moreover, the government's estimate of the fundamental is linear in its own signal, $s_G = \theta + \varepsilon_G$, and in the price P. The government's estimate of the fundamental is consequently

$$E[\theta|s_G, P] = K_\theta\bar{\theta} + K_P\frac{1}{\rho p_Z}(P - p_0) + w(\rho)s_G, \tag{8}$$

where K_θ, K_P and $w(\rho)$ are weights that sum to one. In particular, $w(\rho)$ is the weight the government puts on its own signal in estimating the fundamental, which depends on the information available in the price.[8] By the standard application of Bayes' rule to normal distributions it is given by:

$$w(\rho) \equiv \frac{\tau_G}{\tau_\theta + \rho^2\tau_z + \tau_G}. \tag{9}$$

The weight that the government puts on its own signal is the precision of this signal (τ_G)

[8] Of course, the constants K_θ and K_P also depend on the price informativeness ρ, but for expositional ease we do not make this dependence explicit.

divided by the sum of precisions of the government's signal, the prior information (τ_θ) and the signal from the price ($\rho^2\tau_z$). As one would expect, the government puts more weight on its own signal when it is precise (τ_G is high) and less when the price is informative (ρ is high). Given the policy rule (2), the intervention is

$$T\left(s_G, P\right) = \lambda\hat{\theta} - \lambda w\left(\rho\right)\left(\theta + \varepsilon_G\right) - \lambda K_P \frac{1}{\rho p_Z}\left(P - p_0\right) - \lambda K_\theta \bar{\theta}. \tag{10}$$

Similar to the government, each speculator assigns a normal posterior (conditional on his own signal s_i and price P) to the fundamental θ. Moreover, from (10), each speculator also assigns a normal posterior to the size of the intervention T. Consequently, the well known expression for a CARA individual's demand for a normally distributed stock applies,

$$x_i\left(s_i, P\right) = \frac{E\left[\theta + T|s_i, P\right] - P}{\alpha var\left[\theta + T|s_i, P\right]}. \tag{11}$$

Thus, the amount traded is the difference between the expected value of the security (fundamental + intervention) and the price, divided by the variance of the expected value multiplied by the risk aversion coefficient. Intuitively, speculators want to trade more when they expect a higher gap between the value of the security and the price, but, due to risk aversion, this tendency is reduced by the variance in expected security value.

To characterize the equilibrium informativeness of the stock price, consider simultaneous small shocks of δ to the fundamental θ and $-\delta\rho$ to Z. By construction (see (7)), this shock leaves the price P unchanged. Moreover, the market clearing condition (5) must hold for all realizations of θ and Z. Consequently,

$$\delta\frac{\partial}{\partial\theta}\int x_i\left(s_i, P\right) di = \delta\rho.$$

Substituting in (10) and (11) yields equilibrium price informativeness:

$$\rho = \frac{1}{\alpha}\frac{\frac{\partial}{\partial s_i}E\left[\theta+T|s_i,P\right]}{var\left[\theta+T|s_i,P\right]} = \frac{1}{\alpha}\frac{\left(1-\lambda w\left(\rho\right)\right)\frac{\partial}{\partial s_i}E\left[\theta|s_i,P\right]}{\left(1-\lambda w\left(\rho\right)\right)^2 var\left[\theta|s_i,P\right]+\left(\lambda w\left(\rho\right)\right)^2\tau_G^{-1}}. \tag{12}$$

Here, the informativeness of the price is essentially determined by how much speculators trade on their information about θ. As explained above, this is determined by two factors: the relation between the information and the value of the asset, which appears in the numerator, and the variance in the value of the asset, which appears in the denominator. Regarding the first one, we see in the numerator that a $1 change in the expected fundamental changes expected value by $(1-\lambda w)$, due to the government's intervention based on its signal. The variance of the expected value, which appears in the denominator, is a function of two components: the expected variance due to the fundamental θ and the variance of the noise in government information. The relative importance of these two components is determined by λ, the strength of the government's action, and by w, the extent to which the government relies on its own signal.

Proposition 1 *For $\lambda \leq 1$, a linear equilibrium exists. Equilibrium price informativeness ρ satisfies (12). For any λ sufficiently close to 0, there is a unique linear equilibrium.*

(All proofs are in the appendix.) Note that the original Grossman-Stiglitz model featured a unique linear equilibrium. We can see this in our model by assuming that there is no government intervention, i.e., by setting $\lambda = 0$. In this case, equation (12) has a unique solution given by

$$\rho = \frac{1}{\alpha}\frac{\frac{\partial}{\partial s_i}E\left[\theta|s_i,P\right]}{var\left[\theta|s_i,P\right]} = \frac{1}{\alpha}\frac{\frac{\tau_\varepsilon}{\tau_\theta+\rho^2\tau_z+\tau_\varepsilon}}{\frac{1}{\tau_\theta+\rho^2\tau_z+\tau_\varepsilon}} = \frac{\tau_\varepsilon}{\alpha}. \tag{13}$$

Moreover, as can be easily verified from the proof of Proposition 1, even with government intervention, our model would feature a unique equilibrium if the weight w that the government puts on its own information was exogenous and unaffected by the price informativeness ρ. However, due to the effect of the informativeness of the price on the weight that the government puts on its information in the intervention decision, our model sometimes exhibits multiple equilibria. This is because, as we see in (12), the informativeness ρ affects the weight w, which in turn affects ρ, so we have to solve a fixed-point problem, which sometimes has multiple solutions. Economically, as the price informativeness increases, traders are exposed to less residual risk, which induces them to trade more aggressively resulting in a more informative price. Indeed, for a large enough corrective action ($\lambda >> 0$), we can construct examples where our model has multiple equilibria. Our paper is not the first to show that the uniqueness of equilibrium in Grossman and Stiglitz (1980) is not robust to extensions of the model. For example, Ganguli and Yang (2008) show that introducing private information about the aggregate liquidity shock may lead to multiplicity of equilibria.

Below, we focus on the case where λ is small in absolute value, and so multiplicity does not arise. As we discuss below, the results that we highlight depend on λ being sufficiently small in absolute value. Numerical calculations (see details in Appendix B) suggest that these results hold for a wide range of values of λ. For example, for the case of corrective actions, they hold at least up to a level of $\lambda = 30\%$, and often much higher. This range seems to us to be both economically meaningful and realistic. That is, in the real world, government interventions implied by $\lambda = 30\%$ correspond to very substantial transfers, and so those corresponding to significantly higher values of λ strike us as much less realistic. For this reason we focus on these results.

4 Government policy and price informativeness

In this section we study how the government's decision to use prices as a basis for intervention affects the informativeness of the equilibrium price and what implications this has for the best way to use of market prices. For comparison, consider the benchmark case in which the government completely ignores the price. In this case, the government's estimate of the fundamental is (analogous to (8)),

$$E[\theta|s_G] = \tilde{K}_\theta\bar{\theta} + w_{-P}s_G,$$

where $\tilde{K}_\theta$ is a constant and w_{-P} is the weight the government puts on its own signal when it ignores the price,

$$w_{-P} \equiv \frac{\tau_G}{\tau_\theta + \tau_G}.$$

The government's intervention is then (analogous to (10)),

$$T_{-P}(s_G) = \lambda\hat{\theta} - \lambda w_{-P} \cdot (\theta + \varepsilon_G) - \lambda\tilde{K}_\theta\bar{\theta}.$$

Equilibrium price informativeness when the government ignores the price is then given by (12), with the weight that the government puts on it own signal, $w(\rho)$, replaced by $w_{-P} > w(\rho)$.

Below, we will analyze how the reliance on market price (which shifts the weight on the government's own signal from w_{-P} to $w(\rho)$) affects the informativeness ρ. To understand the results that follow, it is helpful to keep in mind the following three key properties of the standard model *without* government intervention, i.e., where $\lambda = 0$.

Property 1: In the standard model, price informativeness is greater when cash flows depend *less* on the fundamental. To see this, suppose that the traded asset pays $\nu\theta$ instead of θ, where ν is some constant. From (13), the price informativeness is $\frac{\tau_\varepsilon}{\nu\alpha}$. Hence, when the importance of the fundamental is lowered, i.e., $\nu < 1$, price informativeness is increased. Economically, reducing the importance of the fundamental has two opposite effects. It reduces the usefulness of a trader's signal in forecasting cash flows (the numerator in (13)), which causes traders to trade *less* aggressively and pushes price informativeness *down*. It also reduces the risk to which traders are exposed (the denominator in (13)), which causes traders to trade *more* aggressively and pushes price informativeness *up*. As is clear from (13), the second effect is the dominant one, so the net effect is an increase in price informativeness.

Property 2: Any change to the cash flow that is a deterministic function of price has no effect on price informativeness. To see this, consider again (13), and simply replace θ with $\theta + h(P)$, where h is an arbitrary function. It is clear that neither the numerator nor the denominator is affected. Economically, since the price is common knowledge, traders' trading decisions, and hence price informativeness, are determined only by the moments of cash flow after conditioning on the price.

Property 3: In the standard model, price informativeness is unrelated to the tightness of traders' priors (τ_θ) about the fundamental θ. Inspecting (13), we see that changes in τ_θ affect both the usefulness of a trader's signal in forecasting cash flows (the numerator), and the the risk to which traders are exposed (the denominator), but the two effects exactly offset one another.

4.1 The case of corrective actions ($\lambda > 0$)

Returning to the case of government intervention, we now explore the effect of the government's usage of the information in the price when taking a corrective action.

4.1.1 Price informativeness

Property 1 described above implies that corrective actions tend to increase price informativeness: the corrective nature of the intervention reduces the importance of the fundamental in determining cash flows. However, this is true only for corrective actions that are based to some degree on the government's own signal. If instead the government based its decision *only* on the price, Property 2 above implies that price informativeness is the same as without government intervention. Comparing the two cases, the more weight the government puts on the price the less informative prices become. There is a counter effect, however, as the reliance of the government on its own signal introduces another source of variance that speculators are exposed to: the noise in the government's signal. This reduces speculators' incentive to trade and hence price informativeness. Overall, the following proposition shows that the counter effect is weaker when the corrective action is mild (λ is small and positive).

Proposition 2 *For mild corrective actions (λ small and positive) price informativeness is reduced when the government uses the price as a basis of policy.*

As one can see from (12), if instead λ is large and positive, the dominant factor determining a speculator's residual uncertainty about $\theta + T$ is the government's error term ε_G. In this case, if the government puts more weight on its own signal s_G by putting less weight on the price, it only increases a speculator's residual uncertainty, and consequently, it reduces equilibrium price informativeness. As we noted above, however, numerical simulations (see details in Appendix B) show that this will happen only when λ is well above 30%, which we find unrealistic for most cases.

4.1.2 Excess volatility

A direct implication of Proposition 2 is that in the case of a mild corrective action, the government's use of market information increases the excess volatility in stock prices. Excess volatility is usually defined as the fraction of volatility of prices that is not attributable to changes in the fundamental θ. In our framework, given that $P = p_0 + p_Z(\rho\theta + Z)$, excess volatility is given by:

$$\left(\frac{p_Z^2 \tau_Z^{-1}}{\rho^2 p_Z^2 \tau_\theta^{-1} + p_Z^2 \tau_Z^{-1}}\right)^{1/2} = \left(\frac{\tau_\theta}{\rho^2 \tau_Z + \tau_\theta}\right)^{1/2}. \tag{14}$$

It is clear from the above expression that excess volatility is negatively related to price informativeness ρ. This is because when the price provides less precise information about the fundamental, it is affected more by shifts in noise trading, and this leads to excess volatility. Hence, when the government uses the information in the price for its decision on a mild corrective action, it increases excess volatility.

4.1.3 How should the government use market information?

Proposition 2 characterizes the effect of the use of the information in the price on price informativeness. We now examine the implications of this result for how the government should best use market information. It is worth stressing up-front that we stop short of a full analysis of social welfare. Instead, we take as given the government's objective (see page 6), and ask how the government can best meet this objective.

Proposition 2 suggests that the government faces a trade-off. Ex post, using the price allows it to make a better decision. However, doing so decreases the informativeness of the price. If the government can ex ante commit to a policy rule, its best policy balances these two effects.

Formally, the ex post best intervention for the government is given by (10). However, if the government can commit, this is just one of an infinite number of policy rules the government might follow. In particular, consider the class of linear policy rules defined by weights $\tilde{w}$, $\tilde{K}_P$ and $\tilde{K}_\theta$,

$$\tilde{T}\left(s_G, P; \tilde{w}, \tilde{K}_P, \tilde{K}_\theta\right) \equiv \lambda\hat{\theta} - \lambda\tilde{w}s_G - \lambda\tilde{K}_P\frac{1}{\rho p_Z}\left(P - p_0\right) - \lambda\tilde{K}_\theta\bar{\theta}.$$

The government aims to maximize, by choice of weights $\tilde{w}$, $\tilde{K}_P$ and $\tilde{K}_\theta$, its objective:

$$E_{s_G,P}\left[E\left[V\left(\tilde{T}\left(s_G, P; \tilde{w}, \tilde{K}_P, \tilde{K}_\theta\right), \theta\right) - \gamma\left(\tilde{T}\left(s_G, P; \tilde{w}, \tilde{K}_P, \tilde{K}_\theta\right)\right)|s_G, P\right]\right].$$

By construction, for a given price informativeness ρ, the weights $w\left(\rho\right)$, K_P and K_θ maximize

$$E\left[V\left(\tilde{T}\left(s_G, P; \tilde{w}, \tilde{K}_P, \tilde{K}_\theta\right), \theta\right) - \gamma\left(\tilde{T}\left(s_G, P; \tilde{w}, \tilde{K}_P, \tilde{K}_\theta\right)\right)|s_G, P\right]$$

for any realization of s_G and P. Hence, by the envelope theorem, a small increase in $\tilde{w}$ away from the ex post Bayes-rule weight $w\left(\rho\right)$ has an effect only via changes in equilibrium price informativeness ρ. For mild corrective actions, this effect is positive (this is just a local version of Proposition 2), and so a government's commitment to overweight its own signal increases the accuracy of its intervention, and hence increases the government's welfare. The reason is that ex post overweighting of the government's signal generates a first-order improvement of price informativeness, but has only a second-order cost in terms of how effectively the government makes use of available information.[9] Formally:

Proposition 3 *Consider a mild corrective action (λ small and positive), and let ρ be the equilibrium price informativeness if the government uses information in the ex post best way. Then there exists $\tilde{w} > w(\rho)$ such that the government would do better ex ante by committing to place weight $\tilde{w}$ on its own signal.*

While Proposition 3 implies that the government can gain by committing to overweight its own signal and underweight the price ex post, it is clear that it should never go to the extreme of completely ignoring the stock price. This is because the only reason to reduce the weight on the price is to increase price informativeness, but this is of no use if the government does not learn from the information in the price at all. In other words, the government does not care about price informativeness per se; it cares about it only to the extent that it allows it to make better decisions, and this implies using the information in the price to some extent.

An important question regarding the result in Proposition 3 is how such commitment can be implemented. Given that no one sees the government's signal but the government itself, how can the government credibly commit to put more weight on its signal than is ex-post best for it? One way to achieve such commitment is to choose a policymaker who is overconfident about the precision of his own signal. Such a policymaker will put more

[9] A related result is developed by Goldstein, Ozdenoren, and Yuan (2010). In their model, the central bank learns from speculators on the desirability of maintaining a fixed exchange rate regime. This sometimes leads speculators to coordinate on trading on correlated information, reducing the efficiency of the central bank's decision. By putting less weight on market outcomes, the central bank can then reduce the tendency for coordination and increase efficiency. In contrast, here, there is no issue of coordination and correlated information. By committing to place lower weight on market information, the government reduces the exposure of speculators to risk and encourages them to trade more aggressively on their information, making the price more informative.

weight on his signal—and less on both the price and his prior—than implied by Bayes' rule simply because his bias leads him to think that his signal should receive a larger weight. Having such a bias is then beneficial ex ante by making prices more informative.

Finally, Proposition 3 also implies that the government can potentially gain by ex post overweighting *both* its own signal and the price, at the expense of underweighting its prior $\bar{\theta}$. Note, however, that either government overconfidence about the precision of its own information, or underconfidence about the precision of the price, lead to simultaneously overweighting own information s_G and underweighting the price.

4.1.4 The importance of the government's own information

It is tempting to interpret policy proposals to use market information as implying that governments do not need to engage in costly collection of information on their own. For example, in the context of banking supervision, one might imagine that the government could substantially reduce the number of bank regulators. Our framework enables analysis of this issue when the usefulness of market information is endogenous and affected by the government's use of this information. We find that in the case of a mild corrective action, the government's own information exhibits complementarity with the market's information, as the informativeness of the price increases when the government has more precise information and relies less on the price. Hence, the usual argument that market information can easily replace the government's own information is incorrect.

Formally, suppose that the precision of the government's information, τ_G, is a choice variable. What would be the benefits of increasing τ_G? Given that the price aggregates speculators' information imperfectly, the government is using both the price and its private information s_G when making its intervention decision. Then, an increase in the precision of its private signal has a direct positive effect on the quality of the government's overall information about the fundamental θ. More interesting, however, is that an increase in τ_G also has a positive indirect effect, in that more accurate government information leads to more informative prices. The logic follows the previous results on the effect of the government's use of market information on the quality of this information: An increase in τ_G increases the weight w that the government puts on its own information, which, in the case of mild corrective action, increases the equilibrium price informativeness. Hence, the government should be willing to spend more on producing its own information than the direct contribution of this information to its decision making would imply.

The result is summarized in the following proposition.[10]

Proposition 4 *For mild corrective actions (λ small and positive), an increase in the precision of the government's information (τ_G) increases the informativeness of the price.*

4.2 The case of amplifying actions ($\lambda < 0$)

So far, we have considered the case of corrective actions. To recap, in the case of a moderate corrective action, the government reduces price informativeness when it bases interventions on the market price, as opposed to relying solely on its own information. Consequently, to maximize its objective, the government would like to commit to (at least slightly) overweight its own information. Related, the accuracy of its own information (τ_G) is a complement to the use of market prices, since it leads naturally to the government placing more weight on its own information, which increases price informativeness.

The key force driving these results is Property 1 described above: when the fundamental has a weaker effect on the cash flow from the security, as in the case of corrective actions, price informativeness is increased, because traders are exposed to less risk and trade more aggressively. (Recall that, due to Property 2, this occurs only as long as the corrective action is based to some degree on the government's private information.) If instead the

[10] Bond, Goldstein, and Prescott (2010) also note that the government's own information helps the government make use of market information. However, in that model, the market price perfectly reveals the expected value of the firm, and the problem is that the expected value does not provide clear guidance as to the optimal intervention decision. Hence, the government's information can complement the market information in enabling the government to figure out the optimal intervention decision. Here, on the other hand, the fact that the government is more informed encourages speculators to trade more aggressively, and thus leads the price to reflect the expected value more precisely.

action is amplifying, Property 1 generates an effect in the opposite direction—traders are now exposed to more risk and trade less aggressively—and the above results are reversed. Summarizing:

Proposition 5 *Consider an amplifying action $\lambda < 0$:*

(A) Price informativeness is increased when the government uses the price as a basis of policy.

(B) There exists $\tilde{w} < w(\rho)$ such that the government would do better by ex ante committing to place weight $\tilde{w}$ on its own signal.

(C) For $|\lambda|$ sufficiently small, an increase in the precision of the government's information (τ_G) reduces the informativeness of the price.

Note that in the case of amplifying actions, the distinction between moderate and non-moderate actions matters less than in the case of corrective actions. The main result in part (A) holds independently of the size of the amplifying action. This is because the decrease in exposure to government noise when the government relies more on market price strengthens the increased incentive to trade and the increase in price informativeness.

An interesting insight stemming from of part (B) of Proposition 5 is that there is a force that pushes the government towards the adoption of clear (market-based) rules, rather than acting in a discretionary way based on its own information. The implication is that clear rules are desirable when the government's action is amplifying, e.g., when the government shuts down bad banks, but not when it is corrective, e.g., when the government provides support to struggling banks.

Finally, it is interesting to consider how price informativeness varies with the intervention parameter λ (recall that λ is derived from the objective function of the government). Based on Property 1, amplifying actions ($\lambda < 0$) lead to lower price informativeness than the benchmark case of no-intervention, whereas mild corrective actions ($\lambda > 0$ but not too large) lead to greater price informativeness than the benchmark (this can also be seen from (12)).

Consequently, there is a sense in which corrective actions are easier for a government to implement effectively. The result is summarized in the following proposition.

Proposition 6 *Price informativeness is greater in the case of mild corrective actions than for amplifying actions.*

5 Transparency

Governments are often criticized for not being transparent enough about their information and policy goals. But is government transparency actually desirable when the government itself is trying to elicit information from the price? Does the release of information by the government increase or decrease speculators' incentives to trade on their information? We analyze these questions for the case where the government is taking a mild corrective action based on its own information and the information in the price. We find that the results are very different depending on the type of transparency in question, i.e., transparency about the government's information versus about its policy goals.

5.1 Transparency about the government's information

Proposition 7 summarizes the effect that the government's disclosure of its signal s_G has on the informativeness of the price and consequently on the government's objective.

Proposition 7 *For mild corrective actions, the disclosure of the government's signal s_G reduces equilibrium price informativeness and hence the value of the government's objective function.*

This result is rather surprising as it implies that the government's disclosure of its own information is detrimental. Essentially, the fact that the government reveals its information reduces the incentive of speculators to trade on their information, resulting in a lower level of price informativeness. Thus, the government is better off not revealing its information.

To understand this result, recall from (12) that price informativeness is given by

$$\rho = \frac{1}{\alpha} \frac{\frac{\partial}{\partial s_i} E\left[\theta + T | s_i, P, s_G\right]}{var\left[\theta + T | s_i, P, s_G\right]},$$

where we have added s_G to the speculators' information set to account for the government's disclosure of information. Now, given that speculators know the government's signal, conditional on the price P, they know what the government's intervention T will be, and so, given no uncertainty about T,

$$\rho = \frac{1}{\alpha} \frac{\frac{\partial}{\partial s_i} E\left[\theta | s_i, P, s_G\right]}{var\left[\theta | s_i, P, s_G\right]} = \frac{\tau_\varepsilon}{\alpha}.$$

This is lower than the informativeness without transparency, which for mild corrective actions is approximately $\frac{1}{\alpha} \frac{1-\lambda w}{1-2\lambda w} \tau_\varepsilon$ (see (12)).

Economically, transparency reduces speculators' residual uncertainty about the fundamental, but also reduces the extent to which each speculator's private signal affects his forecast of this fundamental. These forces have opposite effects on price informativeness and cancel out with each other. This is essentially Property 3 described above. The result is then driven by a combination of Property 1 and Property 2. As in Proposition 2, for moderate corrective actions, speculators like the reduction in uncertainty induced by the government taking an action that is correlated with their private information (and is not reflected in the price). This effect is lost when the government reveals its signal, as then the government's signal is already reflected in the price, and, conditional on the price, is not correlated anymore with speculators' signals.

Finally, note that the net effect is opposite in the case where the government takes an amplifying action. In this case, revealing the government's signal increases price informativeness and improves the value of the government's objective function.

5.2 Transparency about the government's policy goal

Now, suppose that speculators do not know the government's policy goal. In particular, they do not know exactly the fundamental threshold $\hat{\theta}$, below which the government would like to inject resources into the firm. Suppose that speculators believe that $\hat{\theta}$ is drawn from some normal distribution. Obviously, the government knows $\hat{\theta}$. Proposition 8 summarizes the effect that the government's disclosure of its policy goal $\hat{\theta}$ has on the informativeness of the price and consequently on the value of its objective function.

Proposition 8 *For mild actions (λ sufficiently close to zero),*[11] *the disclosure of the government's policy goal $\hat{\theta}$ increases equilibrium price informativeness and hence the value of the government's objective function.*

This result captures what is perhaps the usual intuition about transparency and the reason why it is strongly advocated. The idea is that when the government reveals its policy goal, it reduces uncertainty for speculators. This encourages them to trade more aggressively, resulting in higher price informativeness. The government is then better off as it can make more informed decisions.

For illustration, note that, just like before, the equilibrium price informativeness is given by the ratio:

$$\frac{1}{\alpha}\frac{\frac{\partial}{\partial s_i}E\left[\theta+T|I\right]}{var\left[\theta+T|I\right]},$$

where I denotes the information available to speculators. The intervention T continues to be given by (10). The only difference from before is that now $\hat{\theta}$ may be unknown (depending on whether the government discloses it or not).

Whether or not the government discloses its policy threshold, the numerator $\frac{\partial}{\partial s_i}E\left[\theta+T|I\right]$ in the price informativeness expression is unchanged from before. This is because the signal

[11] The condition that λ is sufficiently close to zero is needed only to guarentee equilibrium uniqueness (see Proposition 1). However, even when there are multiple equilibria, both the minimum and maximum equilibrium levels of informativeness are higher under transparency about $\hat{\theta}$.

s_i does not tell a speculator anything about the government's policy threshold. In contrast, the denominator $var\left[\theta + T|I\right]$ in case speculators do not know $\hat{\theta}$ is

$$(1-\lambda w)^2 \, var\left[\theta|s_i, P\right] + (\lambda w)^2 \, \tau_G^{-1} + \lambda^2 var\left(\hat{\theta}\right).$$

As a result, the level of informativeness is higher when the government discloses the policy goal, as then speculators are exposed to less risk and are willing to trade more aggressively. Note that this result does not depend on whether the government takes a corrective action or an amplifying action.

Economically, it matters whether the government discloses information about something that the speculators have some information about or not. In the first case, when the government discloses information about the fundamental, this has an ambiguous effect on speculators' incentive to trade, as the information both reduces the value of their signal and the risk they are exposed to. In the second case, when the government discloses information on its policy goal, the effect on trading incentives is unambiguous, since this only reduces the risk that speculators are exposed to.

6 Adding Another Security

Our analysis in previous sections assumed that the traded security is a claim on the value of the firm $\theta + T$. Under the case of a corrective action, we found that the government reduces the informativeness of the price when it uses the information in the price in its intervention decision. An interesting question is whether the government can do better when there are more securities traded in the market, so that the market is closer to completeness. In particular, suppose that in addition to the traditional security, there is a security that provides a claim on the fundamental cash flow of the firm θ. In this section, we analyze the equilibrium outcomes under the assumption that both a security on θ and a security on $\theta + T$ are traded.

The only difference between the version of the model studied in this section and the one in previous sections is that we now assume that speculators can trade two securities; the first one is a claim on $\theta + T$ and the other one is a claim on θ. In each market, there is a noisy supply shock: $-Z_{\theta+T}$ in the market for the $\theta + T$ security and $-Z_\theta$ in the market for the θ security. Both $Z_{\theta+T}$ and Z_θ are distributed normally with mean 0 and standard deviation σ_z (as before, $\tau_z \equiv \frac{1}{\sigma_z^2}$). We denote the prices in the two markets $P_{\theta+T}$ and P_θ, respectively.

To make our analysis as transparent as possible, we assume that the noise shocks $Z_{\theta+T}$ and Z_θ are independent of each other. While this assumption can be relaxed, it has the benefit of making the informational content of the price vector particularly easy to describe. Concretely, the lack of complete correlation between $Z_{\theta+T}$ and Z_θ reflects an assumption that some noise trades are truly random, rather than stemming entirely from hedging motives.

Using the expressions for a CARA individual's demand for a normally distributed asset in a framework with multiple assets (see Admati (1985)), a speculator i's demands for the two securities, $x_{i,\theta+T}$ and $x_{i,\theta}$, are as follows:

$$x_{i,\theta+T}\left(s_i, P_{\theta+T}, P_\theta\right) = \frac{var\left(\theta\right)\left(E\left[\theta+T\right] - P_{\theta+T}\right) - cov\left(\theta, \theta+T\right)\cdot\left(E\left[\theta\right] - P_\theta\right)}{\alpha\left(var\left(\theta+T\right)var\left(\theta\right) - cov\left(\theta, \theta+T\right)^2\right)}, \quad (15)$$

$$x_{i,\theta}\left(s_i, P_{\theta+T}, P_\theta\right) = \frac{var\left(\theta+T\right)\left(E\left[\theta\right] - P_\theta\right) - cov\left(\theta, \theta+T\right)\cdot\left(E\left[\theta+T\right] - P_{\theta+T}\right)}{\alpha\left(var\left(\theta+T\right)var\left(\theta\right) - cov\left(\theta, \theta+T\right)^2\right)}, \quad (16)$$

where all the expectation, variance, and covariance terms are conditional on the information available to speculator i: s_i, $P_{\theta+T}$, and P_θ.

These expressions reveal the complex nature of demands for assets in a framework with multiple correlated assets. Consider the numerator in each of the two expressions. The first term in the numerator reflects the speculative motive for trading: An increase in the expected payoff of the asset relative to its price leads the speculator to increase the quantity of the asset that he demands. The second term in the numerator reflects the hedging motive for trading: If the two assets are positively correlated, an increase in the expected payoff of the other asset relative to its price leads the speculator to decrease the quantity of the asset that he demands, as he uses the asset to hedge against his exposure in the other asset. As we will see, these conflicting motives for trade can severely reduce the informativeness of price of a given asset, and so the overall effect of adding a security on the informativeness of the price system might end up being negative.

To analyze the informativeness of the price system, we again focus on linear equilibria of the form:[12]

$$\begin{aligned} P_{\theta+T} &= \bar{p}_1 + p_{1\theta}\theta + p_{11}Z_{\theta+T} + p_{12}Z_\theta, \\ P_\theta &= \bar{p}_2 + p_{2\theta}\theta + p_{21}Z_{\theta+T} + p_{22}Z_\theta. \end{aligned} \tag{17}$$

A little manipulation implies that the informational content of observing $P_{\theta+T}$ and P_θ is the same as observing the linear transformations:

$$\begin{aligned} \tilde{P}_{\theta+T} &\equiv \frac{\frac{p_{22}}{p_{2\theta}}\frac{P_{\theta+T}-\bar{p}_1}{p_{1\theta}} - \frac{p_{12}}{p_{1\theta}}\frac{P_\theta-\bar{p}_2}{p_{2\theta}}}{\frac{p_{22}}{p_{2\theta}} - \frac{p_{12}}{p_{1\theta}}} = \theta + \rho_{\theta+T}^{-1}Z_{\theta+T}, \\ \tilde{P}_\theta &\equiv \frac{\frac{p_{11}}{p_{1\theta}}\frac{P_\theta-\bar{p}_2}{p_{2\theta}} - \frac{p_{21}}{p_{2\theta}}\frac{P_{\theta+T}-\bar{p}_1}{p_{1\theta}}}{\frac{p_{11}}{p_{1\theta}} - \frac{p_{21}}{p_{2\theta}}} = \theta + \rho_\theta^{-1}Z_\theta, \end{aligned} \tag{18}$$

where

$$\begin{aligned} \rho_{\theta+T} &\equiv \frac{\frac{p_{22}}{p_{2\theta}} - \frac{p_{12}}{p_{1\theta}}}{\frac{p_{22}}{p_{2\theta}}\frac{p_{11}}{p_{1\theta}} - \frac{p_{12}}{p_{1\theta}}\frac{p_{21}}{p_{2\theta}}}, \\ \rho_\theta &\equiv \frac{\frac{p_{11}}{p_{1\theta}} - \frac{p_{21}}{p_{2\theta}}}{\frac{p_{11}}{p_{1\theta}}\frac{p_{22}}{p_{2\theta}} - \frac{p_{21}}{p_{2\theta}}\frac{p_{12}}{p_{1\theta}}}. \end{aligned} \tag{19}$$

Similarly to the parameter ρ in the main model, $\rho_{\theta+T}$ and ρ_θ, together, capture here the informativeness of the price system.

[12] A formal proof of the existence of a linear equilibrium is available from the authors upon request.

Extending the logic in our main model, consider simultaneous small shocks of δ to the fundamental θ, $-\rho_{\theta+T}\delta$ to $Z_{\theta+T}$, and $-\rho_\theta\delta$ to Z_θ. By construction, these shocks leave the prices unchanged. Moreover, the market clearing conditions in both markets must hold for all realizations of θ, $Z_{\theta+T}$, and Z_θ. As a result:

$$\begin{aligned} \rho_{\theta+T} &= \frac{var\left(\theta\right)\frac{\partial}{\partial s_i}E\left[\theta+T\right]-cov\left(\theta,\theta+T\right)\cdot\frac{\partial}{\partial s_i}E\left[\theta\right]}{\alpha\left(var\left(\theta+T\right)var\left(\theta\right)-cov\left(\theta,\theta+T\right)^2\right)}, \\ \rho_\theta &= \frac{var\left(\theta+T\right)\frac{\partial}{\partial s_i}E\left[\theta\right]-cov\left(\theta,\theta+T\right)\cdot\frac{\partial}{\partial s_i}E\left[\theta+T\right]}{\alpha\left(var\left(\theta+T\right)var\left(\theta\right)-cov\left(\theta,\theta+T\right)^2\right)}, \end{aligned} \tag{20}$$

where again all the expectation, variance, and covariance terms are conditional on the information available to a speculator i: s_i, $P_{\theta+T}$, and P_θ.

Now, as in our main model,

$$T\left(s_G,\tilde{P}_{\theta+T},\tilde{P}_\theta\right)=-\lambda w\left(\theta+\varepsilon_G\right)+B(\tilde{P}_{\theta+T},\tilde{P}_\theta), \tag{21}$$

where $B(\tilde{P}_{\theta+T},\tilde{P}_\theta)$ is linear in the two price signals. Hence, we get explicit expressions for the following conditional moments:

$$\begin{aligned} var\left(\theta+T\right) &= \left(1-\lambda w\right)^2 var\left(\theta\right)+\left(\lambda w\right)^2 var\left(\varepsilon_G\right), \\ cov\left(\theta,\theta+T\right) &= \left(1-\lambda w\right)var\left(\theta\right), \\ \frac{\partial}{\partial s_i}E\left[\theta+T\right] &= \left(1-\lambda w\right)\frac{\partial}{\partial s_i}E\left[\theta\right]. \end{aligned}$$

Plugging these expressions in (20), and after some algebra, we get:

$$\begin{aligned} \rho_{\theta+T} &= 0, \\ \rho_\theta &= \frac{\frac{\partial}{\partial s_i}E\left[\theta\right]}{\alpha var\left(\theta\right)}=\frac{\tau_\varepsilon}{\alpha}. \end{aligned} \tag{22}$$

So, the overall informativeness of the price system is $\frac{\tau_\varepsilon}{\alpha}$. This is the same level of informativeness as in a model where the only traded security is a claim on θ. It is lower (higher) than the level of informativeness in a model where the only traded security is a claim on $\theta + T$ and the government takes a moderate corrective (amplifying) action.

Intuitively, traders have information about θ, but not about the noise in the government's signal, ε_G. Consequently, the trade size in the $\theta + T$ security is determined entirely by the trade size in the θ security and the price difference between the two securities; but it is independent of a trader's information s_i. Given this, the price of the $\theta + T$ security reveals no information beyond the price of the θ security. Hence, the informativeness of the price system is identical to what it would be if the only traded security was a claim on θ. Since under a corrective action, the informativeness is higher with only a $\theta + T$ security than with only a θ security (because of the effect discussed earlier, that the government's corrective action based on its own information reduces volatility and encourages trading), adding a θ security on top of a $\theta + T$ security harms informativeness overall and makes the government worse off.

Finally, we have also analyzed a model where the two traded securities are a claim on $\theta + T$ and a claim on T. In such a model, both securities have a level of informativeness of $\frac{\tau_\varepsilon}{\alpha}$. Hence, the comparison with a model with only a $\theta + T$ security under a corrective action yields ambiguous results. On the one hand, adding a T security adds an independent signal, which improves overall informativeness. On the other hand, it reduces the informativeness of the $\theta + T$ security, which reduces informativeness overall.

In summary, adding traded securities might reduce the informativeness of the price system, and hence it is not always a solution to the government's problem of inferring information from prices. The key complication arises due to conflicting trading motives – speculation and hedging – that are introduced into the model once there are multiple securities, which might harm informativeness. We show, via a concrete example, that adding a security may be bad for the government's ability to learn from the price and consequently may reduce the value of its objective function.[13] This insight should be considered on top

[13] For related analysis, see Cao (1999) and Bhattacharya, Reny and Spiegel (1995). Our result is different than those in both papers: Cao (1995) studies the effect on costly information acquisition, while Bhat-

of the fact that adding securities is not easy to implement, given that markets have to be liquid enough and that there should be a reasonable way to verify the payoffs for securities to be implementable.[14]

7 Conclusion

Our paper analyzes how market-based government policy affects the trading incentives of risk-averse speculators in a rational-expectations model of financial markets. We show that when the government takes a moderate corrective action, basing this action on the market price creates more trading risks for speculators. This harms their trading incentives, and hence the ability of the financial market to aggregate information and the informativeness of the price as a signal for government policy. The opposite happens when the government takes an amplifying action.

Our analysis shows that the use of market prices as an input for policy might not come for free and might damage the informational content of market prices themselves. Hence, in some cases the government would be better off limiting its reliance on market prices and increasing their informational content. Yet, the government always benefits from some reliance on market prices. Also, and counter to common belief, transparency by the government might be a bad idea in that it might reduce trading incentives and price informativeness, leading to a lower value for the government's objective function.

While we focus in this paper on market-based government policy, our analysis and results apply more generally for *any* action that is based on the price. For example, similar effects will arise if a corporate-governance action – such as replacement of the CEO – is taken by the board of directors upon a decrease in market valuation. Another example is the idea of contingent capital that is gaining momentum recently as a potential solution to banking crises. Financing banks with contingent capital implies that a bank's debt will be converted

tacharya, Reny and Spiegel's (1995) analysis is based on the complete breakdown of a trading equilibrium.

[14] θ is likely to be non-verifiable, as it is not the actual cash flow generated by the firm. Instead, $\theta + T$ is the actual cash flow, and hence the object that is likely to be verifiable.

into equity upon reduction in its market value. This is in order to allow banks financing flexibility when it is most needed. Since such market-based conversion is essentially a market-based corrective action, our analysis in this paper suggests that it could reduce the information in the price and hence the efficiency of the conversion trigger.

Our model postulates a quadratic objective function for the government, which proves to be very useful for tractability and allows us to focus on the interaction between government actions and market prices. In future research, it would be interesting to derive the government's objective function from first principles, relying on some market friction that makes government intervention desirable. It would also be interesting to consider non-linear equilibria where intervention is a discrete event.[15] It is a significant challenge to consider such extensions while maintaining tractability.

Another direction for future research is to consider different motives for market-based government actions. Our analysis focuses on the informational role of prices, which implies that relying on prices enables the government to make more efficient decisions. Another rationale for market-based actions is that they enable the government to commit to take welfare-improving actions when it has different objectives that might lead it to deviate from maximizing overall welfare. It would be interesting to understand the feedback loop between prices and actions in such a model.

Finally, inferring information from prices might be difficult for other reasons than those highlighted by our paper. In practice, speculators trade on various dimensions of information; only some of them are interesting to the government. Hence, it might be hard for the government to elicit exactly the type of information it desires. Such considerations can be introduced into our model in future research.

[15] Bond, Goldstein, and Prescott (2010) analyze such equilibria, but do not consider the process of price formation, which is our focus here.

References

[1] Admati, A. 1985. A Noisy Rational Expectations Equilibrium for Multi-Asset Securities Markets. *Econometrica* 53: 629-57.

[2] Angeletos, G. M., and A. Pavan. 2007. Efficient Use of Information and Social Value of Information. *Econometrica* 75: 1103-42.

[3] Bernanke, B. S., and M. Woodford. 1997. Inflation Forecasts and Monetary Policy. *Journal of Money, Credit, and Banking* 29: 653-84.

[4] Bhattacharya, U., P. Reny, and M. Spiegel. 1995. Destructive Interference in an Imperfectly Competitive Multi-Security Market. *Journal of Economic Theory* 65: 136-70.

[5] Bond, P., and H. Eraslan. 2008. Information-Based Trade. *Journal of Economic Theory* forthcoming.

[6] Bond, P., I. Goldstein, and E. S. Prescott. 2010. Market-Based Corrective Actions. *Review of Financial Studies* 23: 781-820.

[7] Boot, A., and A. Thakor. 1997. Financial System Architecture. *Review of Financial Studies* 10:693-733.

[8] Cao, H. 1999. The Effect of Derivative Assets on Information Acquisition and Price Behavior in a Rational Expectations Equilibrium. *Review of Financial Studies* 12: 131-63.

[9] Dow, J., I. Goldstein, and A. Guembel. 2010. Incentives for Information Production in Markets where Prices Affect Real Investment. Working Paper.

[10] Dow, J., and G. Gorton. 1997. Stock Market Efficiency and Economic Efficiency: Is There a Connection? *Journal of Finance* 52:1087-129.

[11] Evanoff, D. D., and L. D. Wall. 2004. Subordinated Debt as Bank Capital: A Proposal for Regulatory Reform. *Federal Reserve Bank of Chicago Economic Perspectives* 24:40-53.

[12] Faure-Grimaud, A. 2002. Using Stock Price Information to Regulate Firms. *Review of Economic Studies* 69: 169-90.

[13] Feldman, R., and J. Schmidt. 2003. Supervisory Use of Market Data in the Federal Reserve System. Working Paper, Federal Reserve Bank of Minneapolis.

[14] Fishman, M., and K. Hagerty. 1992. Insider Trading and the Efficiency of Stock Prices. *RAND Journal of Economics* 23:106-22.

[15] Flannery, M. 2009. Stabilizing Large Financial Institutions with Contingent Capital Certificates. Working Paper.

[16] Foucault, T., and T. Gehrig. 2008. Stock Price Informativeness, Cross-Listing, and Investment Decisions. *Journal of Financial Economics* 88:146-68.

[17] Fulghieri, P., and D. Lukin. 2001. Information Production, Dilution Costs, and Optimal Security Design. *Journal of Financial Economics* 61:3-42.

[18] Furlong, F., and R. Williams. 2006. Financial Market Signals and Banking Supervision: Are Current Practices Consistent with Research Findings? *Federal Reserve Bank of San Francisco Economic Review*: 17-29.

[19] Ganguli, J., and L. Yang. 2008. Complementarities, Multiplicity, and Supply Information. *Journal of the European Economic Association* 7: 90-115.

[20] Goldstein, I., and A. Guembel. 2008. Manipulation and the Allocational Role of Prices. *Review of Economic Studies* 75:133-64.

[21] Goldstein, I., E. Ozdenoren, and K. Yuan. 2010. Learning and Complementarities in Speculative Attacks. *Review of Economic Studies* forthcoming.

[22] Grossman, S. J., and J. E. Stiglitz. 1980. On the Impossibility of Informationally Efficient Markets. *American Economic Review* 70:393-408.

[23] Hart, O., and L. Zingales. 2009. A New Capital Regulation for Large Financial Institutions. Working Paper.

[24] Hellwig, M. 1980, "On the Aggregation of Information in Competitive Markets," *Journal of Economic Theory*, 22, 477-98.

[25] Herring, R. J. 2004. The Subordinated Debt Alternative to Basel II. *Journal of Financial Stability* 1:137-55.

[26] Khanna, N., S. L. Slezak, and M. Bradley. 1994. Insider Trading, Outside Search, and Resource Allocation: Why Firms and Society May Disagree on Insider Trading Restrictions. *Review of Financial Studies* 7:575-608.

[27] Krainer, J., and J. A. Lopez. 2004. Incorporating Equity Market Information into Supervisory Monitoring Models. *Journal of Money Credit and Banking* 36:1043-67.

[28] Lehar, A., D. Seppi, and G. Strobl. 2010. Using Price Information as an Instrument of Market Discipline in Regulating Bank Risk. Working Paper.

[29] Leland, H. 1992, "Insider Trading: Should it be Prohibited?", *Journal of Political Economy*, 100, 859-87.

[30] McDonald, R. 2010. Contingent Capital with a Dual Price Trigger. Working Paper.

[31] Morris, S., and H. S. Shin. 2002. Social Value of Public Information. *American Economic Review* 92: 1521-34.

[32] Piazzesi, M. 2005. Bond Yields and the Federal Reserve. *Journal of Political Economy* 113:311-44.

[33] Rochet, J. C. 2004. Rebalancing the Three Pillars of Basel II. *Federal Reserve Bank of New York Policy Review* 10:7-21.

[34] Subrahmanyam, A., and S. Titman. 1999. The Going-Public Decision and the Development of Financial Markets. *Journal of Finance* 54:1045-82.

A Appendix

Proof of Proposition 1: We show that it is possible to choose constants p_0, ρ and p_Z such that $P = p_0 + \rho p_Z \theta + p_Z Z$ is an equilibrium.

Rewriting (8) more explicitly, the government's estimate of the fundamental, conditional on the price and its own signal s_G, is

$$E[\theta|s_G, P] = \frac{\tau_\theta \bar{\theta} + \rho^2 \tau_Z \tilde{P} + \tau_G s_G}{T_G(\rho)},$$

where $\tilde{P} \equiv \frac{1}{\rho p_Z}(P - p_0)$ and $T_G(\rho) \equiv \tau_\theta + \rho^2 \tau_z + \tau_G$ is the precision of the government's estimate of θ. So the government's intervention is

$$T = \lambda \left(\widehat{\theta} - \frac{\tau_\theta \bar{\theta} + \rho^2 \tau_Z \tilde{P} + \tau_G s_G}{T_G(\rho)} \right) = \lambda \left(\widehat{\theta} - \frac{\tau_\theta \bar{\theta} + \rho^2 \tau_Z \tilde{P}}{T_G(\rho)} - w(\rho)\theta - w(\rho)\varepsilon_G \right),$$

where $w(\rho) = \frac{\tau_G}{T_G(\rho)}$ is the weight the government puts on its own signal in estimating θ.

Conditional on seeing signal s_i and price P, a speculator's conditional expectation of the government signal s_G is

$$E[s_G|s_i, P] = E[\theta|s_i, P] = \frac{\tau_\theta \bar{\theta} + \rho^2 \tau_Z \tilde{P} + \tau_\varepsilon s_i}{T_\varepsilon(\rho)},$$

where $T_\varepsilon(\rho) \equiv \tau_\theta + \rho^2 \tau_z + \tau_\varepsilon$ is the precision of the investor's estimate of θ. Hence an investor's estimate of the cash flow net of intervention, $\theta + T$, is

$$E\left[\theta + T | s_i, P\right] = \lambda\left(\hat{\theta} - \frac{\tau_\theta\bar{\theta} + \rho^2\tau_Z\tilde{P}}{T_G(\rho)}\right) + (1 - \lambda w(\rho)) E\left[\theta | s_i, P\right],$$

and the precision of his estimate of $\theta + T$ is

$$\left((1 - \lambda w(\rho))^2 T_\varepsilon(\rho)^{-1} + (\lambda w(\rho))^2 \tau_G^{-1}\right)^{-1}.$$

From (11), total demand by all speculators is

$$\int x_i(s_i, P)\, di = \frac{1}{\alpha}\frac{\lambda\left(\hat{\theta} - \frac{\tau_\theta\bar{\theta} + \rho^2\tau_Z\tilde{P}}{T_G(\rho)}\right) + (1 - \lambda w(\rho))\frac{\tau_\theta\bar{\theta} + \rho^2\tau_Z\tilde{P} + \tau_\varepsilon\theta}{T_\varepsilon(\rho)} - P}{(1 - \lambda w(\rho))^2 T_\varepsilon(\rho)^{-1} + (\lambda w(\rho))^2 \tau_G^{-1}}.$$

This is a linear expression in the random variables θ and Z. Consequently, market clearing (5) is satisfied for all θ and Z if and only if the coefficients on θ and Z both equal zero (the price intercept p_0 is then chosen to make sure total speculator demand equals supply $-Z$), i.e.,

$$-\lambda\frac{\rho^2\tau_Z}{T_G(\rho)} + (1 - \lambda w(\rho))\left(\frac{\rho^2\tau_Z}{T_\varepsilon(\rho)} + \frac{\tau_\varepsilon}{T_\varepsilon(\rho)}\right) - \rho p_Z = 0 \tag{23}$$

and

$$-\rho^{-1}\lambda\frac{\rho^2\tau_Z}{T_G(\rho)} + \rho^{-1}(1 - \lambda w(\rho))\frac{\rho^2\tau_Z}{T_\varepsilon(\rho)} - p_Z + \alpha\left((1 - \lambda w(\rho))^2 T_\varepsilon(\rho)^{-1} + (\lambda w(\rho))^2 \tau_G^{-1}\right) = 0. \tag{24}$$

Subtracting (23) from ρ times (24) yields

$$-(1 - \lambda w(\rho))\frac{\tau_\varepsilon}{T_\varepsilon(\rho)} + \alpha\rho\left((1 - \lambda w(\rho))^2 T_\varepsilon(\rho)^{-1} + (\lambda w(\rho))^2 \tau_G^{-1}\right) = 0, \tag{25}$$

an equation of ρ only (observe that this matches equation (12) in the main text). Note that the pair of equations (23) and (24) hold if and only if the pair (23) and (25) hold. So to complete the proof of equilibrium existence, it suffices to show that there exists ρ solving

(25), since p_Z can then be chosen freely to solve (23).

Since $(1-\lambda w)^2 = 1-\lambda w - \lambda w(1-\lambda w)$, equation (25) can be rewritten as

$$(\alpha\rho - \tau_\varepsilon)(1-\lambda w(\rho)) - \alpha\rho\left(\lambda w(\rho)(1-\lambda w(\rho)) - (\lambda w(\rho))^2 \tau_G^{-1} T_\varepsilon(\rho)\right) = 0.$$

Defining

$$F(\rho, w) \equiv 1 - \frac{\tau_\varepsilon}{\alpha\rho} - \lambda w + \frac{\lambda^2 w^2}{1-\lambda w}\frac{T_\varepsilon(\rho)}{\tau_G},$$

equation (25) is equivalent to

$$F(\rho, w(\rho)) = 0.$$

Note that $w(\rho)$ is decreasing in ρ, with $w(\rho) < 1$ for $\rho = 0$, and $w(\rho) \to 0$ as $\rho \to \infty$. So $F(\rho, w(\rho))$ approaches $-\infty$ as ρ approaches 0, and approaches 1 as $\rho \to \infty$. By continuity, it follows that (25) has a solution, completing the proof of equilibrium existence.

For uniqueness, first note that at $\lambda = 0$, the unique solution of $F(\rho, w(\rho)) = 0$ is $\rho = \frac{\tau_\varepsilon}{\alpha}$. To establish uniqueness for sufficiently small but strictly positive values of λ, proceed as follows. Fix $\bar{\lambda} \in (0,1)$; choose $\underline{\rho}$ such that $F(\rho, w(\rho)) < 0$ for all $\rho \leq \underline{\rho}$ and $\lambda \in [0, \bar{\lambda}]$; and choose $\bar{\rho} > \underline{\rho}$ such that $F(\rho, w(\rho)) > 0$ for all $\rho \geq \underline{\rho}$ and $\lambda \in [0, \bar{\lambda}]$ (the existence of $\underline{\rho}$ and $\bar{\rho}$ with these properties is easily established). At $\lambda = 0$, $\frac{d}{d\rho} F(\rho, w(\rho))\big|_{\rho = \frac{\tau_\varepsilon}{\alpha}} > 0$. Consequently, there exists some $\delta > 0$ such that for all $\lambda \in [0, \bar{\lambda}]$, $\frac{d}{d\rho} F(\rho, w(\rho)) > 0$ for all $\rho \in \left(\frac{\tau_\varepsilon}{\alpha} - \delta, \frac{\tau_\varepsilon}{\alpha} + \delta\right)$. So for all λ sufficiently small, $F(\rho, w(\rho)) = 0$ has a unique solution in $\left(\frac{\tau_\varepsilon}{\alpha} - \delta, \frac{\tau_\varepsilon}{\alpha} + \delta\right)$; by uniform convergence has no solution in the compact set $\left[\underline{\rho}, \frac{\tau_\varepsilon}{\alpha} - \delta\right] \cup \left[\frac{\tau_\varepsilon}{\alpha} + \delta, \bar{\rho}\right]$; and has no solution below $\underline{\rho}$ or above $\bar{\rho}$. Finally, a parallel proof implies uniqueness for the case of λ strictly negative and sufficiently close to 0. ■

Proof of Proposition 2: Let ρ^* and ρ_{-P} denote equilibrium price informativeness for the cases in which the government uses the price in an ex post optimal way and in which the government completely ignores the price, respectively. Let $F(\rho, w)$ be as defined in the proof of Proposition 1.

We now show that for λ positive and sufficiently small, $\rho_{-P} > \rho^*$. As λ approaches 0, both ρ_{-P} and ρ^* approach $\frac{\tau_\varepsilon}{\alpha}$ (and moreover, ρ^* is uniquely defined by Proposition 1). Fix $\delta > 0$, and choose $\hat{\lambda}$ such that if $\lambda \in \left(0, \hat{\lambda}\right)$, then both ρ_{-P} and ρ^* lie within δ of $\frac{\tau_\varepsilon}{\alpha}$. Because $w_{-P} > w(\rho)$, there exists $\check{\lambda} \in \left(0, \hat{\lambda}\right)$ such that if $\lambda \in \left(0, \check{\lambda}\right)$ then $F(\rho, w(\rho)) > F(\rho, w_{-P})$ for all ρ within δ of $\frac{\tau_\varepsilon}{\alpha}$. Consequently, if $\lambda \in \left(0, \check{\lambda}\right)$ then $0 = F(\rho^*, w(\rho^*)) > F(\rho^*, w_{-P})$, which since $F_\rho > 0$ implies $\rho_{-P} > \rho^*$.■

Proof of Proposition 3: From the paragraph prior to the statement of Proposition 3, it suffices to show that a small increase in $\tilde{w}$ above $w(\rho)$ increases equilibrium price informativeness. Let $F(\rho, w)$ be as defined in the proof of Proposition 1, so that $F(\rho, w(\rho)) = 0$. Because $F_\rho > 0$, we must show $F_w(\rho, w(\rho)) < 0$. This is indeed the case for all λ strictly positive and sufficiently close to 0, completing the proof. ■

Proof of Proposition 4: Let $F(\rho, w)$ be as defined in the proof of Proposition 1, so that equilibrium price informativeness satisfies $F(\rho, w(\rho)) = 0$. Hence $\frac{d\rho}{d\tau_G}$ satisfies

$$\begin{aligned} 0 &= \frac{d\rho}{d\tau_G}\left(F_\rho(\rho, w(\rho)) + w'(\rho) F_w(\rho, w(\rho))\right) \\ &\quad + \frac{dw(\rho)}{d\tau_G} F_w(\rho, w(\rho)) + \frac{d}{d\tau_G} F(\rho, w(\rho)). \end{aligned} \tag{26}$$

As in the proof of Proposition 3, $F_w(\rho, w(\rho)) < 0$ for λ strictly positive and sufficiently close to 0. Moreover, $F_\rho > 0$, $w'(\rho) < 0$, $\frac{dw(\rho)}{d\tau_G} > 0$, and $\frac{dF}{d\tau_G} < 0$. Hence $\frac{d\rho}{d\tau_G} > 0$ for λ strictly positive and sufficiently close to 0, completing the proof. ■

Proof of Proposition 5: Part (A) follows on the proof of Proposition 2: for the case of $\lambda < 0$, $F_w > 0$, and so, $0 = F(\rho^*, w(\rho^*)) < F(\rho^*, w_{-P})$, which since $F_\rho > 0$ implies $\rho_{-P} < \rho^*$. Similarly, part (B) follows from straightforward adaptation of the analogous result in Proposition 3. Part (C) also builds on the proof of Proposition 4. Note that

$$\begin{aligned} F_\rho(\rho, w(\rho)) + w'(\rho) F_w(\rho, w(\rho)) &= \frac{\tau_\varepsilon}{\alpha\rho^2} + \text{terms in } \lambda \\ \frac{dw(\rho)}{d\tau_G} F_w(\rho, w(\rho)) &= -\lambda \frac{dw(\rho)}{d\tau_G} + \text{terms in } \lambda^2 \\ \frac{d}{d\tau_G} F(\rho, w(\rho)) &= \text{terms in } \lambda^2. \end{aligned}$$

So for $\lambda < 0$ sufficiently close to 0, it follows from (26) that $\frac{d\rho}{d\tau_G} < 0$, completing the proof. ■

Proof of Proposition 6: Given the main text, it is sufficient to formally show that equilibrium price informativeness is increasing in λ for $\lambda > 0$ and sufficiently small. Let $F(\rho, w)$ be as defined in the proof of Proposition 1, so that equilibrium price informativeness satisfies $F(\rho, w(\rho)) = 0$. Hence $\frac{d\rho}{d\lambda}$ satisfies

$$0 = \frac{d\rho}{d\lambda}\left(F_\rho(\rho, w(\rho)) + w'(\rho) F_w(\rho, w(\rho))\right) + \frac{d}{d\lambda} F(\rho, w(\rho)).$$

As in the proof of Proposition 4, we know $F_\rho > 0$, $w'(\rho) < 0$; and when λ is positive and sufficiently close to 0, $F_w(\rho, w(\rho)) < 0$. Moreover, $\frac{d}{d\lambda} F$ is negative for λ sufficiently close to zero. Hence $\frac{d\rho}{d\lambda} > 0$ for λ positive and sufficiently close to 0. ■

Proof of Proposition 7: See the main text following Proposition 7. ■

Proof of Proposition 8: The equilibrium condition under transparency is (25) (see proof of Proposition 1). The equilibrium condition without transparency has an additional term $\alpha\rho\lambda^2 var\left(\hat{\theta}\right)$ on the lefthand side, but it otherwise identical. The lefthand side of both conditions is negative for ρ sufficiently small, and positive for ρ sufficiently large. Consequently, both the minimum and maximum equilibrium levels of informativeness are higher under transparency. The equilibrium is unique in both cases when λ is sufficiently close to 0 (see Proposition 1), implying the result. ■

B Additional numerical appendix

As we note in the main text, the effect of government corrective actions on price informativeness depends on the size of the corrective action. In the main text we focus on the case in which the corrective action is "mild," or, more mathematically, "sufficiently small." We emphasize in the main text that this does not mean economically small, and refer to numerical simulations that show that corrective actions as large as $\lambda = 30\%$ are still sufficiently small for all our results to hold. Here, we present the details of these numerical simulations.

B.1 Numerical solution of the model

We start by detailing the numerical solution of the model. As shown in the proof of Proposition 1, equilibrium price informativeness ρ solves $F(\rho, w(\rho)) = 0$, where F is as defined in the proof. Dividing by $w(\rho)$ implies that ρ solves

$$0 = \frac{1}{w(\rho)}\left(1 - \frac{\tau_\varepsilon}{\alpha\rho}\right) - \lambda + \frac{\lambda^2}{\frac{1}{w(\rho)} - \lambda}\frac{\tau_z\rho^2 + \tau_\theta + \tau_\varepsilon}{\tau_G},$$

or equivalently

$$0 = \frac{1}{\tau_G}\left(\tau_z\rho^2 + \tau_\theta + \tau_G\right)\left(1 - \frac{\tau_\varepsilon}{\alpha\rho}\right) - \lambda + \lambda^2\frac{\tau_z\rho^2 + \tau_\theta + \tau_\varepsilon}{\tau_z\rho^2 + \tau_\theta + \tau_G - \lambda\tau_G},$$

or equivalently

$$\begin{aligned} 0 \;&=\; \left(\tau_z\rho^2 + \tau_\theta + \tau_G - \lambda\tau_G\right)\left(\tau_z\rho^2 + \tau_\theta + \tau_G\right)(\alpha\rho - \tau_\varepsilon) \\ &\quad -\lambda\alpha\tau_G\rho\left(\tau_z\rho^2 + \tau_\theta + \tau_G - \lambda\tau_G\right) + \lambda^2\alpha\tau_G\rho\left(\tau_z\rho^2 + \tau_\theta + \tau_\varepsilon\right), \end{aligned}$$

or equivalently

$$\begin{aligned} 0 \;&=\; [\tau_z^2\rho^4 + \tau_z(2\tau_\theta + (2-\lambda)\tau_G)\rho^2 + (\tau_\theta + \tau_G)(\tau_\theta + (1-\lambda)\tau_G)](\alpha\rho - \tau_\varepsilon) \\ &\quad +\lambda\alpha\tau_G\rho[(\lambda - 1)\tau_z\rho^2 - (\tau_\theta + (1-\lambda)\tau_G) + \lambda(\tau_\theta + \tau_\varepsilon)]. \end{aligned}$$

Rewriting a final time, the equilibrium condition is equivalent to the fifth-degree polynomial

$$\begin{aligned} 0 \;&=\; \alpha\tau_z^2\rho^5 - \tau_\varepsilon\tau_z^2\rho^4 + [2\tau_\theta + (2-\lambda)\tau_G + \lambda(\lambda-1)\tau_G]\tau_z\alpha\rho^3 \\ &\quad -[2\tau_\theta + (2-\lambda)\tau_G]\tau_\varepsilon\tau_z\rho^2 + A\alpha\rho - B \end{aligned} \qquad (27)$$

where

$$
\begin{aligned}
A &= (\tau_\theta + \tau_G)(\tau_\theta + (1-\lambda)\tau_G) - \lambda\tau_G\left(\tau_\theta + (1-\lambda)\,\tau_G\right) + \lambda^2\tau_G\left(\tau_\theta + \tau_\varepsilon\right) \\
&= (\tau_\theta + (1-\lambda)\tau_G)^2 + \lambda^2\tau_G\left(\tau_\theta + \tau_\varepsilon\right) \\
B &= \tau_\varepsilon(\tau_\theta + \tau_G)(\tau_\theta + (1-\lambda)\tau_G).
\end{aligned}
$$

Solutions to (27) can be found using any standard numerical procedure for finding the roots of polynomials.

B.2 Numerical simulations

The parameters of the model are α, τ_θ, τ_G, τ_ε and τ_Z. Note first that the equilibrium condition $F\left(\rho, w\left(\rho\right)\right) = 0$ is homogeneous of degree zero in the vector of these five parameters. Consequently, it is sufficient to specify the four ratios $\frac{\tau_\theta}{\alpha}$, $\frac{\tau_Z}{\alpha}$, $\frac{\tau_G}{\tau_\theta}$ and $\frac{\tau_\varepsilon}{\tau_G}$.

Let ϕ denote the fraction of price fluctuations that are not attributable to changes in the fundamental θ, for the case in which government intervention is completely absent. From the paper,

$$
\phi = \left(\frac{\tau_\theta}{\frac{\tau_\varepsilon^2\tau_Z}{\alpha^2} + \tau_\theta}\right)^{1/2} = \left(\frac{\tau_\varepsilon^2\tau_Z}{\alpha^2\tau_\theta} + 1\right)^{-1/2} = \left(\frac{\left(\frac{\tau_\varepsilon}{\tau_G}\frac{\tau_G}{\tau_\theta}\frac{\tau_\theta}{\alpha}\right)^2\frac{\tau_Z}{\alpha}}{\frac{\tau_\theta}{\alpha}} + 1\right)^{-1/2},
$$

and so

$$
\frac{\tau_Z}{\alpha} = \frac{\left(\phi^{-2} - 1\right)\frac{\tau_\theta}{\alpha}}{\left(\frac{\tau_\varepsilon}{\tau_G}\frac{\tau_G}{\tau_\theta}\frac{\tau_\theta}{\alpha}\right)^2}.
$$

Consequently, it is sufficient to specify $\frac{\tau_\theta}{\alpha}$, $\frac{\tau_G}{\tau_\theta}$, $\frac{\tau_\varepsilon}{\tau_G}$, together with ϕ.

We simulate the model for values of ϕ (the fraction of price fluctuations that are not attributable to changes in the fundamental θ) of 10%, 50%, and 90%. Likewise, we simulate the model for values of τ_ε/τ_G (the ratio of the precisions of an *individual* speculator's private forecast to the government's) of 10%, 50%, and 90%. In both cases, these ranges more than cover what most people would regard as reasonable values of these parameters.

We have much weaker priors for reasonable values of $\frac{\tau_\theta}{\alpha}$ and $\frac{\tau_G}{\tau_\theta}$. For these parameters, we simply simulate the model over a fine grid of possible values for both parameters, ranging from 1/100 up to 100.

We simulate the model for each possible combination of these four parameters. For each combination of parameter values, we check whether the equilibrium is unique, and whether the derivative F_w (the function F is as defined in the proof of Proposition 1) is negative at the equilibrium value of ρ (this is the condition for which we need λ to be sufficiently small in our analysis).

For values of λ up to $\lambda = 30\%$, we find that both conditions are satisfied *for all* parameter values in the ranges detailed above. As we note in the main text, a corrective action of 30% is economically large, and indeed is considerably above our prior of the likely scale of government interventions. Moreover, we also emphasize that both conditions above are also satisfied for many parameter values even when λ is even higher than 30%.

Is Historical Cost Accounting a Panacea? Market Stress, Incentive Distortions, and Gains Trading

Andrew Ellul
Indiana University, ECGI and CSEF

Chotibhak Jotikasthira
University of North Carolina

Christian T. Lundblad
University of North Carolina

Yihui Wang
Chinese University of Hong Kong

December 12, 2011

Abstract

This paper explores the trading incentives of financial institutions induced by the interaction between regulatory accounting rules and capital requirements by investigating insurance companies' trading behavior during the recent financial crisis. According to insurance regulation, life insurers have a greater degree of flexibility to hold downgraded instruments at historical cost, whereas property and casualty insurers are forced to re-mark many of their downgraded securities to market prices. Using firm-level insurance company transaction and position data, we study the implications of this accounting difference, and document direct evidence of 'gains trading' associated with historical cost accounting during the financial crisis. When faced with severe downgrades among their holdings in asset-backed securities (ABS), life insurers largely continue to hold the downgraded securities at historical cost and instead selectively sell their corporate bond holdings with the highest unrealized gains. This is particularly true for insurers facing regulatory capital constraints and with high ABS exposures. This behavior is largely absent among property and casualty insurers; they instead disproportionately sell their re-marked ABS holdings. Finally, we find that the gains trading among life companies induces significant price declines in the otherwise unrelated corporate bonds that happen to exhibit high unrealized gains.

JEL classifications: G11; G12; G14; G18; G22

Keywords: Regulation; Mark to market; Historical cost accounting; Gains trading; Fire sales; Asset-backed securities (ABS); Corporate bonds; Insurance companies

Acknowledgments: We are grateful for comments received from Utpal Bhattacharya, Mike Burkart, Scott Harrington, Sreenivas Kamma, Peter Kondor, Martin le Roux, Dong Lou, Marco Pagano, Lubos Pastor, Dimitri Vayanos and participants in seminars at Imperial College London, Indiana University, London School of Economics and University of Warwick. We are especially grateful to Robert Hartwig of the Insurance Information Institute for several detailed discussions. This research was conducted, in part, while Lundblad was visiting the Einaudi Institute for Economics and Finance in Rome, Italy. Wang acknowledges the financial support from Rega Technologies Limited.

1. Introduction

This paper explores the trading incentives of financial institutions induced by the interaction between regulatory accounting rules and capital requirements. The theoretical literature (see, for example, Allen and Carletti (2008), Plantin, Sapra and Shin (2008), and Sapra (2008)) argues that mark-to-market (MTM), or fair value, accounting leads to the forced selling of assets by financial institutions during times of market stress, resulting in a downward spiral of liquidity and prices and potential contagion effects for other markets. In contrast, these authors contend that historical cost accounting (HCA) may avoid fire sales and contagion effects.[1] This paper challenges this view by providing new empirical evidence that HCA, along with regulatory capital requirements, induces an altered incentive to "gains trade" where, in order to shore up capital, an institution selectively sells otherwise unrelated assets with high *unrealized* gains. Critically, it is important not to consider the accounting treatment in isolation, but rather how the different treatments interact with capital regulations (Heaton, Lucas, and McDonald (2010)) to influence financial institutions' trading incentives (Laux and Leuz (2009, 2010)).

The role of MTM during the recent financial crisis has generated an intense debate. The accounting rules followed by financial institutions may appear to simply be an issue of measurement and, in frictionless markets, free of any impact on economic fundamentals. However, when markets are illiquid and trading frictions elevated, financial assets may temporarily trade at market prices that are well below fundamental values (Duffie (2010), AFA Presidential Address). In such an environment, write-downs (and the associated deterioration of financial institutions' asset values) will lead to an erosion of their capital base, potentially forcing the liquidation of some assets. Allen and Carletti (2008) argue that in such a market environment, HCA will avoid fire sales because financial institutions would not suffer from a deterioration of their asset valuations in the first place. Plantin, Sapra and Shin (2008) also argue that MTM generates inefficiencies because it injects excessive volatility in prices that naturally degrades their information content and leads to sub-optimal decisions by financial institutions.

HCA may also engender inefficiencies as financial institutions, under HCA, have an incentive to engage in selective asset sales aimed at the early realization of earnings (see Laux and

[1] This is a view that has received support from the banking industry as well. In a letter to the SEC in September 2008, the American Bankers Association was of the opinion that, among several factors that led to the financial crisis, "one factor that is recognized as having exacerbated these problems is fair value accounting."

Leuz (2009) for a discussion of the gains trading incentive under HCA). Indeed, Plantin, Sapra and Shin (2008) recognize that HCA is not immune to these inefficiencies in normal times. In this paper, we focus on the implications of this trading incentive and its impact on financial institutions' trading behavior during times of *market stress*. Below, we argue that it is precisely these times that financial institutions have the highest need to realize gains in order to improve capital positions.

We argue that the crucial issue in the debate surrounding the accounting treatment of financial assets and its impact on financial institutions during periods of market stress relates to the *interaction between the accounting regime and the institutional framework*, specifically regulatory capital requirements. To focus ideas, consider a financial institution that invested heavily in Asset-Backed Securities (ABS) in the years leading up to the financial crisis. The severe downgrades of such instruments that occurred during the 2007-2009 period, taking many such holdings from investment to speculative grades, significantly affected the regulatory capital of various financial institutions holding the downgraded instruments. An institution affected then faced a stark decision: either keep the downgraded instruments and find additional capital elsewhere or sell the downgraded instruments to reduce the required regulatory capital. At the same time, the downgraded instruments likely experienced severe price declines. A crucial input in the institution's decision is the accounting treatment used for the downgraded instruments (as well as the accounting treatment for other assets in its portfolio).

If the downgraded asset is held at market value, the price decline would be automatically reflected in the balance sheet, and the loss will flow to the income statement, impairing the institution's capital. From a purely accounting point of view, the institution will be indifferent between keeping the asset on the balance sheet and selling it. However, considering the regulatory capital dimension, selling the downgraded asset has an advantage as swapping a risky asset for cash reduces the required regulatory capital. The disadvantage of selling is that trades will take place in a market already characterized by severe price declines and illiquidity.

The situation is different if the asset is held under HCA. In line with Allen and Carletti (2008), the decline in value will not be recognized in the balance sheet, but crucially, the institution still has to act because its regulatory capital would have increased as a result of the downgrade. Holding the asset has the advantage of limiting the unfavorable price impact, but additional capital needs to be raised. It is precisely in this situation that the incentive for gains trading arises. The institution may sell other existing assets that have not been downgraded to shore up its capital position. Importantly, the institution faces an altered incentive to do so by *selectively selling* those assets that are held under HCA and have the *largest unrealized gains*. By selling such assets, these unrealized gains can be recognized and flow to its capital.

The question then becomes whether such selective selling engenders significant price pressure in the selected assets with high unrealized gains. If so, one can conclude that HCA – precisely because of the interaction between accounting and capital regulations – does not completely avoid illiquidity spillovers.

We investigate this gains trading mechanism during times of market stress, by examining the behavior of 1,882 insurance companies following severe downgrades within the ABS market. Most importantly, we exploit the different accounting treatments used in determining the required capital for holding speculative-grade assets, under the National Association of Insurance Commissioners (NAIC)'s model law, for life and for property and casualty (P&C) insurance companies. In the case of an ABS downgraded from investment to non-investment grade, P&C insurers have to immediately recognize the value of the ABS as the lower between the amortized value (based on HCA) and the market price (or model price, in case no market price is available). On the other hand, life insurers can continue to hold the downgraded ABS under HCA except in the extreme case when it is classified as 'in or near default' (Class 6).[2] Given these distinctions in accounting treatment and the NAIC's security-level data, the insurance industry presents an interesting laboratory to explore the interplay between accounting and regulatory capital requirements for financial institutions.

We construct a dataset of 34,957 downgrades of non-agency ABS to speculative-grade by S&P over the period 2005-2010 using S&P's Ratings IQuery. We combine information on these securities with *firm-level* observations, provided by the NAIC, on insurance companies' holdings of and transactions in individual ABS and corporate bonds. Further, for each ABS and corporate bond position, insurance companies provide both fair and book values to the NAIC. We obtain data on the financial position and strength of each insurance company from the Street.com.

As of 2007, life and P&C insurance companies held roughly the same amount of non-agency ABS as a percentage of their total bond portfolio (around 5-7%). During the financial

[2] To put the definitions of the asset classes in perspective, a "Class 5" security is one that corresponds to a CCC/Caa credit rating; even in such cases life insurers can continue to hold the asset at HCA while P&C insurers have to recognize the market price if the price falls below the amortized value.

crisis, the downgrades of some of these ABS instruments were severe, with the majority of downgraded securities falling to speculative grade.[3] We find confirmation that the different accounting treatment between life and P&C insurers is triggered when such downgrades occur. For example, from 2004 to 2006, both life and P&C insurance companies hold around 5% of their non-agency ABS positions at market values. In 2008, almost 20% (5%) of the holdings of P&C (life) are held at market values. Given that the exposure of both types of insurers to downgraded ABS is very similar, this evidence provides confirmation of the different accounting regulations across the two groups. Further, the *actual capital* of life companies, due to HCA, is much less affected by the downgrades than the capital of P&C firms (-6% vs. -13% from 2007 to 2008). For both groups, however, the *regulatory capital requirement* increased. The question is then how the different accounting treatments influence the incentives of life and P&C firms to respond to this increase in their required regulatory capital.

Several key empirical results deserve attention. First, during the crisis, we find clear evidence that life firms largely keep the downgraded ABS in their balance sheet and instead engage in gains trading by *selectively* selling corporate bonds. In sharp contrast, P&C firms disproportionately sell their downgraded ABS holdings (35% more likely than life firms). While the selling of the downgraded ABS may take place at fire-sales prices, from an accounting point of view, P&C insurers, having already booked the loss, would be indifferent between holding the asset at the lower value and selling it. Selling the asset has an important advantage from the regulatory capital point of view, as a risky asset will be exchanged for cash thus reducing the capital requirement.

Second, we find that life insurers disproportionately sell the otherwise unrelated corporate bonds that have the highest unrealized gains. Because corporate bonds are also held at historical cost, it is only by the sale transaction that these unrealized gains can be recognized. Following this course of action, life insurers achieved two important objectives: (1) reduce their regulatory capital (exchanging a risky asset for cash) and (2) realize the gain that arises from the HCA treatment. Most importantly, we find that this trading behavior is disproportionately conducted by life insurers that have (a) large exposures to downgraded ABS booked at HCA, and (b) low risk-based capital ratios. Among these insurers, the probability of selling corporate bonds with the highest

[3] For example, out of the 808 ABS held by insurance companies that were downgraded to BB status, 386 were previously rated in the top three credit rating classes, with 299 coming straight from the highest AAA rating class.

unrealized gains increases by more than 50% over the normal selling probabilities.[4] These results are obtained after controlling for standard insurance company and bond characteristics.

Finally, we consider whether the gains trading engaged in by life insurers leads to price pressure in the corporate bond market. If a large number of insurance companies attempt to sell corporate bonds with the largest unrealized gains in a market that is notoriously illiquid, then we should expect significant price pressures. We find that the price at which insurers with the highest pressure sell their bonds is significantly lower than the median price of the same bond during the week of this trade. Further, we also find that the corporate bonds disproportionately targeted by insurers facing the highest propensity to gains trade statistically and economically underperform otherwise similar bonds. The quarterly return is 0.7-1.7% lower as we move from the 25th to 75th percentiles of aggregated gains-trading propensity. No such price impacts are experienced for the corporate bonds selling by P&C insurers.

Overall, these results show that the interactions between accounting treatment, especially HCA, with capital regulations can create unintended consequences where spillover effects and fire sales are not entirely avoided. HCA, through the incentive it creates for gains trading, can still engender price distortions during market stress for bonds that are *completely unrelated* to the original downgraded securities. Thus, using the terminology in Plantin *et al.* (2008) "in such an environment, prices drive measurement, but measurements have an impact on pricing." The results for life insurers are of particular importance to banking institutions since life insurers, in contrast to P&C, have asset and liability structures that resemble those of banks.

Our paper is related to several strands of the literature. We contribute to the growing literature exploring the trading decisions made by institutional investors when faced with a financial shock (for example, Anand *et al.* (2010), Boyson *et al.* (2011), Manconi *et al.* (2011), Hau and Lai (2011), among others). To the best of our knowledge, we are the first to empirically demonstrate the importance of the interaction between accounting and capital regulations on the decisions made by institutional investors and the spillover effects that may occur. One unintended consequence of such an interaction, which we focus on in this paper, is the incentive for gains trading; Beatty, Chamberlain, and Magliolo (1995), Hirst and Hopkins (1998), Hirst *et al.* (2004), Kashyap and Stein (2000), Lee *et al.* (2006), among others, also explore the gains trading (what

[4] The marginal selling probability is calculated by comparing the selling probability between the corporate bonds at the 75th percentile of unrealized gains and the bonds at the 25th percentile.

these articles refer to as the 'cherry picking') phenomenon. In contrast to these earlier efforts, we show that gains trading behavior takes place during periods of market stress and has significant price impacts. Furthermore, we are the first to document such trading behavior *at the security-level*, rather than inferring trading behavior from aggregated data at the institution-level. Finally, our results also contribute to the debate on the choice of accounting system, historical cost vs. marking to market, used in regulating financial institutions.[5] The literature (mostly theory) suggests that during a financial crisis, marking to market may cause distress selling and financial instability (Allen and Carletti (2008), Plantin *et al.* (2008), and Wallison (2008)).[6] We provide evidence in support of Laux and Leuz (2009, 2010)'s conjecture that historical cost accounting is not a panacea either. Historical cost accounting leads to banks' selling of unrelated assets to realize gains, essentially transmitting the shocks from one market to others.

Last, although it is entirely outside the scope of this paper to explore the welfare comparison between HCA and fair value accounting, especially as this relates to the interaction with capital regulation, we acknowledge that HCA may have a positive *net effect* if it helps alleviate aggregate market inefficiency and capital misallocation. Yet, this is not to say that adopting HCA at times of crisis is a panacea for financial institutions.[7] The evidence presented in this paper suggests that the contentious debate about accounting choices cannot exclusively focus on the accounting treatment in isolation but rather has to consider the interaction with the institutional framework in which it is being employed.

The remainder of the paper is organized as follows. Section 2 discusses the sample construction and describes the summary statistics of the data. Section 3 presents our main empirical analysis and discusses the results. Section 4 concludes.

2. The Data

2.1 Sample Construction

We combine three sets of data in our analysis: information on insurance companies, ABS securities and their rating changes, and corporate bonds and their trade prices. We discuss in detail how we

[5] See Goh *et al.* (2009) for a general analysis of the determinants of accounting choice and the effects of fair value disclosure on firms' information environment. See also Eccher, Ramesh, and Thiagarajan (1996), Penman (2007), Petroni and Wahlen (1995), and Wyatt (1991).
[6] See Veron (2008) for an opposing view.
[7] Plantin *et al.* (2008) refer to such welfare comparisons in their conclusions.

assemble the three sets of data below. Our sample period is from 2004 to 2010. This period covers the financial crisis of 2007-2009 and also a non-crisis period that we shall use for comparison. Specifically, we define our crisis period as 2007 Q3 – 2009 Q4, as virtually all significant downgrades of ABS happen during this period.

Our primary data on insurance companies' transactions and year-end positions are from the National Association of Insurance Commissioners (NAIC).[8] The NAIC data provide year-end holdings of invested securities for each insurance firm and detailed transaction information on every trade. Both the position and transaction data provide the identities of the insurance firms and the relevant securities (e.g., 9-digit CUSIP). We merge the year-end position data with transaction data to infer quarter-end positions. Finally, the NAIC data provide detailed information about the *book or adjusted carrying value* and *fair value* of each position held by each insurance company at year-end. We employ this information to infer whether an insurance company holds its ABS and corporate bonds at historical cost or at fair value.

The financial information on each insurance firm is from the Street.com, which provides financial strength ratings. From this source, we obtain annual firm characteristics, such as size, 'capital and surplus', and the risk-based capital (RBC) ratio used for capital regulation. While our Street.com data end in 2007, we use the 2007 to infer subsequent values since these characteristics do not vary much over time. We eliminate small insurers with investment assets less than $13 million (the bottom 1%) and/or with an RBC ratio either below 2 or above 20 to avoid any bias from small or abnormal firms.[9] We also delete all AIG's affiliated insurers and 32 others that provide financial insurance and guarantees for bonds, such as credit default swaps and municipal finance, as these firms were affected by the downgrade of ABS securities through a different channel.[10] Our final sample of insurance firms consists of 13,281 firm-years representing 2,381 firms, among which 709 are life insurers and 1,672 P&C firms.

Our data on ABS ratings are from S&P's Ratings IQuery. We extract all the data in the structured credit subsector in Ratings IQuery, which comprehensively covers initial ratings and histories for all securitized issues rated by S&P from 1991 to 2010. The database records issue

[8] Further details of the NAIC data can be found in Ellul, Jotikasthira, and Lundblad (2011).
[9] Small insurance firms do not have many trading choices. Insurance firms with larger RBC ratios are considered better capitalized. Firms with RBC ratios below two are subject to supervisory intervention. Firms with RBC ratios above 20 are unusual and may behave differently from the average.
[10] We identify bond insurers from Ratings IQuery, which reports financial insurance providers in securitized issues. In addition to AIG, we also exclude Ambac Assurance Corp, MBIA Insurance Corp, Financial Guaranty Insurance Co., etc.

and tranche identity (9-digit CUSIP), gross principle, class, maturity, collateral type, rating, and rating date. With this dataset, we identify 127,719 ABS securities in 13,430 issues.[11] Among all the ABS securities rated by S&P, 65% are mortgage-backed securities, 20% are collateralized debt obligations, and 15% are asset-backed securities backed by consumer loans. We use the list of 9-digit CUSIP of ABS in the rating dataset to identify holdings of ABS by insurers. Over our sample period, 24,452 unique CUSIPs in the portfolios of insurance firms are identified as ABS. The ratings and ratings dates are then used to generate the list of significant ratings downgrades.

The data on corporate bond characteristics and trading are obtained from Mergent Fixed Income Securities Database (FISD) and TRACE. We merge the FISD data with the position and transaction data of insurance firms to identify the corporate bonds being held and transacted as well as the bond characteristics, such as issue size, age, maturity, rating downgrades, and bankruptcy. When we identify downgrades of corporate bonds, we use S&P's ratings whenever they are available, to be consistent with our data source of ABS ratings. When S&P's ratings are missing, we use the ratings from Moody's (or Fitch if Moody's ratings are not available). Data on bond market transaction prices and size are from TRACE, which covers over-the-counter corporate bond market transactions for both investment and speculative grade bonds since 2004. We use the 9-digit CUSIP to merge bonds in FISD and in TRACE.

2.2 Insurance Firms and Their ABS Exposure

Table 1 shows summary statistics on several key financial variables for our sample firms at the end of 2007. A detailed description of the variables can be found in Appendix A.

[Insert Table 1 here]

At the end of 2007, we have complete financial information for 1,344 P&C companies and 538 life companies in the sample. Life firms are larger than P&C firms. Invested assets are $4.7 billion, on average, (median of $393 million) for life firms and $829 million, on average, (median of $119 million) for P&C firms. The average 'capital and surplus' is also larger for life firms at $476

[11] According to SEC (2011), S&P ratings are outstanding for a total of 117,900 ABS securities as of year-end 2010. Note that ABS securities, particularly mortgage-backed securities tend to have long maturities, suggesting that most of the securities in our sample may still exist by year-end 2010. In addition, the majority of the ABS securities were issued after 2000 when this market grew substantially. According the SEC report, S&P rated the largest number of ABS among all rating agencies. The number of S&P's ratings is greater than Moody's 101,546 outstanding ratings.

million (median of $66 million), compared to $384 million (median of $53 million) for P&C firms. In addition, life firms, similar to banks, operate at much higher leverage than P&C firms. Return on equity is at similar levels for the two types of firms.

The capital positions of life and P&C firms are also similar. We use the NAIC risk-based capital ratio (RBC ratio) to measure capitalization. The RBC ratio is the ratio of total adjusted capital to NAIC risk-based capital (RBC), which is the minimum capital under current regulation that an insurance company must maintain given the inherent risks in its operations. It is calculated based on the NAIC's formula which reflects a risk assessment of different asset classes and businesses. Risky assets are weighted more heavily in computing RBC, and higher RBC ratios reflect better capitalization. Insurance companies with RBC ratios below 2 are considered under-capitalized and subject to supervisory interventions. The average life and P&C firms in our sample have RBC ratios of 8.8 and 8.3, respectively. We use the RBC ratio to capture the regulatory constrains that insurance firms may face when their fixed income holdings are downgraded, as downgrades can lead to higher weights assigned on the same assets and hence a lower RBC ratio.

Insurance companies heavily invest in investment-grade bonds, representing 57-60% of their portfolios, on average. The Street.com creates a series of indices to measure insurance companies' liquidity, profitability, and other aspects of their financial conditions. Life and P&C firms are, on average, not systematically different in terms of the Street.com's assessment of profitability and liquidity. Overall, as insurance firms heavily invest in bonds, their trading behavior in this asset class is interesting to analyze.

To see the exposure of the insurance firms to ABS, we report their holdings of ABS over the period 2004-2010 in Table 2.

[Insert Table 2 here]

Life firms have greater exposures to ABS when compared to P&C firms. Based on the left two columns, about 79% (60%) of life (P&C) firms hold ABS. These percentages decline for P&C firms over the crisis period. In the other columns, we report the number and total values of ABS in each year across the firms that hold any ABS. Three features of the data are notable. First, insurance firms' portfolio exposures to ABS were quite large during the crisis. For example, life firms held, on average, 65-68 ABS during 2007-2009, and these securities accounted for about 7% of the par value of their total fixed income holdings, including government, corporate, municipal and all other types of bonds, in addition to securitized instruments. P&C firms held fewer ABS, compared to life firms, but the average exposure was still about 5% in 2007 and 2008.

Second, we note that insurance firms built up their holdings of ABS before the crisis and reduced the exposure afterwards. For life firms, the ABS holdings accounted for 4.6% (median 3.5%) of par value of all bond positions in 2004, increased to 7.4% (median 5.9%) in 2008, and dropped to 5.3% (median 4%) by the end of 2010. P&C firms reduce their exposure earlier and more substantially than life firms. Their relative holdings, measured with par value, were reduced from the maximum of 5.5% (median 3.8%) in 2007 to 3% (median 1.6%) by 2010. We also compute ABS exposure using fair instead of par value. The last four columns show that the fair value of insurance companies' holdings was substantially lower than par value after 2008, suggesting that the ABS holdings were affected more than the other types of bonds held by insurance firms.

Finally, we point out the substantial heterogeneity in ABS exposure across insurance companies. For example, in 2007, the median life firm held only 15 ABS and those in the top percentile held more than 156 such securities. Similarly, the median firm invested 5.9% of the bond portfolio in ABS, and the top percentile held 15.4% in ABS. We will use this heterogeneity to economically identify "gains trading," as those more affected by ABS downgrade and severe price decline have greater incentives to realize gains in other asset classes in order to improve capital positions. There is also significant heterogeneity in the ABS holdings across life and P&C firms. In 2006, the year before the start of the financial crisis, the average life insurance company held 54 different ABS while the average P&C firm held only 11 different ABS.

2.3 Downgrades of ABS Securities and Impact on Insurance Companies

The securitization market expanded substantially before the crisis of 2007-2009. Total ABS issuance grew from $1.5 trillion in 2004 to $2.3 trillion in 2007, according to Asset-Backed Alert. A key development in this market was the collateralized debt obligation (CDO), which, by pooling and tranching, created securities that have much better credit ratings than the collateral assets backing the issues. The better ratings of the ABS attracted investors that face regulatory constraints mechanically tied to credit ratings, such as insurance firms. Following the onset of the financial crisis, ABS were severely downgraded by major rating agencies. In Ratings IQuery, we find 39,464 ABS downgrade actions by S&P in 2008.

We are particularly interested in the downgrades from investment to speculative grades, because these downgrades would force firms to employ higher risk factors thereby triggering larger capital requirements. Figure 1 presents the total number of investment-to-speculative downgrades of ABS on a quarterly basis.

[Insert Figure 1 here]

The massive downgrades started in Q3 of 2007, with 952 downgrades from investment grade to speculative grade. In each of the following four quarters, we observe more than 3,000 such downgrades. In total from Q3 of 2007 to Q4 of 2009, S&P downgraded 34,109 ABS from investment grade to speculative grade.

To gauge the degree to which insurance companies were affected by these downgrades, we count the number of investment to speculative downgrades of ABS that were held by insurance firms. These numbers are also shown in Figure 1 on the right scale. Only a small portion of the downgrades in the early stage of the crisis affected insurers: before the end of 2008, about 5-9% of the downgrades in each quarter affected insurance firms. Insurance companies were, however, more significantly affected by the ABS downgrades in 2009, representing 14-17% of the total number of downgrades of ABS in each quarter of that year. Moreover, we note that the downgrades in the later stage of the crisis were more severe. For example, 325 downgrades in 2008Q4 were straight from AAA to speculative grade, compared to only 94 of such downgrades in all of the preceding quarters. In 2009, a total of 896 downgrades affecting insurance firms were from AAA to speculative grade.

[Insert Table 3 here]

Table 3 shows the rating transitions of severe downgrades, many of which were by several notches, of ABS held by insurance companies. For example, there were 1,254 ABS that were downgraded to a BB rating class (for the sake of brevity, we aggregate all ABS that were downgraded to BB+, BB or BB- in one class) and 455 of these securities were rated as AAA before the downgrade occurred. The same applies to the 1,655 ABS downgraded to the B rating class: 721 ABS were rated as AAA before the downgrade. These dramatic shifts, which likely came as a surprise to insurers, significantly impacted the insurance companies' regulatory capital. In fact, when a bond or bond-like instrument is downgraded from either "class 1" or "Class 2" to "Class 3", which is equivalent to a downgrade from an investment grade rating class to a BB class, the required regulatory capital increases significantly for both life and P&C insurance companies.

2.4 Accounting Treatment of Downgraded ABS Securities

We now explain the rules surrounding the accounting treatment imposed on life and P&C insurance companies when the ABS they held were downgraded. NAIC regulations define 6 different classes by credit ratings, and all fixed income securities held by insurers fall into one of these classes. A particular threshold of importance is between Class 2 and Class 3; the former refers to a security with a BBB rating while the latter refers to a security with a BB rating.

When a fixed income security is downgraded from investment to non-investment grade, P&C insurers have to immediately recognize the value of the bond as the lower between the amortized value (based on HCA) and market price (or model price, in case no market price is available). In the case of life insurers, no such requirement is needed and they can continue holding the downgraded bonds at HCA except in the extreme case when it is classified as 'in or near default' (Class 6). A Class 5 security is one that corresponds to a CCC/Caa credit rating; even in such cases, life insurers can continue holding the security at HCA while P&C insurers have to recognize the market price if the price falls below the amortized value.

In the light of these regulations, the significant ABS downgrades documented in Section 2.3 should generate significant cross-sectional variability in the accounting treatment of downgraded ABS between P&C and life companies. To explore this, we use year-end positions data, which contain the book value and the fair value (the market price or a model price for illiquid assets) reported for each position. We classify as revalued the positions for which the book and the fair values are equal. Others are classified as held at historical cost (i.e. treated under HCA).

Figure 2 reports the percentages of ABS holdings revalued at year-end for both life and P&C over the period 2004-2010.

[Insert Figure 2 here]

The most striking feature is the increase in the percentage of positions that were revalued by P&C companies, far larger than those by life insurers over the same period. P&C companies go from a position where around 5% of their ABS holdings were booked at market values in 2006, similar in magnitude to that of life, to almost 20% of their positions by 2009. Given that the exposures of both life and P&C to downgraded ABS securities were very similar, we attribute the differences in the accounting treatment between life and P&C starting from the end of 2007 to the different accounting regulations imposed on these companies.

An additional question that arises is whether different insurance firms agree on the accounting treatment used to book each ABS that was downgraded (which should be the case if the accounting treatment is determined by the regulation). We address this issue by investigating the revaluations of the ABS positions at the CUSIP level for both life and P&C companies (considering only those ABS that are held by at least two insurance companies within each group). We do not report these results for the sake of brevity; however, the picture that emerges corroborates the evidence in Figure 2.[12]

To better focus on the change in the accounting treatment, we investigate the differences across life and P&C companies in their subsequent accounting treatment of the downgraded ABS. This analysis is shown in Table 4.

[Insert Table 4 here]

Panel A of Table 4 shows the revaluation of ABS that were downgraded from (a) investment grades to speculative grade, and (b) AAA rating to speculative (this being the most severe type of downgrades) over the period 2005-2010. There are striking differences between life and P&C companies. If we consider the most extreme downgrades (from AAA to speculative grade), we see

[12] For example, 91% of all ABS held by at least two life firms are booked at historical cost in 2006, and 92% in 2009. The picture is very different for P&C: in 2006, 88% of all ABS were held at historical cost, but that figure decreases to 65% in 2008, 72% in 2009 and 68% in 2010. Around 4% of ABS were held at market value by all P&C firms in 2006, but this figure rose to 21% in 2008. Finally, we find that there is some disagreement on the same ABS across insurance companies in both groups, but such disagreement is much lower in the life group than P&C group. It is possible that such disagreement arises because of the limited discretion in the hands of each insurance company when determining whether each ABS is to be held until maturity or can be traded and, as a consequence of such decision, the impairment can be judged to be of temporary or permanent nature.

that life insurers had a total of 1,999 different ABS that were downgraded, 1,940 of which were held at book value before the downgrade. Once the downgrade occurred, life insurers kept 78% of those securities at book value and revalued 9% to market values. Compare this to the behavior of P&C that held 970 different ABS that were downgraded, 851 of which were held at book value before the downgrade. Once the downgrade occurred, P&C kept only 45% of those securities at book value, revalued 36% to market value (three times as much as life), and sold 20% of those securities.

One drawback of the NAIC balance sheet data for this particular type of analysis is that the positions are available only at the year-end. It is plausible that revaluations occur at different times within the year. Since we only observe the difference between the book and fair values at year-end, this may lead to some bias against finding revaluations if market prices subsequently drift. This may have happened, for example, during 2009 when many of the extreme downgrades took place relatively early in the year. To address this issue, we consider a subset of downgrades (in Panel B of Table 4) that occurred in the fourth quarter, as these are temporally closer to the year-end measurement we observe and the drift problem may be less important. The results are more striking. Life insurers had a total of 535 different ABS securities that were downgraded (from AAA to speculative grade), 532 of which were held at book value before the downgrade. Once the downgrade occurred, life insurers kept 78% of those securities at book value and revalued 11% to market values. P&C held 243 different ABS that were similarly downgraded, 220 of which were held at book value before the downgrade. Once the downgrade occurred, P&C kept only 16% of those securities at book value, revalued 63% to market values (three times as much as life), and sold 11% of those securities.

Taken together, we conclude that the differences in the regulation governing the accounting treatment between life and P&C draw a clear wedge between the two types of firms. We employ this distinction, and the resulting impact on the way downgraded assets are booked, to explore the incentive for gains trading created by the interplay between accounting and regulatory capital requirements for financial institutions.

2.4 Unrealized Gains and Losses

For each bond position, we calculate the *unrealized gain* as the difference between the position's book and fair values as a percentage of carrying value. As discussed, insurance companies report both the book values and the fair values of all bond positions at year end to the NAIC. Table 5 reports the distribution of the percentage unrealized gains (and losses) separately for life and P&C. Panel A is for ABS, and Panel B is for corporate bonds.

[Insert Table 5 here]

Table 5 Panel A shows that up to the end of 2007, the median unrealized gain for ABS is close to zero, but in 2008 the median unrealized gains for life firms turns into unrealized losses to the tune of -30% with over nine tenths of all ABS positions having unrealized losses and over one tenth having the losses exceeding 75%. These unrealized losses slightly improve in 2009 and 2010, but the overall distribution remains negatively skewed. P&C firms suffer unrealized losses to a much lesser degree, with the median unrealized gain coming back to around zero in 2009 and 2010. This sharp difference between the two groups may be due to the fact that P&C firms are forced to revalue the ABS that are downgraded to speculative grade, essentially truncating the left tail of the unrealized gain distribution. In addition, life firms are likely to avoid selling their ABS in 2008-2009, as doing so would have a significant negative effect on their income and capital.

Panel B of Table 5 shows that the distribution of unrealized gains for corporate bonds also suffers a negative shift in 2008, but quickly returns to normal in 2009. Interestingly, over a quarter of corporate bond positions have unrealized gains in 2008, suggesting that there is more flexibility to potentially realize these gains by trading. In 2009, over half of the corporate bond positions have unrealized gains for life and over three quarters for P&C. As a result, we will use corporate bonds as the main asset class for studying gains trading among insurance companies.

3. Empirical Methodologies and Results

3.1 Difference in Accounting Treatment of ABS

As seen earlier, P&C firms revalue a larger proportion of their ABS positions than do life firms during the crisis. To ensure that this finding is indeed due to regulatory differences, rather than the difference in, say, the credit quality of the ABS held by the two types of insurers, we estimate a logit model for the probability that an ABS position is revalued controlling for credit quality and other distinct characteristics of the ABS:

$$\Pr(M_{i,j,t}=1)=l(\delta_0+\delta_P P_j+\delta_X X_{i,t}+\delta_W W_t) \quad (1)$$

where $l(\cdot)$ denotes the logistic probability function, $M_{i,j,t}$ is an indicator variable that equals one if the insurance company *j* (holding bond *i*) revalues bond *i* by the end of year *t* and zero otherwise, P_j is an indicator variable that equals one if the insurance company *j* is a P&C insurer and zero otherwise, $X_{i,t}$ is a vector of bond *i*'s static characteristics (e.g. issue size) and time-varying characteristics (e.g. remaining maturity) at the end of year *t*, W_t is a vector of time-specific variables, and δ's are the corresponding vectors of coefficients to be estimated. It is important to highlight that in all specifications, we include ratings group[13] ($X_{i,t}$) and U.S. state of incorporation fixed effects (in P_j), as well as either a crisis indicator or year fixed effects (in W_t).

We estimate the model for the crisis (2007-2009)[14] and non-crisis (2004-2006, 2010) periods, both separately as well as together. We are interested in how P&C firms' incremental propensity to revalue their ABS positions differs across the crisis (where many ABS are severely downgraded) and non-crisis periods. Given our use of interaction terms in a number of specifications, it is difficult within a logit specification to interpret parameter estimates for the property and crisis indicators in isolation. We therefore estimate and report the marginal effect of being a P&C firm on the probability of revaluation by making 2,000 repeated draws from the (multivariate normal) distribution of parameter estimates and calculating a simulated sample of probability difference between life and P&C firms (a) in the crisis vs. non-crisis periods and (b) for downgraded ABS vs. others. We then use the simulated sample of marginal effects to construct 90%, 95%, and 99% confidence intervals to determine statistical significance.[15] The results are shown in Table 6.

[Insert Table 6 here]

[13] The ratings fixed effects are equivalent to the NAIC class 1, 2, 3, separately, and classes 4-6 put together.
[14] Note that the logit is estimated for revaluations that can only be observed at a year end frequency.
[15] This methodology is standard for a non-linear model with many indicator variables (see Bratsberg, Raaum, and Roed (2010), for example) and/or interaction terms (see Scheve and Slaughter (1999), for example). See Norton, Wang, and Ai (2004) for a detailed discussion of both the problem and the STATA program they write to address it; however, in most of our settings, we cannot use their program directly.

In column (1) of Table 6, we report the simplest specification in which we examine the behavior of P&C relative to life insurers during the non-crisis period, including ratings group, U.S. state of incorporation, and year fixed effects. In column (2), we have the same specification for the crisis period. We find that while the P&C indicator carries a positive coefficient for both the crisis and the non-crisis periods, it is only statistically significant during the crisis period, with the marginal effect being much larger during the crisis period (1.9% vs. 8.1%). Considering that the average probability of a revaluation of ABS held by all insurance companies during the crisis period is 3.6% (during non-crisis period it amounts to 5.2%), it is very evident that P&C insurers do revalue significantly more than life insurers.

We find similar results in columns (3) and (4) when we consider an alternative specification that includes ABS-level control variables (while still using all fixed effects as before). In columns (5) to (8), we consider the entire sample from 2004 to 2010 together and introduce (a) a crisis indicator and (b) an interaction effect between the P&C indicator and the crisis indicator. In column (5), for example, we show that P&C companies have a higher propensity to revalue assets compared to life companies, and this difference increases during the financial crisis consistent with the time-separated results. The results hold strongly in all specifications, largely unaffected by any control variables we include.

In columns (3), (4), (7), and (8), we also investigate the propensity to revalue downgraded ABS, by adding a downgrade indicator that takes the value of 1 when the ABS rating falls from investment to speculative grades. Recall that by regulation, P&C companies are required to book speculative-grade ABS at the lower between amortized cost and market value while life companies are required to do so only when the ABS are near or in default. Thus, the difference in revaluation probability should come out most strongly among the downgraded ABS. Indeed, we find that the marginal effects of P&C indicator on revaluation are multiple times higher for the downgraded ABS than for others, during both the crisis (21.6%) and non-crisis (59.6%) periods. This striking result, consistent with our finding in Table 4, suggests that the wave of ABS downgrades during 2007-2009 is ideal for investigating the implications of HCA vs. fair value accounting.

Overall these results show clearly that the different regulatory accounting treatments imposed on P&C and life companies have a significant impact on their revaluation behavior. To be clear, the higher propensity of P&C to revalue should be understood more as an outcome of regulations on the accounting treatment rather than voluntary choice. If it is true that accounting is not simply a veil but rather constitutes an important influence on real decisions in markets characterized by frictions, then we should expect to see differences in trading behavior across insurance firms that have different revaluation propensities. Specifically, we expect P&C to behave differently from life when faced with the stark decision of how to react to a high number of ABS downgrades given that these downgrades have a direct impact on their regulatory capital.

3.2 Selling of Downgraded ABS

In this section, we assess whether the P&C firms' revaluation of downgraded ABS to market values (which we have shown truncates the distribution of unrealized losses) makes them more likely to directly sell the downgraded ABS relative to their life counterparts. We model the probability of selling the downgraded ABS *within 3 months after the downgrade having occurred* as a logistic function:

$$\Pr(S_{i,j,k}=1)=l(\kappa_0+\kappa_P P_j+\kappa_V V_{i,j,k}+\kappa_X X_{i,k}+\kappa_W W_k) \qquad (2)$$

where $l(\cdot)$ denotes the logistic probability function, $S_{i,j,k}$ is an indicator variable that equals one if the insurance company *j* (holding downgraded bond *i*) sells the downgraded bond *i* within 3 months after downgrade event *k* and zero otherwise, P_j is an indicator variable that equals one if the insurance company *j* is a P&C insurer and zero otherwise, $V_{i,j,k}$ is an indicator variable that equals one if the insurance company *j* holds the downgraded bond *i* at market value at the year-end before event *k* and zero otherwise, $X_{i,t}$ is a vector of bond *i*'s static characteristics and time-varying characteristics just before event *k*, W_k is a vector of time-specific variables for each event *k*, and κ's are the corresponding vectors of coefficients to be estimated. In all specifications, we include ratings group, U.S. state of incorporation and year fixed effects.

We estimate the model separately for (a) all downgraded ABS, (b) only the ABS that were downgraded from investment to speculative grade, and (c) only the ABS that were downgraded from AAA to speculative grade. The marginal effects of the P&C and revaluation indicators are calculated as previously described. The results are shown in Table 7.

[Insert Table 7 here]

From column (1), we find that P&C companies have a higher propensity to sell the downgraded ABS (rather than keeping these ABS on their book) relative to life. This result is confirmed using the specification that includes ABS-level control variables (remaining maturity and the log of the issue size) and a revaluation indicator variable that captures the insurance company's decision to revalue the downgraded ABS. The revaluation indicator variable is used to control for the likelihood that insurance companies are more likely to sell positions that have been re-booked at market price because once they do so, they should be largely indifferent between keeping the asset on their balance sheet or selling it. In fact, the marginal effect of the revaluation indicator is positive and statistically significant in all specifications, showing that revalued positions are indeed more likely to be sold. It should be noted that these trading dynamics cannot be explained by any differences in regulations across different U.S. states since we include state of incorporation fixed effects. Further, these effects are not driven by ABS-level characteristics, such as liquidity, since we include ABS-level controls.[16]

The main result that P&C firms are more likely to sell downgraded ABS than life firms is confirmed when we investigate (a) only the ABS that were downgraded from investment to speculative grades (columns (3) and (4)) and (b) only the ABS that were downgraded from AAA to speculative grade (columns (5) and (6)). We want to highlight that for the last set (most severe downgrades), the effects of the P&C indicator are about twice as large as those for the other types of downgrades. For these downgrades (AAA to speculative grade), and controlling for the revaluation effects, we find that the selling probability of P&C firms exceeds the selling probability of life firms by 2.5%. Considering that the average selling probability is 8.3% for the ABS downgraded from AAA to speculative grade, we can say that, P&C firms have a 72 percentage points higher propensity to sell than life firms. The same selling behavior is observed when we consider less severe downgrades from investment to speculative (35 percentage points) and all downgrades (34 percentage points).

These results, obtained at the individual ABS-level, are a confirmation of the broad industry trend shown in Table 2. Recall that the average holdings of ABS of life insurers was

[16] In fact, the marginal effects of issue size show that large-issue ABS are more likely to be sold, possibly due to their superior liquidity.

6.48% in 2006 and was reduced to 4.74% in 2009, while the average holdings of ABS of P&C went from 5.08% to 2.75% over the same period.

In sum, the results from Table 7 exhibit a sharp contrast between P&C and life companies: P&C firms disproportionately sell their downgraded ABS holdings. This may very well be a consequence of the regulatory accounting treatment we document in Section 3.1. Since P&C companies are forced to book the losses that result from severe downgrades, they would be indifferent between holding the asset at the lower value and selling it. More importantly, these trading dynamics may be the result of the interactions between those accounting rules and the regulatory capital requirements. For P&C insurers, selling the asset has an important advantage from the regulatory capital point of view, as they are exchanging a risky asset for cash thereby reducing their capital requirements.

3.3 Propensity to Gains Trade

In this section, we assess insurance companies' propensity to gains trade, defined as selectively selling the positions in the book that have high unrealized gains, when they hold these positions at book value. So far we have established that P&C insurers are more likely to revalue the downgraded ABS to market prices and have a higher propensity to sell them relative to life insurers. Given that both P&C and life have roughly similar exposures to downgraded ABS securities, and thus a similar impact on their regulatory capital, this begs the question as to the actions taken by life insurers in response to this hit on their regulatory capital.

In continuing to hold their downgraded ABS positions, life companies may have the advantage of limiting an unfavorable price impact, but *additional capital may still be required*. It is precisely here that gains trading becomes important. As a life firm attempts to sell its existing assets that have not been downgraded to shore up its capital, it has an incentive to do so by *selectively selling* those assets that have the *largest unrealized gains*. Only by selling these assets can these large unrealized gains be recognized and be applied to its capital.

Life companies engaging in gains trading have different asset classes from which to choose. It would be natural to consider government bonds which feature heavily in insurance companies' portfolio. However, selling government bonds does little to assist from a regulatory capital point of view because they are free from capital requirement. Insurance companies also hold equities, but we do not consider these positions since equities are held at market value. The other natural positions to consider are (mostly investment-grade) corporate bonds[17], a significant number of which *are* carried at unrealized gains, as shown in Table 5. An advantage of selling corporate bonds is that the insurance company will exchange a risky asset for cash.

We model the probability of selling each corporate bond position as a logistic function:

$$\Pr(S_{i,j,q}=1)=l(\gamma_0+\gamma_Z Z_{i,j,q}+\gamma_X X_{i,q}+\gamma_Y Y_{j,q}+\gamma_W W_q) \quad (3)$$

where $l(\cdot)$ denotes the logistic probability function, $S_{i,j,q}$ is an indicator variable that equals one if the insurance company *j* (holding bond *i*) sell bond *i* in calendar quarter *q* and zero otherwise, $Z_{i,j,q}$ is the percentile (ranging from 0 to 1) of unrealized gain of corporate bond *i* in the portfolio of insurance company *j* at the year-end prior to quarter *q*, $X_{i,q}$ is a vector of bond *i*'s static characteristics and time-varying characteristics at the beginning of quarter *q*, $Y_{j,q}$ is a vector of financial and risk characteristics of insurance company *j* at the year-end prior to quarter *q*, W_q is a vector of time-specific variables for quarter *q*, and γ's are the corresponding vectors of coefficients to be estimated. The results are shown in Table 8.

[Insert Table 8 here]

We start by comparing the results during the crisis period for life insurers, shown in columns (1) and (2), and compare them with those for P&C insurers, shown in columns (7) and (8). Considering the first row, *assuming that the insurance firms do not have high ABS exposure and low RBC ratio* (hence all interaction terms are zero), we find that life companies have a positive and statistically significant propensity to sell corporate bonds at higher levels of unrealized gains. Interestingly, the same coefficient is negative for P&C firms. When we compare the propensity of the life and P&C insurers during the non-crisis periods (results shown in columns (3) and (4) for life and columns (9) and (10) for P&C), we also find a significant difference between the two groups. In normal times, life insurers are actually less likely to sell

[17] On balance, there are various reasons to believe that corporate bonds will be preferred to equity for gains trading. First, insurance companies are significant investors in corporate bonds. Schultz (2001) and Ellul *et al.* (2011) estimate that insurance companies collectively hold between one-third and forty percent of investment-grade corporate bonds. At a weight of around 8%, equities represent a much smaller segment of insurance companies' portfolios. Finally, and most importantly, equities are held at market value, and thus gains trading is not possible.

bonds with high unrealized gains (possibly to avoid tax)[18] while P&C are found to be insensitive to any type of gains trading during the non-crisis period. Overall, life insurers are found to engage in gains trading *only* during the crisis period, while they tend to sell their corporate bonds with the *lowest unrealized gains* during the non-crisis period. No such behavior is detected for P&C companies.

We next investigate the impact of each insurance company's exposure to ABS assets *held at book value* on the propensity to engage in gains trading. We have argued that this is precisely a factor that may lead insurance companies to gains trade. We find that while life insurers with high ABS exposure (held at book values) sell more frequently corporate bonds (possibly just turning over their portfolios more often) in general (second row), they do not appear more likely to gains trade during the crisis (third row of columns (1) and (2)). These life firms tend to sell corporate bonds with the lowest unrealized gains during the non-crisis period (third row of columns (3) and (4)). These life firms thus change their behavior from the non-crisis period (selling of corporate bonds with lowest unrealized gains) to the crisis period. Interestingly, the results for P&C also confirm that gains trading is associated with historical cost accounting. Recall that the variable "High ABS exposure dummy" refers to ABS held at book value. Thus, even in P&C companies, high exposure to ABS held at book value should still induce gains trading.

The other important dimension to consider in the decision to gains trade is the regulatory capital pressure that each insurance company faces at times of severe downgrades. We explore this additional dimension by investigating the impact of the firm-level risk-based capital (RBC) ratio. Recall from Section 2 that the RBC ratio is the ratio of total adjusted capital to NAIC risk-based capital. It is important to note that the insurance literature views RBC ratios as *indicative* of financial health rather than categorical (e.g. above or below two is not a sole criterion for regulatory scrutiny) and the higher is the ratio the lower are the regulatory constraints.

To fully understand the dynamics of the interactions between accounting treatment and regulatory capital, we use three different variables: (a) an indicator variable to capture insurance companies with low RBC, defined as the RBC of insurers in the bottom quartile of the RBC ratio distribution, (b) an interaction term between the low RBC ratio indicator and the unrealized gains percentile, and (c) a triple interaction term between the low RBC ratio indicator, the unrealized

[18] See Jin (2006).

gains percentile and the high ABS exposure held at book values indicator variable. We find evidence indicative of the impact of the interaction between accounting treatment and capital regulations. While life insurers with low RBC ratio and high ABS exposure held at book value are more likely to engage in gains trading during the crisis, we find no such behavior for P&C companies.

So far, we have investigated the trading behavior of insurance firms during crisis and non-crisis periods, separately. We also estimate the propensity to engage in gains trading for the entire sample for life insurers, in columns (5) and (6), and for P&C insurers, in columns (11) and (12). In such specifications we introduce a crisis indicator variable and interact this indicator with other variables of interest. This pooled specification broadly confirms the evidence in the time-separate estimations. First, life insurers engage in gains trading during the crisis period, whereas the coefficient estimate for P&C is found to be negative. Second, the quadruple interaction term (in row 14) where we interact the crisis indicator with the low RBC ratio indicator, the unrealized gains percentile, and the high ABS exposure held at book values indicator confirms that life companies facing regulatory constraints and holding significant ABS positions at book values disproportionately sell corporate bonds with high unrealized gains during the crisis. No such action is observed for P&C companies.

Given the non-linear nature of the logistic function and our heavy use of interaction terms, a more appropriate way to understand the sign, magnitude, and economic significance of these results, is to investigate the marginal effects (estimated via simulation). Panel B of Table 8 provides the estimates of the marginal effects using the models for the entire sample period (shown in columns (5) and (6) for life and columns (11) and (12) for P&C). We estimate the marginal effects on gains trading considering two cases for each group: (a) insurers with low ABS exposure held at book values and high RBC ratios (denoted as "High ABS exposure dummy = 0 and Low RBC ratio dummy = 0" in Panel B), and (b) insurers with high ABS exposure held at book values and low RBC ratios (denoted as "High ABS exposure dummy = 1 and Low RBC ratio dummy = 1" in Panel B). The average selling probability of a corporate bond held by life companies over the entire period is 4.4%. Consider two similar corporate bonds held by a life firm with high ABS exposure and low RBC ratio, one with unrealized gain at the 25th percentile of the firm's portfolio and the other with unrealized gain at the 75th percentile. Using the estimates from model (5), the second bond is significantly less likely than the first bond to be sold during the non-crisis period (by 1.1%) but is significantly more likely during the crisis (by 1.3%). The difference is equivalent to an increase of the probability of selling by 2.4%, almost 55 percentage points of the average selling probability for the entire period. Similar results are obtained when we estimate the marginal effects using the model shown in columns (6) in Panel A. In sharp contrast, all the marginal effects for corporate bonds held by P&C companies are statistically insignificant, confirming the indicative results found in Panel A. While we find that life companies with low exposure to ABS held at book value and high RBC ratio also engage in gains trading during the crisis, the effect is much smaller than that found among the group of life companies with high ABS exposure and low RBC ratio. This result is consistent with the importance of the interaction between regulatory capital and accounting treatments.

It is also important to note that the probability of gains trading of life insurers is robust to the inclusion of a host of control variables that may be associated with selling for unrelated reasons. The first notable variable is liquidity. The corporate bond literature has found that bid-ask spreads increase with bond age and decrease with bond issue size (see Edwards, Harris and Piwowar (2007)).[19] In all specifications, we include the log of the corporate bond age and the log of the corporate bond size issue and thus liquidity considerations should not drive our results. However, it is also important to note that we find that insurance companies are more likely to sell younger bonds and bonds with larger issue size. One interpretation of these results is that life insurers actively try to minimize any negative price impact by avoiding the sale of the most illiquid corporate bonds.

Other significant control variables include the proportion of risky assets in an insurer's portfolio (in order to capture the insurer's risk appetite or its capacity to bear risk), an indicator that measures whether the downgrade is into the speculative class, and an indicator that captures whether the bond issuer filed for bankruptcy during the quarter. We find no evidence that insurance companies with higher risk appetite or higher capacity to bear risk, as proxied by their risky asset holdings, are more likely to engage in the selling of corporate bonds. Bonds that are downgraded to the speculative grade are more likely to be sold. This may be due to the differential degrees of negative information across rating classes or the regulation that imposes

[19] See also Hong and Warga (2000) and Schultz (2001). Driessen (2005) uses bond age to identify the liquidity component of credit spreads.

much higher capital charge for holding speculative-grade bonds. Finally, as expected, bonds of the issuers that went into bankruptcy are more likely to be sold.

The above results are obtained after the inclusion of year-quarter dummies, U.S. state (of incorporation) dummies and rating group dummies. Thus, our results cannot be driven by market-wide conditions that occur during the time of downgrades, other regulations faced by insurance companies, which differ across U.S. states, or ratings of the bonds.

3.4 Price Impact of Gains Trade

The final question we address is whether the selective selling in Section 3.3 creates enough pressure in the targeted bonds to generate price distortions. In the case that such price pressures do occur, one can conclude that HCA – precisely because of its interactions with capital regulations – does not avoid spillovers and fire sales, as claimed by the existing theoretical literature.

We evaluate the price impact in two ways. First, for each sale transaction, we compare the sale price with the market median obtained from TRACE. Like any other investors, insurance companies pay transaction costs when they sell a bond to a dealer (as they demand liquidity); their sale price is, on average, lower than the market value. Since only bond positions with sufficiently large unrealized gain can be used for gains trading, insurers may have to concede even more to sell this limited number of positions. Second, at the bond level, we compare the quarterly return across bonds that are subject to different degrees of gains trading. If insurance companies demand liquidity when they sell bonds to realize gain, elevated gains trading should put more pressure on the overall market price. We thus expect the bonds most targeted for gains trading to underperform otherwise similar bonds.

At the transaction level, we measure the price impact of each transaction as the percentage discount of the sale price relative to the median market price during the week, $\ln(P_{i,j,w}/\bar{P}_{i,w})$, where $P_{i,j,w}$ is the transaction price of bond *i* sold by insurance company *j* in week *w* and $\bar{P}_{i,w}$ is the median market price of bond *i* in week *w*. To ensure the reliability of our median price, we only use the bond-weeks in which there are at least 3 transactions in the same bond. We model the price impact as

$$\ln(P_{i,j,w} / \overline{P}_{i,w}) = \beta_0 + \beta_Z Z_{i,j,w} + \beta_X X_{i,w} + \beta_Y Y_{j,w} + \beta_W W_w + \varepsilon_{i,j,w} \tag{4}$$

where $Z_{i,j,w}$ is a vector of characteristics of a sale transaction of bond *i* by insurance company *j* in week *w*, $X_{i,w}$ is a vector of bond *i*'s static characteristics and time-varying characteristics at the beginning of week *w*, $Y_{j,w}$ is a vector of financial and risk characteristics of insurance company *j* at the year-end prior to week *w* (including the dummies for the company's domicile state), W_w is a vector of time-specific variables for week *w*, and β's are the corresponding vectors of coefficients to be estimated. Wherever the specification allows us, we include ratings group, U.S. state of incorporation, and year fixed effects.

Our interested variable is the marginal effect of unrealized gain percentile (relative to mean) on the selling probability, which is part of the vector $Z_{i,j,w}$. We measure this marginal effect using the model similar to the models shown in columns (2) in Table 8, estimated cross-sectionally for each calendar quarter. This marginal effect is specific to each bond position so that even the same bond may have different marginal effects depending on the insurance company holding the bond, the price at which the bond is acquired, and the period in which we measure the selling probability. We estimate the above model of percentage price discount by OLS, and cluster the standard errors by bond issuer. The results are shown in Table 9.

[Insert Table 9 here]

Table 9 columns (1) and (4) show that for both life and P&C insurers, the price impact of selling is significantly higher during the crisis. Columns (2), (3), (5), and (6) analyze whether this increased price impact is related to gains trading. To do so, we include the interaction variable between the crisis indicator and the gains trading selling pressure. We find that the coefficient estimates are negative and statistically significant for life companies but are not significant for P&C. This result clearly shows that the gains trading engaged by life companies is potentially expensive. As we move from the 25^{th} to 75^{th} percentiles of the gains-trading selling pressure for life companies during the crisis, the price impact increases from 62 basis points (specification in column (3)) to 77 basis points (specification in column (2)). This price impact is sizable, considering that the mean one-way transaction costs are about 15 basis points in normal markets (see Edwards *et al*. (2007)). It is important to stress that our estimate of price impact is obtained after controlling for various bond-level, insurance-level, market-wide, and transaction-specific effects.

If a large number of insurance companies attempt to gains trade using the same corporate bonds in an illiquid market, then we should expect the bonds to suffer significant price pressures. To investigate whether this is the case, we move from the transaction level to the bond level. Using the transaction prices from TRACE, we calculate the quarterly return of a bond as the logged change in price from the last day of the previous quarter to the last day of the current quarter. This return measure is far from perfect. First, corporate bonds do not trade every day so the last day on which we observe trades for each bond may often be a few days before quarter-end. We however find that for the bonds in our data that pass our screen, approximately 90% of the last trading days that we use fall in the last month of the quarter. Second, the holding period over which we measure the bond return may be greater than one or two quarters (if the bond is last traded a few quarters back). This problem affects less than 5% of the observations. We address the first two problems, which result in irregular holding periods, by measuring the values of (some) control variables over the same period in which the bond return is measured. Finally, a bond may trade a few times in a day and the trade prices can be very different.[20] Since corporate bond prices are more accurate for larger trades, we use the size-weighted average of all trade prices on the last day of a quarter.

Our model of quarterly bond return is as follows:

$$R_{i,q} = \beta_0 + \beta_X X_{i,q} + \beta_W W_q + \xi_{i,q} \tag{5}$$

Where $R_{i,q}$ is the return of bond *i* in quarter *q*, $X_{i,q}$ is a vector of bond static and time-varying characteristics at the beginning of quarter *q*, W_q is a vector of time-specific variables for quarter *q*, and β's are the corresponding vectors of coefficients to be estimated.

Our variable of interest is the selling pressure associated with unrealized gains, which we measure in two broad ways. First, we use the bond-level averaged marginal effects of unrealized

[20] See Bao, Pan, and Wang (2011), Feldhutter (2011), and Jotikasthira (2008), for example.

gain on selling probability. The marginal effect for each bond position is calculated as the predicted value (relative to mean) obtained from the logit model similar to column (2) in Table 8, estimated separately for each calendar quarter. For each bond in each quarter, we then take (value-weighted or equally-weighted) averages of the marginal effects across the positions of all insurance companies in the particular bond. Second, we also consider the percentage unrealized gains, averaged across all positions in the bond. Although this measure does not reflect the fact that the same unrealized gains may have different effects on selling propensity for different insurance companies, it is not affected by possible misspecifications of our logit model. These selling pressures from gains trading enter the model as part of the vector $X_{i,q}$.

We distinguish the effect of gains trading from other effects of unrealized gains by interacting our measures of selling pressure with the crisis indicator. The selling pressure from gains trading should only operate during the crisis where insurance companies are hit by ABS downgrades and the huge unrealized losses of ABS positions in their portfolios. We only include the pressure from life firms, as we have shown in Table 8 that P&C firms do not appear to significantly gains trade.

We estimate the above model of quarterly bond return by OLS, and cluster the standard errors by bond issuer. We include standard control variables for the fundamental movement in the bond price, using maturity-matched Treasury and ratings- and maturity-matched credit spread returns. We use the interpolated constant maturity Treasury bond/note from the Fed to calculate the Treasury return. The spread return is the corporate bond index return minus the Treasury return, where we use the ratings- and maturity-matched Bank of America-Merrill Lynch bond index as our primary source. Since the maturity group for the index is broad, we also adjust for the duration difference between the index and the bond of interest to ensure accuracy. Finally, we control for bond-specific characteristics, particularly bond age and issue size, which are key determinants of bond liquidity. Ratings group and calendar quarter dummies are also included. The results are shown in Table 10.

[Insert Table 10 here]

We find clear evidence that the corporate bonds disproportionately targeted by life insurance companies for gains trading statistically and economically underperform otherwise similar bonds. The coefficient estimate of the interaction variable of interest (the crisis indicator interacted with gains-trading selling pressure) is always negative whether we use an equal-weighted or a value-weighted methodology and whether we use marginal effects of unrealized gain on selling probability (columns (1) to (4)) or the percentage unrealized gains (columns (5) to (8)). The quarterly return is 0.7-1.7% lower as we move from the 25th to 75th percentiles of aggregated gains trading. Given that the inter-quartile range of quarterly bond abnormal return during the crisis is approximately 4.90%, the magnitude of the selling pressure has economic significance. It is important to put this result in perspective: the selling pressure is generated by gains trading engaged in by life companies. This originates, in part, from their exposure to ABS held at book value. We are documenting significant price declines in the corporate bond market, demonstrating spillover effects from downgraded ABS to otherwise unrelated corporate bonds through the interaction between the accounting treatment and regulatory capital requirements.

4. Conclusions

While the theoretical literature has argued that historical cost accounting may insulate financial institutions from the price distortions associated with market stress, we provide new empirical evidence supporting the view that historical cost accounting, along with regulatory capital requirements, induces an altered incentive to "gains trade" – that is, to selectively sell otherwise unrelated assets with high *unrealized* gains. Given (a) the distinction in regulatory accounting treatment across life and property and casualty insurance companies and (b) the availability of security-level data on all positions held, traded, and booked by insurers, the insurance industry presents an interesting laboratory to explore the interplay between accounting and regulatory capital requirements for financial institutions.

In contrast to property and casualty insurers, life insurers have a greater degree of regulatory flexibility to hold downgraded instruments at historical cost. When faced with severe downgrades among their holdings in asset-backed securities (ABS) during the financial crisis, life insurers, particularly those facing both regulatory capital constraints and high ABS exposures, largely continue to hold the downgraded securities at historical cost and instead selectively sell their unrelated corporate bond holdings with the highest unrealized gains. This behavior is largely absent among property and casualty insurers. As the observed gains trading by life insurers induces significant price declines for the corporate bonds that happen to exhibit high unrealized gains, we conclude that historical cost accounting does not completely avoid illiquidity spillovers.

References

Allen F. and E. Carletti, 2008, Mark-to-market accounting and cash-in-the-market pricing, *Journal of Accounting and Economics* 45, 358-378.

Anand, A., P. J. Irvine, A. Puckett, and K. Venkataraman, 2010, Market crashes and institutional trading, working paper, University of Georgia.

Bao, J., J. Pan, and J. Wang, 2011, The illiquidity of corporate bonds, *Journal of Finance*, forthcoming.

Beatty, A., S. L. Chamberlain, and J. Magliolo. 1995. Managing financial reports of commercial banks: The influence of taxes, regulatory capital earnings, *Journal of Accounting Research* 33, 231-261.

Boyson, N., J. Helwege, and J. Jindra, 2011, Crises, liquidity shocks, and fire sales at financial institutions, working paper, University of South Carolina.

Bratsberg, B., O. Raaum and K. Røed, 2010, When minority labor migrants meet the welfare state, *Journal of Labor Economics* 28, 633-676.

Driessen, J., 2005, Is default event risk priced in corporate bonds? *Review of Financial Studies* 18, 165-195.

Duffie, D., 2010, Presidential address: Asset price dynamics with slow-moving capital, *Journal of Finance* 65, 1237-1267.

Eccher, E. A., K. Ramesh, and S. R. Thiagarajan, 1996, Fair value disclosures by bank holding companies, *Journal of Accounting and Economics* 22, 79-117.

Edwards, A., L. Harris, and M. Piwowar, 2007, Corporate bond market transaction costs and transparency, *Journal of Finance* 63, 1421-1451.

Ellul, A., C. Jotikasthira, and C. Lundblad, 2011, Regulatory pressure and fire sales in the corporate bond market, *Journal of Financial Economics*, 2011, 101, 596-620.

Feldhutter, P., 2011, The same bond at different prices: Identifying search frictions and selling pressures, *Review of Financial Studies*, forthcoming.

Goh, B.W., Ng, J., and Yong, K.O., 2009, Market pricing of banks' fair value assets reported under SFAS 157 during the 2008 economic crisis, working paper, Singapore Management University.

Heaton, J., D. Lucas, and R. McDonald, 2010, Is mark-to-market accounting destabilizing? Analysis and implications for policy, *Journal of Monetary Economics* 57, 64-75.

Hirst, D., and P. Hopkins, 1998, Comprehensive income reporting and analysts' valuation judgments, *Journal of Accounting Research* 36 (Supplement), 47–75.

Hirst, D., P. Hopkins, and J. Wahlen, 2004, Fair values, income measurement, and bank analysts' risk and valuation judgments, *The Accounting Review* 79, 453–472.

Hong, G., and A. Warga, 2000, An empirical study of bond market transactions, *Financial Analysts Journal* 56, 32-46.

Jin, L., 2006, Capital gains tax overhang and price pressure, *Journal of Finance* 61, 1399-1431.

Jotikasthira, C., 2008, Does search friction really matter? Evidence from the corporate bond market, Working paper, University of North Carolina at Chapel Hill.

Kashyap, A. K. and J. C. Stein, 2000, What do a million observations on banks say about the transmission of monetary policy? *American Economic Review* 90, 407-428.

Laux, C., and C. Leuz, 2009, The crisis of fair-value accounting: Making sense of the recent debate, *Accounting, Organizations and Society* 34, 826–834.

Laux, C., and C. Leuz, 2010, Did fair-value accounting contribute to the financial crisis? *Journal of Economic Perspectives* 24, 93–118.

Lee, Y-J, K.R. Petroni, and M. Shen (2006), Cherry picking, disclosure quality, and comprehensive income reporting choices: The case of property-liability insurers, *Contemporary Accounting Research* 23, 655–692.

Manconi, A, M. Massa, and A. Yasuda, 2010, The behavior of intoxicated investors: The role of institutional investors in propagating the crisis of 2007-2008, *Journal of Financial Economics*, forthcoming.

Norton, E., H, Wang, and C. Ai, 2004, Computing interaction effects and standard errors in logit and probit models, *The STATA Journal* 4, 154-167.

Penman, S. H., 2007, Financial reporting quality: Is fair value a plus or a minus? *Accounting and Business Research* Special Issue: International Accounting Policy Forum, 33-44.

Petroni, K.R., and J.M. Wahlen, 1995, Fair values of equity and debt securities and share prices of property-liability insurers, *Journal of Risk and Insurance,* 62, 719-737.

Plantin, G., H. Sapra, and H. Shin, 2008, Marking-to-market: Panacea or Pandora's box? *Journal of Accounting Research* 46, 435-460.

Sapra, H., 2008, Do accounting measurement regimes matter? A discussion of mark-to-market accounting and liquidity pricing, *Journal of Accounting and Economics* 45, 379-387.

Scheve, K. F., and M. J. Slaughter, 1999, What determines individual trade-policy preferences? NBER working paper.

Schultz, P., 2001, Corporate bond trading costs: A peek behind the curtain, *Journal of Finance* 56, 677-698.

SEC, 2011, 2011 Summary Report of Commission Staff's examinations of each nationally recognized statistical rating organization, U.S. Securities and Exchange Commission, September 2011.

Veron, N., 2008, Fair value accounting is the wrong scapegoat for this crisis, *European Accounting Review* 5, 63-69.

Wallison, P.J., 2008, Fair value accounting: A critique, Financial Services Outlook, American Enterprise Institute for Public Policy Research, July 2008.

Wyatt, A. (1991). The SEC says: mark to market! *Accounting Horizons* 5, 80-84.

Appendix A: Descriptions of Variables

Variable	Specific to	Definition
% risky assets	Insurer-year	Percentage of investment assets invested in any of the following asset classes: non-investment grade bonds, common and preferred stocks, non-performing mortgages, real estate, and other investments. According to the Street.com and NAIC, the target capital percentages for these assets are greater than or equal to those of the least risky class of non-investment grade bonds (BB).
ABS exposure	Insurer-year	Percentage of bond portfolio invested in asset-backed securities (ABS) held at book value, measured in market value terms. Securities (9-digit CUSIPs) are classified as ABS, based on the list of rated securities from S&P's RatingIquery. High ABS exposure dummy equals 1 for ABS exposures above the annual median, and 0 otherwise.
Bankruptcy dummy	Bond-quarter	Dummy variable equal to 1 if the issuer of the bond files for bankruptcy during the quarter, and 0 otherwise.
Bond age	Bond-quarter	Time from issuance to the beginning of quarter of interest or the beginning of quarter in which the interested transactions fall (depending on specifications), measured in years.
Bond return	Bond-quarter	Log of change in prices from the last day when there are any trades of a bond in the previous quarter to the last day in the current quarter, scaled by a factor of 100. Abnormal bond return is calculated as the residual of pooled OLS regression of bond return on maturity matched treasury return and maturity- and rating-matched corporate bond index return over the same quarter. Corporate bond index return is calculated using Bank of America-Merrill Lynch bond index, adjusted for duration difference between the index and the bond of interest.
Calendar quarter fixed effects	Quarter	Set of dummy variables for calendar quarters in which the observations fall.
Capital and surplus	Insurer-year	The insurance company's statutory net worth (including paid-in capital or unimpaired surplus and additional funds in surplus) in millions of dollars through the most recent year end.
Crisis dummy	Quarter	Dummy variable equal to 1 if the calendar quarters are in the 2007-2009 crisis period, and 0 otherwise. The crisis period is defined based on the volume of ABS downgrades, and covers 2007Q3 to 2009Q4.
Downgrade dummy	Bond-quarter	Dummy variable equal to 1 if the bond is downgraded from investment to non-investment grades during the quarter, and 0 otherwise. S&P ratings are used wherever available. Moody's ratings are used when S&P ratings are unavailable.
Incremental effect of unrealized gain on selling probability	Position-quarter or Bond-quarter	Position-quarter incremental effect of unrealized gain on selling probability is predicted value for each bond position (relative to mean) obtained from the logit model similar to model (2) in Table 8, estimated separately for each calendar quarter. Bond-quarter incremental effect of unrealized gain on selling probability is value-weighted or equally-weighted average of position-quarter incremental effects, where the average is taken across all insurers holding the bond at the beginning of the quarter.
Issue size	Bond	Offering amount of the bond, measured in million dollars.
Leverage	Insurer-year	Debt as percentage of total assets, all measured at book values.
Liquidity index *,**	Insurer-year	The Street.com's index that measures the insurance company's ability to raise cash to settle claims. The inability to raise cash may arise when the company is owed a great deal of money from its agents or reinsurers, or it cannot sell its investments at the prices at which the investments are valued in the company's financial statements. Low liquidity index dummy equals 1 for liquidity index values below 5, and 0 otherwise.
Maturity	Bond-quarter	Maturity of the bond at the beginning of quarter of interest or the beginning of quarter in which the interested transactions fall (depending on specifications), measured in years.
NAIC risk-based	Insurer-year	Ratio of total adjusted capital (capital, surplus, and applicable valuation reserves)

Variable	Specific to	Definition
capital ratio (RBC Ratio)		to NAIC risk-based capital (RBC). RBC is the *minimum* amount of capital that the insurance company must maintain based on the inherent risks in its operations. RBC is calculated based on the NAIC's formula which reflects its assessment of risks of different asset classes and businesses. For example, a company with RBC ratio of 1.0 has capital equal to its RBC. Insurance companies with higher RBC ratios are considered better capitalized. Insurance companies with RBC ratio below 2.0 are subject to supervisory interventions. The levels of supervisory actions depend on the level of RBC ratio. Low RBC ratio dummy equals 1 for RBC ratios below the annual median, and 0 otherwise.
Rating group fixed effects	Bond-quarter	Set of dummy variables for credit rating groups, defined by the NAIC's capital requirement in the RBC ratio formula. The groups are, in order of credit quality, A and above, BBB, BB, and B and below. S&P ratings are used wherever available. Moody's ratings are used when S&P ratings are unavailable.
Revalue dummy	Position-year	Dummy variable equal to 1 if the position has the book value that is equal to its reported fair or market value, and 0 otherwise.
ROE	Insurer-year	Return on equity, measured as net income divided by book value of equity at the beginning of the year. Positive ROE dummy equals 1 if ROE is greater than zero, and 0 otherwise.
Sell dummy (dependent variable in logit)	Position-quarter	Dummy variable equal to 1 if part or all of the position is sold during the quarter, and 0 otherwise.
Selling probability	Position-quarter or Bond-quarter	Position-quarter selling probability is predicted selling probability for each bond position obtained from the logit model similar to model (2) in Table 8, estimated separately for each calendar quarter. Bond-quarter selling probability is value-weighted or equally-weighted average of position-quarter selling probabilities, where the average is taken across all insurers holding the bond at the beginning of the quarter.
State fixed effects	Insurer-year	Set of dummy variables for insurers' domicile states.
Unrealized gain (and loss)	Position-year	Difference between insurer's reported fair value and book-adjusted carrying value of the position at previous year end, measured as percentage of book value. Unrealized gain percentile is the percentile rank, ranging from 0 to 1, of the position's dollar unrealized gain within the insurer's portfolio at previous year end.

* The Street.com may not evaluate some insurance companies for one or more of the following reasons: (i) total assets are less than $1 million, (ii) premium income for the current year is less than $100,000, (iii) the company functions almost exclusively as a holding company rather than as an underwriter, or (iv) the Street.com does not have enough information to reliably evaluate the company.

** Scores range from 1 to 10. Scores of 7 to 10, 5 to 6.9, 3 to 4.9, and 2.9 and below are considered "strong", "good", "fair", and "weak", respectively.

Table 1: Summary Statistics of Insurance Companies

This table presents descriptive characteristics of all insurance firms in our sample at the end of 2007. Our sample restrict to insurance companies with invested assets not less than $13 million and RBC ratio between 2 and 20. We also exclude 33 bond insurers including AIG, AMBAC, MBIA, etc. Definition of the variables can be found in the Appendix A.

Panel A: Life Firms

	Mean	10thPct	50thPct	90thPct	Std.Dev.
Number of Firms	538				
Invested Assets ($ million)	4,738	27	393	9,414	16,600
Capital and Surplus ($ million)	476	8	66	1,070	1,329
Leverage	0.80	0.52	0.87	0.96	0.18
Return on Equity (ROE)	0.10	-0.08	0.09	0.31	0.29
NAIC Risk-Based Capital Ratio (RBC ratio)	8.81	4.48	8.15	14.37	3.82
Holding of Investment-Grade Bonds (%)	57.22	32.70	57.35	84.00	20.66
Holding of Risky Assets (%)	15.54	1.59	11.36	32.17	16.12
Profitability Index	5.78	1.90	6.30	8.60	2.42
Liquidity Index	6.57	4.90	6.70	8.80	1.73

Panel B: Property and Casualty Firms

	Mean	10thPct	50thPct	90thPct	Std.Dev.
Number of Firms	1,344				
Invested Assets ($ million)	829	22	119	1,272	3,996
Capital and Surplus ($ million)	384	11	53	565	2,172
Leverage	0.60	0.42	0.62	0.75	0.14
Return on Equity (ROE)	0.12	0.01	0.11	0.25	0.13
Risk-Adjusted Capital Ratio 1 (RACR1)	2.34	0.72	1.86	3.92	4.85
Holding of Investment-Grade Bonds (%)	63.16	35.72	64.19	91.13	21.45
Holding of Risky Assets (%)	17.51	0.00	12.05	41.86	19.40
Profitability Index	6.31	3.30	6.50	8.80	2.07
Liquidity Index	6.76	5.70	6.80	8.30	1.40

Table 2: Summary Statistics of Insurance Companies' Holding of ABS Securities

This table summarizes the holding of ABS securities of insurance companies at year-end. We identify ABS positions by matching insurance firms' bond holding positions at year-end to a list of ABS securities identified from S&P's Ratings IQuery using 9-digit CUSIP. S&P's Ratings IQuery comprehensively covers initial ratings and histories for all securitized issues rated by S&P from 1991 to 2010. The statistics on the number and size of ABS holdings are reported only for firms investing in ABS. The size of the ABS holdings is the par (or fair) value of the identified ABS securities held by a firm relative to the par (or fair) value of all fixed income positions in this firm. We report the mean, median, top, and bottom percentile across firms at each year-end.

		No. of insurance firms		No. of ABS securities held by each firm				% ABS holding in all bond positions (par value)				% ABS holding in all bond positions (fair value)			
	year	All	Firms Hold ABS	Mean	10th Pct	Median	90th Pct	Mean	10th Pct	Median	90th Pct	Mean	10th Pct	Median	90th Pct
Life	2004	618	471	32.41	1	10	71	4.64%	0.57%	3.46%	8.96%	4.55%	0.57%	3.52%	8.73%
	2005	589	463	42.56	2	11	100	5.19%	0.63%	4.14%	10.96%	5.19%	0.66%	4.18%	10.91%
	2006	574	454	54.49	2	14	126	6.49%	0.67%	5.09%	14.07%	6.48%	0.70%	5.20%	14.14%
	2007	552	436	65.12	2	15	156	7.29%	0.76%	5.87%	15.45%	6.92%	0.75%	5.71%	14.34%
	2008	546	428	67.81	2	16	181	7.42%	0.72%	5.95%	16.26%	5.45%	0.58%	4.27%	11.84%
	2009	530	417	65.06	2	15	178	6.45%	0.52%	4.82%	13.70%	4.74%	0.46%	3.54%	10.25%
	2010	507	388	62.00	1	13	178	5.31%	0.39%	4.02%	11.61%	4.19%	0.34%	3.36%	8.96%
Property & Casualty	2004	1338	778	7.70	1	5	17	4.06%	0.54%	3.33%	8.40%	4.05%	0.54%	3.26%	8.54%
	2005	1353	813	9.11	1	5	19	4.42%	0.60%	3.21%	8.78%	4.42%	0.59%	3.21%	8.78%
	2006	1346	831	10.77	1	6	23	5.12%	0.52%	3.70%	11.57%	5.08%	0.53%	3.64%	11.22%
	2007	1374	864	11.99	1	7	26	5.49%	0.61%	3.79%	11.59%	5.26%	0.56%	3.70%	11.21%
	2008	1420	857	11.99	1	6	26	4.88%	0.54%	3.13%	10.77%	3.70%	0.43%	2.38%	8.08%
	2009	1404	813	10.66	1	4	22	3.62%	0.30%	2.17%	7.85%	2.75%	0.27%	1.73%	6.12%
	2010	1385	673	8.42	1	3	18	2.95%	0.23%	1.59%	6.95%	2.35%	0.19%	1.29%	5.37%

Table 3: Ratings Downgrades of ABS Securities Held by Insurance Firms

This table reports the change of ratings in S&P's downgrades of ABS securities that was held by any insurance companies at the end of year proceeding the downgrades. We include downgrades occurring during the financial crisis, i.e., from the third quarter of 2007 to the end of 2009.

Rating Before Downgrade	Rating After Downgrade: AA	A	BBB	BB	B	CCC	CC	C	D	Total	% of Total
AAA	948	666	573	455	721	131	7	0	2	3503	40.1%
AA		451	329	136	166	132	21	0	0	1235	14.2%
A			572	268	224	206	34	0	1	1305	15.0%
BBB				395	276	307	80	0	5	1063	12.2%
BB					268	308	59	0	0	635	7.3%
B						514	95	0	2	611	7.0%
CCC							340	3	19	362	4.1%
CC								0	10	10	0.1%
C									3	3	0.0%
Total	948	1117	1474	1254	1655	1598	636	3	42	8727	100.0%
% of Total	10.9%	12.8%	16.9%	14.4%	19.0%	18.3%	7.3%	0.0%	0.5%	100.0%	

Table 4: Accounting Treatment of Downgraded ABS

This table reports statistics on insurance companies' accounting treatment of downgraded ABS. Two types of downgrade are considered: (a) from investment to non-investment grades and (b) from AAA to non-investment grade. All downgrades during 2005 and 2010 are considered in Panel A, but only the downgrades in the fourth quarter of each year are considered in Panel B. Over the year in which the downgrade occurs, each position held at the beginning of year is classified into three groups: (i) kept at historical cost (HCA), (ii) kept but revalued to the market, and (iii) sold. The percentages of these groups are reported, conditional on the beginning-of-year (previous year-end) accounting treatment.

Panel A: All Downgrades in 2005-2010

Accounting Treatment in Previous Year	Life				Property & Casualty			
	Total Number	Treatment after Downgrade			Total Number	Treatment after Downgrade		
		HCA	Revalued	Sell		HCA	Revalued	Sell
A-1: Investment to Non-Investment Grades								
HCA	5,337	70%	15%	14%	1,588	40%	39%	21%
Revalued	694	43%	30%	27%	495	27%	35%	39%
Total	6,031				2,083			
A-2: AAA to Non-Investment Grade								
HCA	1,940	78%	9%	13%	851	45%	36%	20%
Revalued	59	75%	14%	12%	119	19%	34%	47%
Total	1,999				970			

Panel B: All Downgrades in Fourth Quarter

Accounting Treatment in Previous Year	Life				Property & Casualty			
	Total Number	Treatment after Downgrade			Total Number	Treatment after Downgrade		
		HCA	Revalued	Sell		HCA	Revalued	Sell
B-1: Investment to Non-Investment Grades								
HCA	1,235	73%	14%	13%	327	20%	60%	20%
Revalued	47	47%	26%	28%	59	20%	31%	49%
Total	1,282				386			
B-2: AAA to Non-Investment Grade								
HCA	532	78%	11%	11%	220	16%	63%	20%
Revalued	3	67%	0%	33%	23	0%	48%	52%
Total	535				243			

Table 5: Summary Statistics of Unrealized Gains/Losses of ABS and Corporate Bond Positions

This table presents descriptive characteristics on the distribution of unrealized gains/losses on ABS (Panel A) and corporate bonds (Panel B) held by insurance firms. For each bond position, unrealized gain is the difference between the position's fair value and book-adjusted carrying value, measured as a percentage of book-adjusted carrying value.

Panel A: Unrealized Gains/Losses of ABS

	Life Firms						Property and Casualty Firms					
Year	# of Positions	10thPct	25thPct	50thPct	75thPct	90thPct	# of Positions	10thPct	25thPct	50thPct	75thPct	90thPct
2004	14,196	-1.5%	-0.2%	0.8%	4.4%	10.1%	5,642	-1.7%	-0.8%	0.0%	1.8%	5.9%
2005	18,519	-3.3%	-1.8%	-0.1%	1.0%	5.3%	7,072	-3.2%	-2.0%	-0.7%	0.1%	1.8%
2006	23,180	-2.7%	-1.1%	0.0%	1.0%	3.7%	8,508	-2.7%	-1.4%	-0.2%	0.2%	1.4%
2007	27,819	-17.0%	-7.9%	-2.0%	0.2%	2.8%	9,886	-7.4%	-1.7%	-0.3%	0.6%	1.9%
2008	27,507	-75.3%	-55.4%	-30.3%	-8.5%	-0.3%	8,416	-54.6%	-31.9%	-12.1%	-1.9%	0.0%
2009	24,927	-60.8%	-38.0%	-16.0%	-0.1%	7.6%	6,733	-35.4%	-15.0%	-1.5%	1.7%	9.5%
2010	21,056	-36.2%	-18.5%	-2.5%	4.8%	23.6%	4,281	-16.0%	-4.0%	1.4%	6.9%	28.5%

Panel B: Unrealized Gains/Losses of Corporate Bonds

	Life Firms						Property and Casualty Firms					
Year	# of Positions	10thPct	25thPct	50thPct	75thPct	90thPct	# of Positions	10thPct	25thPct	50thPct	75thPct	90thPct
2004	156,282	-0.9%	0.8%	4.0%	9.3%	15.3%	69,897	-1.3%	-0.1%	2.0%	6.4%	11.6%
2005	155,060	-3.5%	-1.6%	0.6%	5.0%	11.0%	67,288	-3.4%	-2.2%	-0.5%	2.2%	7.0%
2006	147,888	-4.0%	-2.0%	0.1%	3.3%	8.0%	66,790	-3.5%	-2.2%	-0.3%	1.6%	5.3%
2007	146,744	-6.8%	-2.6%	0.1%	2.8%	6.9%	63,947	-3.6%	-1.1%	0.3%	2.1%	4.8%
2008	144,740	-31.5%	-18.0%	-6.6%	0.0%	4.9%	63,542	-17.7%	-7.6%	-1.8%	1.0%	4.5%
2009	139,580	-6.1%	-0.1%	4.2%	7.8%	13.1%	69,066	-0.5%	1.8%	4.7%	7.6%	11.9%
2010	124,450	-0.9%	2.7%	7.1%	11.8%	17.4%	61,337	0.3%	2.4%	6.3%	10.4%	14.9%

Table 6: Probability of Revaluing ABS

This table reports logit estimates for the probability that an insurance company revalues its ABS position. The dependent variable is a dummy that equals one if the insurance company revalues the ABS at a particular year-end, and zero otherwise. Panel A reports the coefficient estimates. Panel B reports the marginal effects of being a P&C insurance company (over being a life company), evaluated for (i) the crisis vs. non-crisis periods and (ii) the ABS being downgraded from investment to non-investment grades vs. not during the year. All other variables are held at the sample means. Standard errors, clustered by insurance company, are in parentheses. *, **, and *** refer to statistical significance at 10%, 5%, and 1% levels. Variable definitions are in Appendix A.

Panel A: Coefficient Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Non-Crisis	Crisis	Non-Crisis	Crisis	All	All	All	All
P&C dummy	0.346	1.440***	0.531**	1.616***	0.360	0.379	0.560**	0.597**
	(0.245)	(0.140)	(0.215)	(0.155)	(0.252)	(0.277)	(0.231)	(0.255)
Downgrade dummy			0.400	0.387***			0.709***	0.660***
			(0.276)	(0.107)			(0.157)	(0.124)
P&C dummy x Downgrade dummy			1.645***	0.276*			0.387**	0.364**
			(0.424)	(0.165)			(0.168)	(0.162)
Crisis dummy					-0.441*		-0.517*	
					(0.263)		(0.276)	
P&C dummy x Crisis dummy					0.994***	0.998***	0.940***	0.929***
					(0.286)	(0.323)	(0.292)	(0.322)
ln(maturity)			0.266***	0.429***			0.306***	0.335***
			(0.058)	(0.056)			(0.046)	(0.047)
ln(issue size)			-0.191***	-0.263***			-0.211***	-0.228***
			(0.026)	(0.025)			(0.021)	(0.019)
Rating group fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
State fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES	NO	YES	NO	YES
Observations	104,482	108,757	102,301	106,373	213,448	213,448	208,881	208,881
Pseudo R-squared	0.103	0.210	0.122	0.238	0.129	0.147	0.153	0.172

Panel B: Marginal Effects of P&C Dummy

	(1) Non-Crisis	(2) Crisis	(3) Non-Crisis	(4) Crisis	(5) All	(6) All	(7) All	(8) All
Crisis dummy = 0	0.019				0.019	0.019		
Crisis dummy = 1		0.081***			0.083***	0.072***		
Crisis dummy = 0 and Downgrade dummy = 0			0.030**				0.030**	0.030**
Crisis dummy = 0 and Downgrade dummy = 1			0.596**				0.148**	0.129**
Crisis dummy = 1 and Downgrade dummy = 0				0.089***			0.103***	0.077***
Crisis dummy = 1 and Downgrade dummy = 1				0.216***			0.342***	0.241***

Table 7: Probability of Selling Downgraded ABS

This table reports logit estimates for the probability that an insurance company sells its downgraded ABS within 3 months after the downgrade. The first two columns are based on all downgrades. Columns (3) and (4) are based on the downgrades from an investment grade to a non-investment grade. Columns (5) and (6) are based on the downgrades from AAA to a non-investment grade. Panel A reports the coefficient estimates. Panel B reports the marginal effects of (i) the company being a P&C insurance company (over being a life company) and (ii) the position being revalued by the company at previous year-end. All other variables are held at the sample means. Standard errors, clustered by insurance company, are in parentheses. *, **, and *** refer to statistical significance at 10%, 5%, and 1% levels. Variable definitions are in Appendix A.

	(1) All Downgrades	(2) All Downgrades	(3) Investment to Non-Investment	(4) Investment to Non-Investment	(5) AAA to Non-Investment	(6) AAA to Non-Investment
Panel A: Coefficient Estimates						
P&C dummy	0.431***	0.334***	0.431***	0.348**	0.722***	0.698***
	(0.115)	(0.121)	(0.146)	(0.156)	(0.205)	(0.215)
Revalue dummy		0.397***		0.561***		0.809**
		(0.131)		(0.202)		(0.361)
ln(maturity)		0.063		-0.303		-1.058***
		(0.100)		(0.247)		(0.376)
ln(issue size)		0.042		0.073*		0.151**
		(0.034)		(0.044)		(0.064)
Rating group (before downgrade) fixed effects	YES	YES	YES	YES	YES	YES
State fixed effects	YES	YES	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES	YES	YES
Observations	22,471	21,571	6,910	6,646	2,582	2,450
Pseudo R-squared	0.0814	0.0852	0.112	0.120	0.116	0.126
Panel B: Marginal Effects						
P&C dummy	0.032***	0.025***	0.035***	0.028**	0.062***	0.060***
Revalue dummy		0.031***		0.050**		0.083*

Table 8: Gains Trading and Probability of Selling Corporate Bonds

This table reports logit estimates for the effects of unrealized gain on the probability that an insurance company will sell the bond during (i) non-crisis and (ii) crisis periods. The dependent variable is a dummy that equals one if the insurance company holding the bond at the beginning of the quarter sells the bond during the quarter, and zero otherwise. Panel A reports the coefficient estimates. Panel B reports the effects of moving from the 25th to 75th percentiles of unrealized gains, evaluated for (i) the crisis vs. non-crisis periods, (ii) the insurance companies in the top quartile of exposure to ABS not revalued at the previous year-end, and (iii) the insurance companies in the bottom quartile of RBC ratio. All other variables are held at the sample means. Standard errors, clustered by insurance company, are in parentheses. *, **, and *** refer to statistical significance at 10%, 5%, and 1% levels. Variable definitions are in Appendix A.

Panel A: Coefficient Estimates

	Life Firms						Property and Casualty Firms					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Crisis	Crisis	Non-Crisis	Non-Crisis	All	All	Crisis	Crisis	Non-Crisis	Non-Crisis	All	All
Main variables												
(1) Unrealized gain percentile	0.224**	0.219**	-0.375***	-0.404***	-0.387***	-0.403***	-0.318***	-0.289***	-0.077	-0.081	-0.070	-0.068
	(0.104)	(0.099)	(0.074)	(0.073)	(0.075)	(0.075)	(0.106)	(0.095)	(0.063)	(0.063)	(0.063)	(0.062)
(2) High ABS exposure dummy	0.229*	0.251*	0.338***	0.268***	0.332***	0.267***	0.194**	0.213***	0.220***	0.241***	0.212***	0.244***
	(0.124)	(0.129)	(0.080)	(0.080)	(0.079)	(0.078)	(0.085)	(0.079)	(0.073)	(0.074)	(0.074)	(0.074)
(1) x (2)	0.013	-0.008	-0.323***	-0.311***	-0.318***	-0.314***	0.390**	0.309**	0.032	0.050	0.034	0.030
	(0.151)	(0.141)	(0.118)	(0.120)	(0.119)	(0.120)	(0.156)	(0.151)	(0.121)	(0.109)	(0.122)	(0.111)
(3) Low RBC ratio dummy	0.317**	0.252*	0.098	-0.023	0.087	-0.005	-0.011	-0.028	0.077	0.062	0.081	0.067
	(0.152)	(0.141)	(0.072)	(0.090)	(0.070)	(0.083)	(0.094)	(0.094)	(0.081)	(0.079)	(0.081)	(0.077)
(1) x (3)	-0.111	-0.136	0.332**	0.404**	0.329**	0.377**	0.181	0.158	0.150	0.212*	0.164	0.206
	(0.166)	(0.157)	(0.155)	(0.162)	(0.153)	(0.158)	(0.172)	(0.176)	(0.127)	(0.128)	(0.125)	(0.126)
(1) x (2) x (3)	0.342*	0.428*	-0.120	-0.132	-0.133	-0.095	-0.437	-0.313	-0.195	-0.300	-0.202	-0.266
	(0.207)	(0.253)	(0.280)	(0.284)	(0.271)	(0.286)	(0.319)	(0.327)	(0.205)	(0.184)	(0.202)	(0.192)
(4) Revalue dummy	0.599***	0.638***	0.517***	0.651***	0.566***	0.684***	0.212***	0.181**	0.256***	0.264***	0.250***	0.263***
	(0.109)	(0.115)	(0.090)	(0.170)	(0.078)	(0.147)	(0.075)	(0.075)	(0.063)	(0.058)	(0.064)	(0.057)
Crisis dummy					-0.591***						-0.082	
					(0.080)						(0.071)	
Crisis dummy x (1)					0.703***	0.694***					-0.221**	-0.203**
					(0.115)	(0.108)					(0.105)	(0.103)
Crisis dummy x (2)					-0.081	-0.011					-0.044	-0.063
					(0.115)	(0.109)					(0.096)	(0.095)
Crisis dummy x (1) x (2)					0.295*	0.299*					0.389**	0.347**
					(0.172)	(0.171)					(0.170)	(0.166)
Crisis dummy x (3)					0.226	0.197					-0.076	-0.115
					(0.154)	(0.143)					(0.103)	(0.104)
Crisis dummy x (1) x (3)					-0.441	-0.459*					0.026	-0.013
					(0.273)	(0.235)					(0.203)	(0.207)

Cont'd next page

Cont'd from previous page

	Life Firms						Property and Casualty Firms					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Crisis	Crisis	Non-Crisis	Non-Crisis	All	All	Crisis	Crisis	Non-Crisis	Non-Crisis	All	All
Crisis dummy x (1) x (2) x (3)					0.468*	0.398*					-0.263	-0.158
					(0.283)	(0.241)					(0.338)	(0.340)
Crisis dummy x (4)					0.136	0.012					-0.064	-0.085
					(0.098)	(0.151)					(0.094)	(0.103)
<u>Bond control variables</u>												
ln(bond age)	-0.217***	-0.197***	-0.223***	-0.208***	-0.239***	-0.198***	-0.180***	-0.179***	-0.189***	-0.185***	-0.181***	-0.177***
	(0.028)	(0.027)	(0.023)	(0.023)	(0.021)	(0.021)	(0.032)	(0.030)	(0.023)	(0.023)	(0.023)	(0.021)
ln(maturity)	-0.322***	-0.326***	-0.127***	-0.127***	-0.195***	-0.190***	-0.105***	-0.114***	0.046*	0.045*	-0.002	-0.004
	(0.034)	(0.033)	(0.023)	(0.023)	(0.022)	(0.020)	(0.033)	(0.031)	(0.027)	(0.026)	(0.022)	(0.021)
ln(issue size)	0.356***	0.347***	0.307***	0.293***	0.310***	0.311***	0.242***	0.236***	0.231***	0.223***	0.232***	0.228***
	(0.016)	(0.015)	(0.012)	(0.012)	(0.012)	(0.011)	(0.015)	(0.015)	(0.012)	(0.012)	(0.011)	(0.011)
Bankruptcy dummy	1.881***	1.907***	2.329***	2.412***	1.953***	2.017***	1.534***	1.560***	1.961***	1.940***	1.531***	1.610***
	(0.097)	(0.100)	(0.152)	(0.158)	(0.080)	(0.085)	(0.279)	(0.281)	(0.296)	(0.315)	(0.234)	(0.246)
Downgrade dummy	1.061***	1.067***	1.472***	1.482***	1.309***	1.285***	1.491***	1.502***	1.496***	1.529***	1.413***	1.491***
	(0.076)	(0.075)	(0.047)	(0.048)	(0.043)	(0.045)	(0.079)	(0.080)	(0.073)	(0.073)	(0.058)	(0.059)
<u>Insurance control variables</u>												
ln(capital and surplus)	0.026	0.028	0.027	0.002	0.025	0.012	-0.002	0.026	0.012	0.020	0.014	0.028
	(0.032)	(0.033)	(0.027)	(0.025)	(0.025)	(0.024)	(0.029)	(0.033)	(0.021)	(0.020)	(0.020)	(0.020)
% risky assets	0.007	0.007	0.006	0.007**	0.006*	0.007**	0.004	0.004	-0.000	-0.000	-0.000	-0.000
	(0.005)	(0.005)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.000)	(0.000)	(0.000)	(0.000)
Leverage	0.491	0.501	0.382	0.448	0.407	0.482*	0.281	0.031	0.013	-0.049	0.030	-0.059
	(0.417)	(0.384)	(0.323)	(0.305)	(0.293)	(0.271)	(0.453)	(0.419)	(0.323)	(0.306)	(0.300)	(0.291)
ROE	-0.062	-0.237*	0.034	-0.039	0.030	-0.107	0.027	-0.006	-0.386	-0.457*	-0.276	-0.368
	(0.144)	(0.129)	(0.128)	(0.133)	(0.103)	(0.102)	(0.459)	(0.475)	(0.250)	(0.270)	(0.236)	(0.250)
Rating group fixed effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
State fixed effects	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Quarter fixed effects	YES	YES	YES	YES	NO	YES	YES	YES	YES	YES	NO	YES
Observations	1,109,964	1,109,964	1,609,938	1,609,938	2,719,902	2,719,902	526,873	526,873	774,144	774,144	1,301,017	1,301,017
Pseudo R-squared	0.056	0.066	0.037	0.048	0.040	0.052	0.043	0.052	0.037	0.043	0.034	0.044

Table 8, cont'd: Gains Trading and Probability of Selling Corporate Bonds

Panel B: Effects of Moving from the 25th to 75th Percentiles of Unrealized Gain

	Life Firms (Mean Selling Probability = 0.044)			Property and Casualty Firms (Mean Selling Probability = 0.054)		
	Non-Crisis	Crisis	Difference	Non-Crisis	Crisis	Difference
Models (5) and (11)						
High ABS exposure dummy = 0 and Low RBC ratio dummy = 0	-0.006***	0.004**	0.010***	-0.002	-0.005**	-0.004
High ABS exposure dummy = 1 and Low RBC ratio dummy = 1	-0.011***	0.013***	0.024***	-0.002	-0.003	-0.001
Models (6) and (12)						
High ABS exposure dummy = 0 and Low RBC ratio dummy = 0	-0.005***	0.005**	0.011***	-0.001	-0.005**	-0.004
High ABS exposure dummy = 1 and Low RBC ratio dummy = 1	-0.007***	0.015***	0.022***	-0.003	-0.002	0.000

Table 9: Trade Price Impact of Gains Trading on Corporate Bonds

This table reports coefficients of regressions of relative sale prices on estimated gains-trading selling pressure faced by insurance firms. For each sale transaction, relative sale price is calculated as the logged ratio of an insurance firm's sale price over the median market trade price for the bond during the week (from TRACE). Only the weeks in which each bond trades at least 3 times are used. Gains-trading selling pressure is measured as the incremental effect of unrealized gain on predicted selling probability of an insurance firm's position in the bond during the applicable calendar quarter. The incremental effect of unrealized gain on selling probability is the change in predicted probability as the unrealized gain percentile increases from 0.5 to the actual value, under the logit model (2) in Table 8, estimated separately for each calendar quarter. Standard errors, clustered by bond issuer level, are in parentheses. *, **, and *** refer to statistical significance at 10%, 5%, and 1% levels. Variable definitions are in Appendixes A.

	Life Firms			Property and Casualty Firms		
	(1)	(2)	(3)	(4)	(5)	(6)
Main variables						
Crisis dummy	-0.005*	-0.001		-0.006***	-0.005***	
	(0.003)	(0.002)		(0.002)	(0.002)	
Gains-trading selling pressure		0.003	-0.036		-0.002	0.007
		(0.035)	(0.053)		(0.016)	(0.023)
Crisis dummy x Gains-trading selling pressure		-0.249***	-0.202**		-0.097	-0.080
		(0.069)	(0.092)		(0.078)	(0.076)
Bond control variables						
ln(bond age)	0.001	0.001	0.001	0.001*	0.001*	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
ln(maturity)	0.001**	0.001	0.001	0.002**	0.002**	0.002**
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
ln(issue size)	-0.000	-0.000	-0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Bankruptcy dummy	-0.080	-0.072	-0.072	-0.056	-0.055	-0.060
	(0.059)	(0.057)	(0.057)	(0.043)	(0.042)	(0.042)
Downgrade dummy	-0.015**	-0.016**	-0.016**	-0.008	-0.008	-0.008
	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)

Cont'd next page

Cont'd from previous page

	Life Firms			Property and Casualty Firms		
	(1)	(2)	(3)	(4)	(5)	(6)
Insurance control variables						
ABS exposure	-0.014	-0.011	-0.014	0.018	0.015	0.014
	(0.021)	(0.020)	(0.019)	(0.018)	(0.017)	(0.017)
ln(RBC ratio)	0.004	0.003	0.004	-0.001	-0.002	-0.002
	(0.005)	(0.005)	(0.005)	(0.001)	(0.001)	(0.001)
ln(capital and surplus)	-0.002*	-0.002*	-0.002*	-0.002**	-0.002**	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
% risky assets	-0.000	-0.000	-0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Leverage	-0.001	0.001	0.001	-0.014*	-0.014*	-0.014*
	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)
ROE	0.000	-0.001	-0.001	-0.000	-0.000	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Market and transaction control variables						
ln(trade size)	0.003**	0.003**	0.003**	0.003**	0.003**	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
ln(number of sale trades in quarter)	0.003**	0.003***	0.003***	0.004***	0.004***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
ln(median market trade price)	-0.097**	-0.100**	-0.101**	-0.073**	-0.073**	-0.075**
	(0.045)	(0.045)	(0.046)	(0.035)	(0.035)	(0.036)
ln(range of market trade price)	-0.002*	-0.001	-0.002*	-0.003***	-0.003***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Rating group fixed effects	YES	YES	YES	YES	YES	YES
State fixed effects	YES	YES	YES	YES	YES	YES
Calendar quarter fixed effects	NO	NO	YES	NO	NO	YES
Observations	102,399	102,399	102,399	66,590	66,590	66,590
R-squared	0.044	0.047	0.048	0.041	0.041	0.044

Table 10: The Impact of Gains Trading on Corporate Bond Return

This table reports coefficients of regressions of quarterly bond return on average gains-trading selling pressure from insurance firms. To be included in the regression, the bonds must be held by at least 9 life insurance firms (25th percentile) at the end of previous year. Quarterly bond return is the log of change in prices on the last day when there are any trades of a bond from the previous quarter. Treasury return is the return on maturity-matched Treasury bond/note, proxied by the interpolated constant maturity Treasury bond/note from the Fed. Spread return is the maturity- and rating-matched corporate bond index return minus Treasury return. Corporate bond index return is calculated using Bank of America-Merrill Lynch bond index, adjusted for duration difference between the index and the bond of interest. For each bond in each quarter, gains-trading selling pressure is measured as either the incremental effect of unrealized gain on selling probability or the percentage unrealized gain, (value-weighted or equally-weighted) averaged across all positions of life insurance companies in the bond. The incremental effect of unrealized gain on selling probability is the change in predicted probability as the unrealized gain percentile increases from 0.5 to the actual value, under the logit model (2) in Table 8, estimated separately for each calendar quarter. Standard errors, clustered by bond issuer level, are in parentheses. *, **, and *** refer to statistical significance at 10%, 5%, and 1% levels. Variable definitions are in Appendix A.

	Average Incremental Selling Probability of Unrealized Gain				Average Percentage Unrealized Gain			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Equal Weighted		Value Weighted		Equal Weighted		Value Weighted	
Gains-trading selling pressure	5.304	10.838	-4.076	0.807	0.604	-0.600	0.833	0.160
	(8.828)	(11.112)	(10.597)	(13.454)	(1.300)	(2.694)	(1.348)	(2.155)
Crisis dummy x	-116.745***	-134.698***	-131.824***	-150.364***	-6.595**	-7.536*	-14.607***	-18.819***
Gains-trading selling pressure	(22.368)	(27.606)	(29.819)	(35.103)	(3.288)	(4.383)	(1.637)	(2.038)
Treasury return	0.643***	0.660***	0.643***	0.660***	0.633***	0.639***	0.626***	0.634***
	(0.034)	(0.033)	(0.033)	(0.033)	(0.038)	(0.037)	(0.037)	(0.037)
Spread return	0.598***	0.604***	0.602***	0.607***	0.596***	0.594***	0.584***	0.581***
	(0.037)	(0.037)	(0.035)	(0.035)	(0.043)	(0.043)	(0.042)	(0.042)
ln(bond age)	-0.075	-0.302	-0.051	-0.307*	-0.025	0.108	0.028	-0.613**
	(0.066)	(0.185)	(0.064)	(0.186)	(0.063)	(0.356)	(0.064)	(0.272)
ln(issue size)	0.098*		0.084*		-0.049		-0.013	
	(0.051)		(0.049)		(0.047)		(0.045)	
ln(maturity)	-0.085*	-0.156	-0.066	-0.131	-0.059	-0.108	-0.075	-0.308
	(0.052)	(0.266)	(0.049)	(0.264)	(0.048)	(0.353)	(0.048)	(0.310)
Downgrade dummy	-3.308*	-2.025	-3.339*	-2.089	-4.036***	-2.159	-4.510***	-2.687*
	(1.982)	(1.620)	(1.965)	(1.609)	(1.546)	(1.589)	(1.558)	(1.598)

Cont'd next page

Cont'd from previous page

	Average Incremental Selling Probability of Unrealized Gain				Average Percentage Unrealized Gain			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Equal Weighted		Value Weighted		Equal Weighted		Value Weighted	
Bankruptcy dummy	-13.760*	-10.955	-13.291*	-10.455	-17.141*	-14.286	-17.259*	-14.192
	(8.335)	(8.965)	(7.750)	(8.315)	(9.812)	(10.648)	(9.593)	(10.420)
Rating group fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Calendar quarter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bond fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Cluster at bond issuer level	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	84,234	84,234	84,234	84,234	84,234	84,234	84,234	84,234
R-squared	0.313	0.320	0.314	0.321	0.315	0.319	0.323	0.328
Number of bonds		8,272		8,272		8,272		8,272

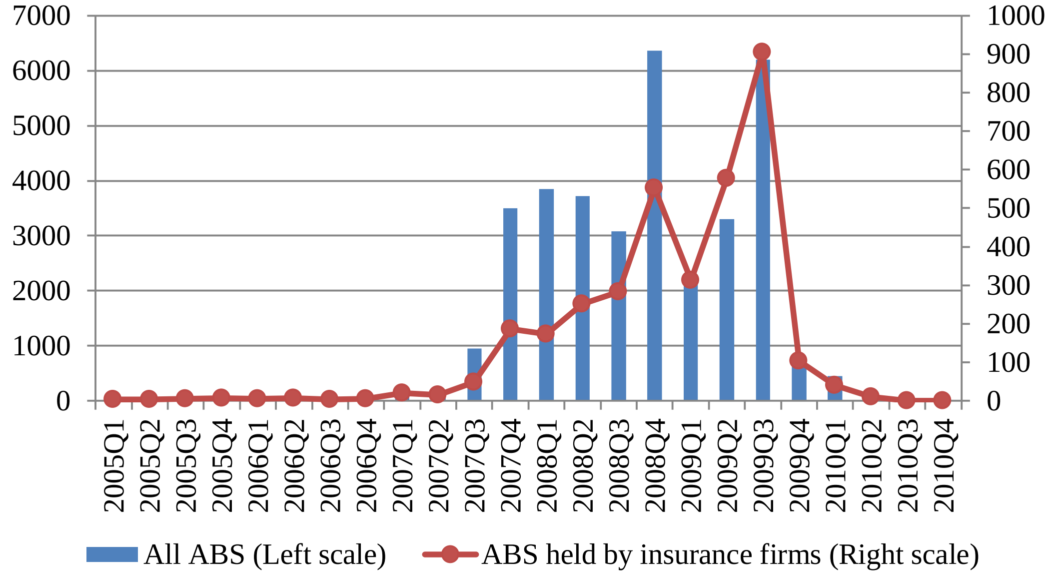


Figure 1: Number of Downgrades of ABS by S&P from Investment Grade to Speculative Grade
This figure presents the number of downgrades of ABS securities from an investment grade to a speculative grade by S&P on quarterly basis. The bar shows the number of such downgrades of all ABS securities included in S&P's Ratings IQuerry. We count only the downgrades affecting insurance companies (i.e., downgrades of the ABS securities held by any insurance firms) with the connected dots.

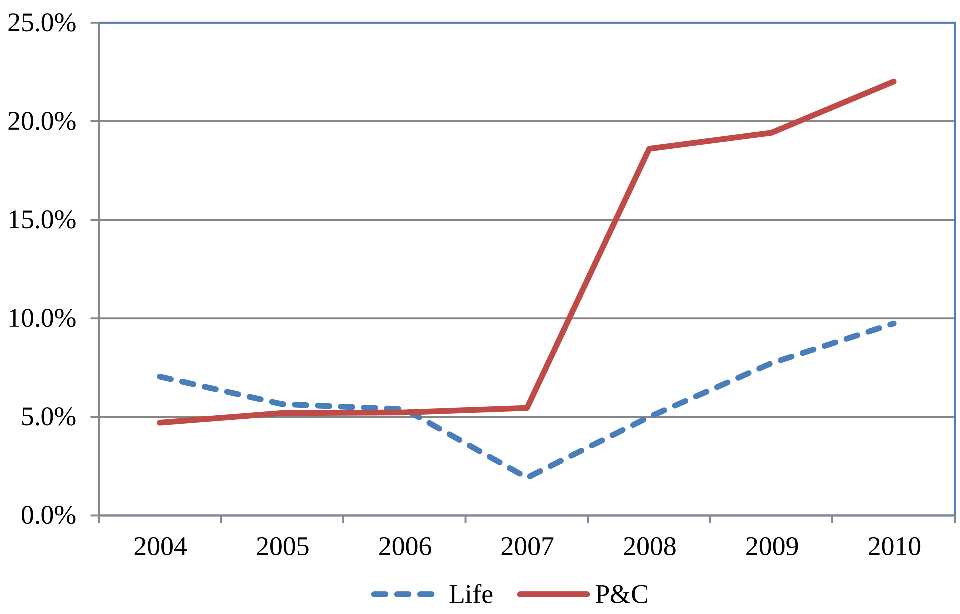


Figure 2: Fraction of ABS Positions Revalued at Year End
This figure presents the number of ABS positions revalued to the market value as a percentage of all ABS positions across all life and P&C firms at the end of 2004-2010. We classify a position as revalued to the market value if the book or adjusted carrying value equals the fair value reported at year-end.

Financial Intermediary Capital*

Adriano A. Rampini
Duke University

S. Viswanathan
Duke University

First draft: July 2010
This draft: March 2012

Abstract

We propose a dynamic theory of financial intermediaries as collateralization specialists that are better able to collateralize claims than households. Intermediaries require capital as they can borrow against their loans only to the extent that households themselves can collateralize the assets backing the loans. The net worth of financial intermediaries and the corporate sector are both state variables affecting the spread between intermediated and direct finance and the dynamics of real economic activity, such as investment, and financing. The accumulation of net worth of intermediaries is slow relative to that of the corporate sector. A credit crunch has persistent real effects and can result in a delayed or stalled recovery. We provide sufficient conditions for the comovement of the marginal value of firm and intermediary capital.

Keywords: Collateral; Financial intermediation; Financial constraints; Investment

*We thank Nittai Bergman, Doug Diamond, Emmanuel Farhi, Itay Goldstein, Bengt Holmström, Nobu Kiyotaki, David Martinez-Miera, Alexei Tchistyi, and seminar participants at the IMF, the MIT theory lunch, Boston University, the Federal Reserve Bank of New York, the Stanford University macro lunch, the Federal Reserve Bank of Richmond, the 2010 SED Annual Meeting, the 2010 Tel Aviv University Finance Conference, the 2011 Jackson Hole Finance Conference, the 2011 FIRS Annual Conference, the 2011 WFA Annual Meeting, the 2011 CEPR European Summer Symposium in Financial Markets, the 2011 FARFE Conference, and the 2012 AEA Conference for helpful comments. This paper subsumes the results on financial intermediation in our 2007 paper "Collateral, financial intermediation, and the distribution of debt capacity," which is now titled "Collateral, risk management, and the distribution of debt capacity" (Rampini and Viswanathan (2010)). Address: Duke University, Fuqua School of Business, 100 Fuqua Drive, Durham, NC, 27708. Rampini: Phone: (919) 660-7797. Email: rampini@duke.edu. Viswanathan: Phone: (919) 660-7784. Email: viswanat@duke.edu.

1 Introduction

The capitalization of financial intermediaries is arguably critical for economic fluctuations and growth. We provide a dynamic model in which financial intermediaries are collateralization specialists and firms need to collateralize promises to pay with tangible assets. Financial intermediaries are modeled as lenders that are able to collateralize a larger fraction of tangible assets than households who lend to firms directly, that is, are better able to enforce their claims. Financial intermediaries require net worth as their ability to refinance their collateralized loans from households is limited, as they, too, need to collateralize their promises. The net worth of financial intermediaries is hence a state variable and affects the dynamics of the economy. Importantly, both firm and intermediary net worth play a role in our model and jointly affect the dynamics of firm investment, financing, and loan spreads. Spreads on intermediated finance are high when both firms and financial intermediaries are poorly capitalized and in particular when intermediaries are moreover poorly capitalized relative to firms. One of our main results is that intermediaries accumulate net worth more slowly than the corporate sector. This has important implications for economic dynamics. For example, a credit crunch, that is, a drop in intermediary net worth, has persistent real effects and can result in a delayed or stalled recovery.

In our model, firms can raise financing either from households or from financial intermediaries. Firms have to collateralize their promises to pay due to limited enforcement.[1] Both households and intermediaries extend collateralized loans, but financial intermediaries are better able to collateralize promises and hence are able to extend more financing per unit of tangible assets collateralizing their loans. Financial intermediaries in turn are able to borrow against their loans, but only to the extent that other lenders themselves can collateralize the assets backing the loans. Intermediaries thus need to finance the additional amount that they are able to lend out of their own net worth. Since intermediary net worth is limited, intermediated finance commands a positive spread.

The determinants of the capital structure for firms and intermediaries differ. Firms' capital structure is determined by the extent to which the tangible assets required for production can be collateralized. Intermediaries' capital structure is determined by the extent to which their collateralized loans can be collateralized themselves. In other words, firms issue promises against tangible assets whereas intermediaries issue promises against collateralized claims, which are in turn backed by tangible assets.

Intermediaries are essential in our economy in the sense that allocations can be

[1] Rampini and Viswanathan (2010, 2012) provide a dynamic model with collateral constraints which are explicitly derived in an environment with limited enforcement.

achieved with financial intermediaries, which cannot be achieve otherwise. Financial intermediaries have constant returns in our model and hence there is a representative financial intermediary. We first consider the equilibrium spread on intermediated finance in a static environment with a representative firm.[2] Importantly, the spread on intermediated finance critically depends on both firm and intermediary net worth. Given the (representative) firm's net worth, spreads are higher when the intermediary is less well capitalized. However, spreads are particularly high when firms are poorly capitalized, and intermediaries are poorly capitalized relative to firms at the same time. Poor capitalization of the corporate sector per se does not imply high spreads, as low firm net worth reduces the demand for loans from intermediaries. Given the net worth of the intermediary sector, a reduction in the net worth of the corporate sector may reduce spreads as the intermediaries can more easily accommodate the reduced loan demand.

Our model allows the analysis of the dynamics of intermediary capital. A main result is that the accumulation of net worth of intermediaries is slow relative to that of the corporate sector. We first consider the deterministic dynamics of intermediary net worth and the spread on intermediated finance. In a deterministic steady state, intermediaries are essential, have positive net worth, and the spread on intermediated finance is positive. Dynamically, if firms and intermediaries are initially poorly capitalized, both firms and intermediaries accumulate net worth over time. Importantly, firms in our model accumulate net worth faster than financial intermediaries, because the marginal and in particular the average return on net worth for financially constrained firms is relatively high due to the high marginal product of capital. Financial intermediaries accumulate net worth at the interest rate earned on intermediated finance, which is at most the marginal return on net worth of the corporate sector and may be below when the collateral constraint for intermediated finance binds. Thus, intermediaries, with constant returns to scale, earn at most the marginal return on all their net worth, whereas firms, with decreasing returns to scale, earn the average return on their net worth.

Suppose that firms are initially poorly capitalized also relative to financial intermediaries. Then the dynamics of the spread on intermediated finance are as follows. Because the firms are poorly capitalized, the current demand for intermediated finance is low and the spread on intermediated finance is zero. Intermediaries save net worth by lending to households to meet higher future corporate loan demand. As the firms accumulate more net worth, their demand for intermediated finance increases, and intermediary finance

[2] In Appendix A, we analyze the choice between intermediated and direct finance in the cross section of firms in a static environment. More constrained firms borrow more from intermediaries, which is empirically plausible and similar to the results in Holmström and Tirole (1997).

becomes scarce and the spread rises. The spread continues to rise as long as the firm's collateral constraint for intermediated finance binds. Once the spread gets so high that the collateral constraint is slack, the spread declines again as both firms and intermediaries accumulate net worth. Since intermediary net worth accumulates more slowly, firms may temporarily accumulate more net worth and then later on re-lever as they switch to more intermediated finance when intermediaries become better capitalized. Eventually, the spread on intermediated finance declines to the steady state spread as intermediaries accumulate their steady state level of net worth.

A credit crunch, modeled as a drop in intermediary net worth, has persistent real effects in our model. While small drops to intermediary net worth can be absorbed by a cut in dividends, larger shocks reduce intermediary lending and raise the spread on intermediated finance. Real investment drops, and indeed drops even if the corporate sector is well capitalized, as the rise in the cost of intermediated finance raises firms' cost of capital. Remarkably, the recovery of investment after a credit crunch can be delayed, or stall, as the cost of intermediated finance only starts to fall once intermediaries have again accumulated sufficient net worth.

In a stochastic economy, we provide sufficient conditions for the marginal value of intermediary and firm net worth to comove. For example, if intermediary net worth is sufficiently low, these values comove and indeed move proportionally. Thus, the marginal value of intermediary net worth may be high exactly when the marginal value of firm net worth is high, too.

Few extant theories of financial intermediaries provide a role for intermediary capital. Notable is in particular Holmström and Tirole (1997) who model intermediaries as monitors that cannot commit to monitoring and hence need to have their own capital at stake to have incentives to monitor. In their analysis, firm and intermediary capital are exogenous and the comparative statics with respect to these are analyzed. Holmström and Tirole conclude that "[a] proper investigation ... must take into account the feedback from interest rates to capital values. This will require an explicitly dynamic model, for instance, along the lines of Kiyotaki and Moore [1997a]." We provide a dynamic model in which the joint evolution of firm and intermediary net worth and the interest rate on intermediated finance are endogenously determined. Diamond and Rajan (2001) and Diamond (2007) model intermediaries as lenders which are better able to enforce their claims due to their specific liquidation or monitoring ability in a similar spirit to our model, but do not consider equilibrium dynamics. In contrast, the capitalization of financial intermediaries plays essentially no role in liquidity provision theories of financial intermediation (Diamond and Dybvig (1983)), in theories of financial intermediaries as delegated, diversified monitors (Diamond (1984), Ramakrishnan and Thakor (1984), and Williamson (1986)) or in coalition based theories (Townsend (1978) and Boyd and Prescott (1986)).

Dynamic models in which net worth plays a role, such as Bernanke and Gertler (1989) and Kiyotaki and Moore (1997a), typically consider the role of firm net worth only, although dynamic models in which intermediary net worth matters have recently been considered (see, for example, Gertler and Kiyotaki (2010), who also summarize the recent literature, and Brunnermeier and Sannikov (2010)). However, to the best of our knowledge, we are the first to consider a dynamic model in which both firm and intermediary net worth are critical and jointly affect the dynamics of financing, spreads, and economic activity.

In Section 2 we describe the model. Section 3 studies how the spread on intermediated finance varies with firm and intermediary net worth in a simplified static version of the model. The dynamics of intermediary capital are analyzed in Section 4. We first consider the deterministic steady state and dynamics of firm and intermediary capital, and the dynamic effects of a credit crunch. We then provide sufficient conditions for the comovement of the marginal value of intermediary and firm net worth in a stochastic economy. Section 5 concludes. All proofs are in Appendix B.

2 Model

We consider a model in which promises to pay need to be collateralized due to limited enforcement. There are three types of agents: households, financial intermediaries, and agents that run firms; we discuss these in turn. We consider an environment with a representative firm. Time is discrete and the horizon infinite. There is an exogenous state $s \in S$, which determines the firm's productivity, that follows a Markov chain with transition probability $\Pi(s, s')$, where S is a finite state space.[3]

2.1 Households

There is a continuum of households (of measure 1) in the economy which are risk neutral and discount future payoffs at a rate $R > 1$ where $R^{-1} > \beta$ and $\beta \in (0, 1)$ is the discount rate of agents who run firms, that is, households are more patient than the agents who run firms. These lenders are assumed to have a large endowment of funds in all dates and states, and have a large amount of collateral and hence are not subject to enforcement

[3]In a slight abuse of notation, we denote the cardinality of S by S as well.

problems but rather are able to commit to deliver on their promises. They are willing to provide any state-contingent claim at an expected rate of return R so long as such claims satisfy the firms' and intermediaries' collateral constraints.

2.2 Financial intermediaries as collateralization specialists

There is a continuum of financial intermediaries (of measure 1) which are risk neutral, subject to limited liability, and discount future payoffs at β_i where $\beta_i \in (\beta, R^{-1})$. Financial intermediaries are *collateralization specialists.* Intermediaries are able to seize up to fraction $\theta_i \in (0,1)$ of the (resale value of) collateral backing promises issued to them; we assume that $\theta_i > \theta$ where $\theta \in (0,1)$ is the fraction of collateral that households can seize. The left-hand side of Figure 1 illustrates this, interpreting the fraction θ as structures, which both households and intermediaries can collateralize, and the fraction $\theta_i - \theta$ as equipment, which only financial intermediaries can collateralize. Financial intermediaries can in turn issue claims against their collateralized loans. Lenders to financial intermediaries can lend to intermediaries up to the amount of the collateral backing the intermediaries' loans that they themselves can seize. Consider the problem of a representative financial intermediary[4] with current net worth w_i and given the state of the economy $Z \equiv \{s, w, w_i\}$ which includes the exogenous state s as well as two endogenous state variables, the net worth of the corporate sector w and the net worth of the intermediary sector w_i. The state-contingent interest rate on intermediated finance R_i' depends on state s' and the state Z of the economy, as shown below, but we suppress the argument for notational simplicity.

The intermediary maximizes the discounted value of future dividends by choosing a dividend payout policy d_i, state-contingent loans to households l', state-contingent intermediated loans to firms l_i', and state-contingent net worth w_i' next period to solve

$$v_i(w_i, Z) = \max_{\{d_i, l', l_i', w_i'\} \in \mathbb{R}_+^{1+3\#Z}} d_i + \beta_i E\left[v_i(w_i', Z')\right] \tag{1}$$

subject to the budget constraints

$$\begin{aligned} w_i &\geq d_i + E[l'] + E[l_i'], &(2)\\ Rl' + R_i' l_i' &\geq w_i'. &(3) \end{aligned}$$

[4] We consider a representative financial intermediary since intermediaries have constant returns to scale in our model and hence aggregation in the intermediation sector is straightforward. The distribution of intermediaries' net worth is hence irrelevant and only the aggregate capital of the intermediation sector matters.

We denote variables which are measurable with respect to the next period, that is, depend on the state s', with a prime; that is, we use the shorthand $w' \equiv w(s')$ and analogously for other variables.

Note that we state the intermediary's problem as if the intermediary only lends the additional amount it can collateralize. This simplifies the notation and analysis. We do not need to consider the intermediary's collateral constraint explicitly, as the firms' collateral constraint for financing ultimately provided by the households already ensures that this constraint is satisfied, rendering the additional constraint redundant. However, whenever the intermediary is essential in the sense that the allocation cannot be supported without an intermediary, the interpretation is that the firms' claims are held by the intermediary and the intermediary in turn refinances the claims with households to the extent that they can collateralize the claims themselves. In contrast, we interpret financing which does not involve the intermediary as direct or unintermediated financing.

The first order conditions, which are necessary and sufficient, can be written as

$$\mu_i = 1 + \eta_d, \tag{4}$$

$$\mu_i = R\beta_i\mu_i' + R\beta_i\eta', \tag{5}$$

$$\mu_i = R_i'\beta_i\mu_i' + R_i'\beta_i\eta_i', \tag{6}$$

$$\mu_i' = v_{i,w}(w_i', Z'), \tag{7}$$

where the multipliers on the constraints (2) through (3) are μ_i and $\Pi(Z, Z')\beta_i\mu_i'$, and η_d, $\Pi(Z, Z')R\beta_i\eta'$, and $\Pi(Z, Z')R_i'\beta_i\eta_i'$ are the multipliers on the non-negativity constraints on dividends and direct and intermediated lending; the envelope condition is $v_{i,w}(w_i, Z) = \mu_i$.

2.3 Corporate sector

There is a representative firm which is risk neutral and subject to limited liability and discounts the future at rate β. The representative firm (which we at times refer to simply as the firm or the corporate sector) has limited net worth w and has access to a standard neoclassical production technology $A'f(k)$ where $A' > 0$ is the stochastic total factor productivity, $f(\cdot)$ is the production function, and k is the amount of capital the firm deploys next period, which depreciates at the rate $\delta \in (0, 1)$. We assume that the production function $f(\cdot)$ is strictly increasing and strictly concave and satisfies the usual Inada condition. Total factor productivity A' depends on the exogenous state s' next period, that is, $A' \equiv A(s')$. We suppress the dependence on s' and use the short-hand A' throughout as discussed above. The firm can raise financing from both households and intermediaries by issuing one-period collateralized state-contingent claims b' to households and b_i' to intermediaries.

We write the representative firm's problem recursively. The firm maximizes the discounted expected value of future dividends by choosing a dividend payout policy d, capital k, state-contingent promises b' and b'_i to households and intermediaries, and state-contingent net worth w' for the next period, taking the state-contingent interest rates on intermediated finance R'_i and their law of motion as given, to solve:

$$v(w,Z) = \max_{\{d,k,b',b'_i,w'\}\in\mathbb{R}^2_+\times\mathbb{R}^S\times\mathbb{R}^{2S}_+} d + \beta E\left[v(w',Z')\right] \tag{8}$$

subject to the budget constraints

$$w + E\left[b' + b'_i\right] \;\geq\; d + k, \tag{9}$$

$$A' f(k) + k(1-\delta) \;\geq\; w' + Rb' + R'_i b'_i, \tag{10}$$

and the collateral constraints

$$\theta k(1-\delta) \;\geq\; Rb', \tag{11}$$

$$(\theta_i - \theta)k(1-\delta) \;\geq\; R'_i b'_i, \tag{12}$$

where θ is the fraction of tangible assets, that is, capital, that households can collateralize while θ_i is the fraction of tangible assets that intermediaries can collateralize. Since the firm issues state-contingent claims to both households and intermediaries and pricing of the state-contingent loans is risk neutral, it is the expected value of the claims that enters the budget constraint in the current period, equation (9). Depending on the realized state next period, the firm repays Rb' to households and $R'_i b'_i$ to financial intermediaries as the budget constraint for the next period, equation (10), shows. The interest rate on direct finance R is constant as discussed above. The middle and right-hand side of Figure 1 illustrate the collateral constraints (11) and (12). Note that the expectation operator $E[\cdot]$ denotes the expectation conditional on state Z, but the dependence on the state is again suppressed to simplify notation.

Importantly, to simplify the analysis we use notation that keeps track separately of the claims that are ultimately financed by households (b') and the claims that are financed by intermediaries out of their own net worth b'_i. In particular, whenever the firm borrows from financial intermediaries and issues strictly positive promises $R'_i b'_i$, the corresponding promises Rb' should be interpreted as being financed by the intermediary who in turn refinances them by issuing equivalent promises to households. Thus, we do not distinguish between claims financed by households directly, and claims financed by households indirectly by lending to financial intermediaries against collateral backing intermediaries' loans. This allows a simple formulation of the collateral constraints: firms can borrow up to fraction θ of the resale value of their capital by issuing claims to households (whether these are held directly or are indirectly financed via the intermediary) and can borrow up to the difference in collateralization rates, $\theta_i - \theta$, additionally by issuing claims which are financed by intermediaries out of their own net worth. We elaborate on the enforcement and settlement of claims below.[5]

The first order conditions, which are necessary and sufficient, can be written as

$$\mu = 1 + \nu_d, \tag{13}$$

$$\mu = E\left[\beta\left(\mu'\left[A' f_k\left(k\right) + \left(1-\delta\right)\right] + \left[\lambda'\theta + \lambda_i'(\theta_i - \theta)\right]\left(1-\delta\right)\right)\right], \tag{14}$$

$$\mu = R\beta\mu' + R\beta\lambda', \tag{15}$$

$$\mu = R_i'\beta\mu' + R_i'\beta\lambda_i' - R_i'\beta\nu_i', \tag{16}$$

$$\mu' = v_w(w', Z'), \tag{17}$$

where the multipliers on the constraints (9) through (12) are μ, $\Pi(Z,Z')\beta\mu'$, $\Pi(Z,Z')\beta\lambda'$, and $\Pi(Z,Z')\beta\lambda_i'$, and ν_d and $\Pi(Z,Z')R_i'\beta\nu_i'$ are the multipliers on the non-negativity constraints on dividends and intermediated borrowing;[6] the envelope condition is $v_w(w,Z) = \mu$.

2.4 Enforcement and settlement

Rampini and Viswanathan (2010, 2012) study an economy with limited enforcement and show that the optimal allocation can be implemented with complete markets in one period ahead Arrow securities subject to state-by-state collateral constraints. These collateral constraints are similar to the collateral constraints in Kiyotaki and Moore (1997a), except that they are state-contingent. The borrowers' and intermediaries' collateral constraints we analyze in this paper are in a similar spirit, although we do not derive them explicitly from limited enforcement constraints here.

An important additional aspect that arises in the context with financial intermediation is the enforcement of claims intermediaries issue against loans they hold. Our formulation of the contracting problem with separate constraints for promises ultimately issued to households and promises financed by intermediaries themselves allows us to

[5]A model with two types of collateral constraints is also studied by Caballero and Krishnamurthy (2001) who consider international financing in a model in which firms can raise funds from domestic and international financiers subject to separate collateral constraints.

[6]We use $\Pi(Z,Z')$ for the transition probability of the state of the economy in a slight abuse of notation. We ignore the constraints that $k \geq 0$ and $w' \geq 0$ as they are redundant, due to the Inada condition and the fact that the firms can never credibly promise their entire net worth next period (which can be seen by combining (10) at equality with (11) and (12).

sidestep this issue. Nevertheless, it is important to be explicit about our assumptions about enforcement. We assume that collateralized promises can be used as collateral to back other promises, to the extent that other lenders themselves can enforce payment on such promises. Specifically, per unit of the resale value of tangible assets, firms in our model can borrow a fraction θ from households and a fraction θ_i from intermediaries. Intermediaries in turn can use the collateralized claims they own to back their own promises to other lenders. However, per unit of collateral value backing their loans, intermediaries can only refinance fraction θ from other lenders, which is less than the repayment they themselves can enforce, that is, θ_i. Thus, intermediaries are forced to finance the difference, $\theta_i - \theta$, out of their own net worth. In contrast, an intermediary can promise the entire value θ_i to other intermediaries, that is, the interbank market is frictionless in our model, which is why we are able to consider a representative financial intermediary.

In terms of limited enforcement, the assumption is that firms can abscond with all cash flows and a fraction $1-\theta$ of collateral backing promises to households and a fraction $1-\theta_i$ of collateral backing promises to financial intermediaries. Financial intermediaries in turn can abscond with their collateralized claims except to the extent that the collateral backing their claims is in turn collateral backing their own promises to households, that is, they can abscond with $\theta_i - \theta$ per unit of collateral. If a financial intermediary were to default on its promises, its lenders could enforce a claim up to the fraction θ of collateral backing the intermediary's loans directly from corporate borrowers.

2.5 Equilibrium

We now define an equilibrium in our economy. An equilibrium determines both aggregate economic activity and the cost of intermediated finance in our economy.

Definition 1 (Equilibrium) *An equilibrium is an allocation $x \equiv [d, k, b', b'_i, w']$ for the representative firm and $x_i \equiv [d_i, l', l'_i, w'_i]$ for the representative intermediary for all dates and states and a state-contingent interest rate process R'_i for intermediated finance such that (i) x solves the firm's problem in (8)-(12) and x_i solves the intermediary's problem (1)-(3) and (ii) the market for intermediated finance clears in all dates and states*

$$l'_i = b'_i. \tag{18}$$

Note that equilibrium promises are default free, as the promises satisfy the collateral constraints (11) and (12), which ensures that neither firms nor financial intermediaries are able to issue promises on which it is not credible to deliver. While this is of course the implementation that we study throughout, we emphasize that the promises traded in our economy are contingent claims and that these contingent claims may be implemented in practice with noncontingent claims on which issuers are expected and in equilibrium indeed do default (see Kehoe and Levine (2006) for an implementation with equilibrium default in this spirit).

2.6 Endogenous minimum down payment requirement

Define the *minimum down payment requirement* $\wp$ when the firm borrows the maximum amount it can from households only as $\wp = 1 - R^{-1}\theta(1-\delta)$.[7] Similarly, define the minimum down payment requirement when the firm borrows the maximum amount it can from both households (at interest rate R) and intermediaries (at state-contingent interest rate R_i') as $\wp_i(R_i') = 1 - [R^{-1}\theta + E[(R_i')^{-1}](\theta_i - \theta)](1-\delta)$ (illustrated on the right-hand side of Figure 1). Note that the minimum down payment requirement, at times referred to as the margin requirement, is endogenous in our model. Using this definition and equations (14) through (16) the firm's investment Euler equation can then be written concisely as

$$1 \geq E\left[\beta \frac{\mu'}{\mu} \frac{A' f_k(k) + (1-\theta_i)(1-\delta)}{\wp_i(R_i')}\right]. \tag{19}$$

2.7 User cost of capital with intermediated finance

We can extend Jorgenson's (1963) definition of the user cost of capital to our model with intermediated finance. Define the premium on internal funds ρ as $1/(R+\rho) \equiv E[\beta\mu'/\mu]$ and the premium on intermediated finance ρ_i as $1/(R+\rho_i) \equiv E[(R_i')^{-1}]$. Using (14) through (16) the user cost of capital u is

$$u \equiv r + \delta + \frac{\rho}{R+\rho}(1-\theta_i)(1-\delta) + \frac{\rho_i}{R+\rho_i}(\theta_i - \theta)(1-\delta), \tag{20}$$

where $r + \delta$ is the frictionless user cost derived by Jorgenson (1963) and $r \equiv R - 1$. The user cost of capital exceeds the user cost in the frictionless model, because part of investment needs to be financed with internal funds which are scarce and hence command a premium ρ (the second term on the right hand side) and part of investment is financed with intermediated finance which commands a premium ρ_i, as the funds of intermediaries are scarce as well (the last term on the right hand side).[8]

[7] We use the character $\wp$, a fancy script p, for down *payment* (`\wp` in LaTeX and available under miscellaneous symbols).

[8] Alternatively, the user cost can be written in a weighted average cost of capital representation as $u \equiv R/(R+\rho)(r_w + \delta)$ where the weighted average cost of capital r_w is defined as $r_w \equiv (r+\rho)\wp_i(R_i') +$

Internal funds and intermediated finance are both scarce in our model and command a premium as collateral constraints drive a wedge between the cost of different types of finance. The premium on internal finance is higher than the premium on intermediated finance, as the firm would never be willing to pay more for intermediated finance than the premium on internal funds.

Proposition 1 (Premia on internal and intermediated finance) *The premium on internal finance ρ (weakly) exceeds the premium on intermediated finance ρ_i*

$$\rho \geq \rho_i \geq 0,$$

and the two premia are equal, $\rho = \rho_i$, iff the collateral constraint for intermediated finance does not bind for any state next period, that is, $E[\lambda_i'] = 0$. Moreover, the premium on internal finance is strictly positive, $\rho > 0$, iff the collateral constraint for direct finance binds for some state next period, that is, $E[\lambda'] > 0$.

When all collateral constraints are slack, there is no premium on either type of finance, but typically the inequalities are strict and both premia are strictly positive, with the premium on internal finance strictly exceeding the premium on intermediated finance.

3 Effect of intermediary capital on spreads

In this section we study how the choice between intermediated and direct finance varies with firm and intermediary net worth in a static (one period) version of our model with a representative firm.[9] We further simplify but considering the deterministic case, although the results in this section do not depend on this assumption.[10] The equilibrium spread on intermediated finance depends on both firm and intermediary net worth. Given firm net worth, spreads are higher when the intermediary is less well capitalized. Importantly, the spread on intermediated finance depends on the relative capitalization of firms and

$rR^{-1}\theta(1-\delta)+(r+\rho_i)(R+\rho_i)^{-1}(\theta_i-\theta)(1-\delta)$. The cost of capital r_w is a weighted average of the fraction of investment financed with internal funds which cost $r+\rho$ (first term on the right hand side), the fraction financed with households funds at rate r (second term), and the fraction financed with intermediated funds at rate $r+\rho_i$ (third term).

[9]The capital structure implications for the cross section of firms with different net worth is analyzed in Appendix A.

[10]With one period only, the interest rate on intermediated finance is independent of the state s', as the marginal value of net worth next period for financial intermediaries and firms equals 1 for all states, that is, $\mu' = \mu_i' = 1$, rendering the model effectively deterministic.

intermediaries. Spreads are particularly high when firms are poorly capitalized and intermediaries are relatively poorly capitalized at the same time. Poor capitalization of the corporate sector does not per se imply high spreads, as firms' limited ability to pledge may result in a reduction in firms' loan demand which intermediaries with given net worth can more easily accommodate.[11]

The representative intermediary solves

$$\max_{\{d_i,l',l'_i,w'_i\}\in\mathbb{R}^4_+} d_i + \beta_i w'_i \tag{21}$$

subject to (2) through (3). The representative firm solves

$$\max_{\{d,k,b',b'_i,w'\}\in\mathbb{R}^2_+\times\mathbb{R}\times\mathbb{R}^2_+} d + \beta w' \tag{22}$$

subject to (9) through (12). An equilibrium is defined in Definition 1. In addition to the equilibrium allocation, the spread on intermediated finance, $R'_i - R$, is determined in equilibrium.

The following proposition summarizes the characterization of the equilibrium spread. Figures 2 through 4 illustrate the results. The key insight is that the spread on intermediated finance depends on both the firm and intermediary net worth. Importantly, low capitalization of the corporate sector does not necessarily result in a high spread on intermediated finance. Indeed, it may reduce spreads. Similarly, while low capitalization of the intermediation sector raises spreads, spreads are substantial only when the corporate sector is poorly capitalized and intermediaries are poorly capitalized relative to the corporate sector at the same time.

Proposition 2 (Firm and intermediary net worth) ***(i)*** *For $w_i \geq w_i^*$, intermediaries are well capitalized and there is a minimum spread on intermediated finance $\beta_i^{-1} - R > 0$ for all levels of firm net worth.* ***(ii)*** *Otherwise, there is a threshold of firm net worth $\underline{w}(w_i)$ (which depends on w_i) such that intermediaries are well capitalized and the spread on intermediated finance is $\beta_i^{-1} - R > 0$ as long as $w \leq \underline{w}(w_i)$. For $w > \underline{w}(w_i)$, intermediated finance is scarce and spreads are higher. For $w_i \in [\bar{w}_i, w_i^*)$, spreads are increasing in w until w reaches $\hat{w}(w_i)$, at which point spreads stay constant at $\hat{R}'_i(w_i) - R \in (\beta_i^{-1} - R, \beta^{-1} - R]$. For $w_i \in (0, \bar{w}_i)$, spreads are increasing in w until w reaches $\hat{w}(w_i)$, then decreasing in w until $\bar{w}(w_i)$ is reached, at which point spreads stay constant at $\beta^{-1} - R$. As $w_i \to 0$, $\hat{w}(w_i) \to 0$.*

[11] Note that in Holmström and Tirole (1997) aggregate investment only depends on the sum of firm and intermediary capital.

Figure 2 displays the cost of intermediated finance as a function of firm net worth (w) and intermediary net worth (w_i). Figure 3 displays the contours of the various areas in Figure 2. Figure 4 displays the cost of intermediated finance as a function of firm net worth for different levels of intermediary net worth, and is essentially a projection of Figure 2. When financial intermediaries are well capitalized the spread on intermediated finance is at its minimum, $\beta_i^{-1} - R > 0$. This is the case when financial intermediary net worth is high enough ($w_i \geq w_i^*$) so that they can accommodate the loan demand of even a well capitalized corporate sector or when corporate net worth is relatively low so that the financial intermediary sector is able to accommodate demand despite its low net worth ($w \leq \underline{w}(w_i)$). When intermediary capital is below w_i^* and the corporate sector is not too poorly capitalized ($w > \underline{w}(w_i)$), spreads on intermediated finance are higher. Indeed, when intermediary capital is in this range, higher firm net worth initially raises spreads as loan demand increases (until firm net worth reaches $\hat{w}(w_i)$). This effect can be substantial when $w_i < \bar{w}_i$. Indeed, interest rates in our example increase to around 200% when financial intermediary net worth is very low, albeit our example is not calibrated. If firm net worth is still higher, spreads decline as the marginal product of capital and hence firms' willingness to borrow at high interest rates declines. When corporate net worth exceeds $\bar{w}(w_i)$, the cost on intermediated finance is constant at β^{-1}, which equals the shadow cost of internal funds of well capitalized firms.

To sum up, spreads are determined by firm and intermediary net worth jointly. Spreads are higher when intermediary net worth is lower. But firm net worth affects both the demand for intermediated loans and, via investment, the collateral available to back such loans. When collateral constraints bind, lower firm net worth reduces spreads.

4 Dynamics of intermediary capital

Our model allows the analysis of the dynamics of intermediary capital and indeed the joint dynamics of the capitalization of the corporate and intermediary sector. We first characterize a deterministic steady state and then analyze the deterministic dynamics of firm and intermediary capitalization. Both firms and intermediaries accumulate capital over time, but the corporate sector initially accumulates net worth faster than the intermediary sector, which has important implications for the dynamics of spreads on intermediated finance. We also study the dynamic effects of a credit crunch, and show that the economy may be slow to recover. Finally, we provide sufficient conditions for the marginal values of firm and intermediary net worth to commove.

4.1 Intermediaries are essential in a deterministic economy

We first show that intermediaries always have positive net worth, that is, they never choose to pay out their entire net worth as dividends if the economy is deterministic or eventually deterministic, that is, deterministic from some time $T < +\infty$ onward.

Proposition 3 (Positive intermediary net worth) *Financial intermediaries always have positive net worth in an equilibrium in a deterministic or eventually deterministic economy.*

Since intermediaries always have positive net worth, the interest rate on intermediated finance R_i' must in equilibrium be such that the representative firm never would want to lend at that interest rate, as the following lemma shows:

Lemma 1 *In any equilibrium,* ***(i)*** *the cost of intermediated funds (weakly) exceeds the cost of direct finance, that is, $R_i' \geq R$;* ***(ii)*** *the multiplier on the collateral constraint for direct finance (weakly) exceeds the multiplier on the collateral constraint for intermediated finance, that is, $\lambda' \geq \lambda_i'$; and* ***(iii)*** *the constraint that the representative firm cannot lend at R_i' never binds, that is, $\nu_i' = 0$ w.l.o.g. Moreover, in a deterministic economy,* ***(iv)*** *the constraint that the representative intermediary cannot borrow at R_i' never binds, that is, $\eta_i' = 0$; and* ***(v)*** *the collateral constraint for direct financing always binds, that is, $\lambda' > 0$.*

We define the essentiality of intermediaries as follows:

Definition 2 (Essentiality of intermediation) *Intermediation is* ***essential*** *if an allocation can be supported with a financial intermediary but not without.*[12]

The above results together imply that financial intermediaries must always be essential. First note that firms are always borrowing the maximal amount from households. If firms moreover always borrow a positive amount from intermediaries, then they must achieve an allocation that would not otherwise be feasible. If $R_i' = R$, then the firm must be collateral constrained in terms of intermediated finance, too, that is, borrow a positive amount. If $R_i' > R$, then intermediaries lend all their funds to the corporate sector and in equilibrium firms must be borrowing from intermediaries. We have proved the following:

Proposition 4 (Essentiality of intermediaries) *In an equilibrium in a deterministic economy, financial intermediaries are always essential.*

[12]This definition is analogous to the definition of essentiality of money in monetary theory (see, e.g., Hahn (1973)).

4.2 Intermediary capitalization and spreads in a steady state

We define a deterministic steady state in the economy with an infinite horizon as follows:

Definition 3 (Steady state) *A deterministic steady state equilibrium is an equilibrium with constant allocations, that is,* $x^* \equiv [d^*, k^*, b'^*, b_i'^*, w'^*]$ *and* $x_i^* \equiv [d_i^*, l'^*, l_i'^*, w_i'^*]$.

In the deterministic steady state, intermediaries are essential, have positive capital, and spreads are positive.

Proposition 5 (Steady state) *In a steady state, intermediaries are essential, have positive net worth, and pay positive dividends. The spread on intermediated finance is* $R_i'^* - R = \beta_i^{-1} - R > 0$. *Firms borrow the maximal amount from intermediaries. The relative (ex dividend) intermediary capitalization is*

$$\frac{w_i^*}{w^*} = \frac{\beta_i(\theta_i - \theta)(1-\delta)}{\wp_i(\beta_i^{-1})}.$$

The relative (ex dividend) intermediary capitalization, that is, the ratio of the representative intermediary's net worth (ex dividend) relative to the representative firm's net worth (ex dividend), is the ratio of the intermediary's financing (per unit of capital) to the firm's down payment requirement (per unit of capital). In a steady state, the shadow cost of internal funds of the firm is $\beta^{-1} - 1$ while the shadow cost of internal funds of the intermediary is $\beta_i^{-1} - 1$ and equals the interest rate on intermediated finance $R_i'^* - 1$. Since $\beta_i > \beta$, intermediated finance is cheaper than internal funds for firms in the steady state, and firms borrow as much as they can. In a steady state equilibrium, financial intermediaries have positive capital and pay out the steady state interest income as dividends $d_i^* = (R_i'^* - 1)l_i'^*$. Both firms and intermediaries have positive net worth in the steady state despite the fact that their rates of time preference differ and both are less patient than households.

4.3 Deterministic dynamics of intermediary capital and spreads

Consider the dynamics of both firm and intermediary capitalization in an equilibrium converging to the steady state. We show that the equilibrium dynamics evolve in two main phases, an initial one in which the corporate sector pays no dividends and a second one in which the corporate sector pays dividends. Intermediaries do not pay dividends until the steady state is reached, except that they may pay an initial dividend (at time 0), if they are well capitalized relative to the corporate sector at time 0. We first state these results formally and then provide an intuitive discussion of the equilibrium dynamics.

Proposition 6 (Deterministic dynamics) *Given w and w_i, there exists a unique deterministic dynamic equilibrium which converges to the steady state characterized by a no dividend region (ND) and a dividend region (D) (which is absorbing) as follows:*

Region ND $w_i \leq w_i^*$ *(w.l.o.g.) and* $w < \bar{w}(w_i)$, *and (i)* $d = 0$ $(\mu > 1)$, *(ii) the cost of intermediated finance is*

$$R_i' = \max\left\{R, \min\left\{\frac{(\theta_i - \theta)(1-\delta)\left(\frac{w}{w_i}+1\right)}{\wp}, \frac{A'f_k\left(\frac{w+w_i}{\wp}\right) + (1-\theta)(1-\delta)}{\wp}\right\}\right\},$$

(iii) investment $k = (w + w_i)/\wp$ *if* $R_i' > R$ *and* $k = w/\wp_i(R)$ *if* $R_i' = R$, *and (iv)* $w'/w_i' > w/w_i$, *that is, firm net worth increases faster than intermediary net worth.*

Region D $w \geq \bar{w}(w_i)$ *and (i)* $d > 0$ $(\mu = 1)$. *For* $w_i \in (0, \bar{w}_i)$, *(ii)* $R_i' = \beta^{-1}$, *(iii)* $k = \bar{k}$ *which solves* $1 = \beta[A'f_k(\bar{k}) + (1-\theta)(1-\delta)]/\wp$, *(iv)* $w_{ex}'/w_i' < w_{ex}/w_i$, *that is, firm net worth (ex dividend) increases more slowly than intermediary net worth, and (v)* $\bar{w}(w_i) = \wp\bar{k} - w_i$. *For* $w_i \in [\bar{w}_i, w_i^*)$, *(ii)* $R_i' = (\theta_i - \theta)(1-\delta)k/w_i$, *(iii)* k *solves* $1 = \beta[A'f_k(k) + (1-\theta)(1-\delta)]/(\wp - w_i/k)$, *(iv)* $w_{ex}'/w_i' < w_{ex}/w_i$, *that is, firm net worth (ex dividend) increases more slowly than intermediary net worth, and (v)* $\bar{w}(w_i) = \wp_i(R_i')k$. *For* $w_i \geq w_i^*$, $\bar{w}(w_i) = w^*$ *and the steady state of Proposition 5 is reached with* $d = w - w^*$ *and* $d_i = w_i - w_i^*$.

Figure 5 displays the contours of the two regions in terms of firm net worth w and intermediary net worth w_i and Figure 6 illustrates the dynamics of firm and intermediary net worth, the interest rate on intermediated finance, and investment over time. The representative intermediary's dividend policy is characterized as follows:

Lemma 2 (Initial intermediary dividend) *The representative intermediary pays at most an initial dividend and no further dividends until the steady state is reached. If* $w_i > w_i^*$, *the initial dividend is strictly positive.*

To understand the intuition, suppose both firms and financial intermediaries are initially poorly capitalized, and assume moreover that firms are poorly capitalized even relative to financial intermediaries. The dynamics of financial intermediary net worth are relatively simple, since as long as no dividends are paid (which is the case until the steady state is reached, except possibly at time 0), the intermediaries' net worth evolves according to the law of motion $w_i' = R_i'w_i$, that is, intermediary net worth next period is simply intermediary net worth this period plus interest income. When no dividends are paid, intermediaries lend out all their funds at the interest rate R_i'. Of course, the interest rate R_i' needs to be determined in equilibrium.

Given our assumptions, the corporate sector's net worth, investment and loan demand evolve in several phases, which are reflected in the dynamics of the equilibrium interest rate. If firms are initially poorly capitalized even relative to financial intermediaries, as we assume, loan demand is low and intermediaries are relatively well capitalized. In this case, except for a potential initial dividend, intermediaries conserve net worth to meet future loan demand by lending some of their funds to households (see Panel B3 of Figure 6) and spreads are zero, that is, $R_i' = R$ (see Panel B1). In fact, the intermediaries' lending to households exceeds their lending to the corporate sector early on. Corporate investment is then $k = w/\wp_i(R)$. Intermediaries accumulate net worth at rate R in this phase while the corporate sector accumulates net worth at a faster rate, given the high marginal product; thus, the net worth of the corporate sector rises relative to the net worth of intermediaries. In Figure 6, this phase last from time $t = 0$ to $t = 3$, except that the intermediary pays an initial dividend at $t = 0$, since Figure 6 considers an initial drop in corporate net worth only.

Eventually, the increased net worth of the corporate sector raises loan demand so that intermediated finance becomes scarce. The corporate sector then borrows all the funds intermediaries are able to lend and invests $k = (w + w_i)/\wp$. The interest rate on intermediated finance is determined by the collateral constraint, which is binding, and equals $R_i' = (\theta_i - \theta)(1 - \delta)\left(w/w_i + 1\right)/\wp$. Note that since corporate net worth increases faster than intermediary net worth, the interest rate on intermediated finance rises in this phase. As the corporate sector accumulates net worth, it can pledge more and the equilibrium interest rate rises. In Figure 6, this occurs between $t = 3$ and $t = 4$.

As the net worth and investment of the corporate sector continues to rise faster than intermediary net worth, the increase in firms' collateral means that firms' ability to pledge no longer constrains their ability to raise intermediated finance. Intermediated finance is scarce in this phase because of limited intermediary net worth, however, and so spreads are high but declining. The law of motion of investment is as in the previous phase $k = (w + w_i)/\wp$, while the equilibrium interest rate on intermediated finance is determined by $R_i' = [A' f_k(k) + (1 - \theta)(1 - \delta)]/\wp$. Both firm and intermediary net worth continue to increase, and hence investment increases and the equilibrium interest rate on intermediated finance decreases. In Figure 6, this occurs between $t = 4$ and $t = 5$.

Eventually, the interest rate on intermediated finance reaches β^{-1}, the shadow cost of internal funds of the corporate sector. At that point, corporate investment stays constant and firms start to pay dividends. However, intermediaries continue to accumulate net worth and the economy is not yet in steady state. As intermediaries accumulate net worth, the corporate sector reduces its net worth by paying dividends. Essentially, the corporate sector relevers as the supply of intermediated finance increases when financial intermediary net worth increases. This is the case at $t = 5$ and $t = 6$ in Figure 6.

Once intermediary capital is sufficiently high to accommodate the entire loan demand of the corporate sector at an interest rate β^{-1}, the cost of intermediated funds decreases further. As the interest rate on intermediated finance is now below the shadow cost of internal funds of the corporate sector, the collateral constraint binds again. Investment increases due to the reduced cost of intermediated financing. This phase lasts from $t = 7$ to $t = 9$ in Figure 6. Eventually, intermediaries accumulate their steady state level of net worth and the cost of intermediated finance reaches β_i^{-1}, the intermediaries' shadow cost of internal funds. The steady state is reached at $t = 9$ in Figure 6.

We emphasize two key aspects of the dynamics of intermediary capital, beyond the fact that intermediary and firm net worth affect the dynamics jointly. First, intermediary capital accumulates more slowly than corporate net worth in our model. Second, the interest rate on intermediated finance is low when intermediaries conserve net worth to meet the higher loan demand later on when the corporate sector is temporarily relatively poorly capitalized. And vice versa, the corporate sector accumulates additional net worth and spreads remain higher (and investment lower than in the steady state) as the corporate sector "waits" for intermediary net worth to rise and eventually reduce spreads, at which point firms relever. The second two observations of course are a reflection of the relatively slow pace of intermediary capital accumulation.

4.4 Dynamics of a credit crunch

Suppose the economy experiences a *credit crunch*, which we model here as an unanticipated one-time drop in intermediary net worth w_i. We assume that the economy is otherwise deterministic and is in steady state when the credit crunch hits. Figure 7 illustrates the effects of such a credit crunch on interest rates, net worth, intermediary lending, and investment. The effect of a credit crunch depends on its size. Intermediaries can absorb a small enough credit crunch simply by cutting dividends. But a larger drop in intermediary net worth results in a reduction in lending and an increase in the spread on intermediated finance. Moreover, the higher cost of intermediated finance increases the user cost of capital (20) (as the premium on internal finance is either unchanged or increases) and so investment drops. Thus, a credit crunch has real effects in our model. Remarkably, investment drops even if the corporate sector is still well capitalized (that is, even if $w'^* > \bar{w}$). The reason is that the cost of capital increases even if the corporate sector is well capitalized, as intermediaries' capacity to extend relatively cheap financing is reduced. In that case, the credit crunch results in a jump in the interest rate on intermediated finance to $R_i' = \beta^{-1} > R_i'^* = \beta_i^{-1}$ and an immediate drop in investment (and capital, which drops to $\bar{k} < k^*$). The real effects in our model are moreover persistent, even if the corporate sector remains well capitalized. Indeed, the recovery of the real economy can be delayed. After a sufficiently large credit crunch, investment and capital remain constant at the lower level, and spreads remain constant at the elevated level, until the intermediary sector accumulates sufficient capital to meet the loan demand. At that point, intermediary interest rates start to fall and investment begins to recover, until the economy eventually recovers fully.

If the corporate sector is no longer well capitalized after the credit crunch, the spread on intermediated finance rises further and investment drops even more. This is the case in Figure 7 at time 0 (see Panel B1 and B4). Moreover, after an initial partial recovery, the recovery stalls, potentially for a long time (from time 1 to time 23 in Figure 7), in the sense that the interest rate on intermediated finance remains at $R_i' = \beta^{-1}$ and investment remains constant below its steady state level (in fact, capital remains constant at $\bar{k}$), until the intermediaries accumulate sufficient capital. Then the recovery resumes.

If net worth of both the intermediaries and the corporate sector drop at the same time, for example, because of a one-time depreciation shock to capital, then investment and output fall more substantially. The dynamics of the recovery from such a downturn are as described in Section 4.3. It is noteworthy, though, that the spreads on intermediated finance may or may not go up in such a general downturn, and in fact may well go down despite the scarcity of intermediary capital. The point is that the lower net worth of the corporate sector reduces loan demand, possibly by more than the drop in intermediary net worth reduces loanable funds. If corporate loan demand drops sufficiently, intermediaries may pay a one time dividend when the downturn hits, and then cut dividends to zero until the economy recovers.

4.5 Comovement of firm and intermediary capital

Do the marginal value of firm and intermediary net worth comove? We consider this question in a stochastic economy which is deterministic from time 1 onward. Importantly, this allows both firms and intermediaries to engage in risk management at time 0 and hedge the net worth available to them in different states $s' \in S$ at time 1. We first show that the representative firm optimally engages in incomplete risk management, that is, the collateral constraint for direct finance against at least one state $s' \in S$ must bind. We then provide sufficient conditions for the marginal value of net worth of the representative firm and the representative intermediary to comove.

Proposition 7 (Comovement of the value of firm and intermediary capital) *In an economy that is deterministic from time* 1 *onward and has constant expected productivity,* ***(i)*** *the representative firm must be collateral constrained for direct finance against at least one state at time* 1*;* ***(ii)*** *the marginal value of firm and intermediary net worth comove, in fact* $\mu(s')/\mu(s'_+) = \mu_i(s')/\mu_i(s'_+)$, $\forall s', s'_+ \in S$*, if* $\lambda_i(s') = 0$, $\forall s' \in S$*.* ***(iii)*** *Suppose moreover that there are just two states, that is,* $S = \{\hat{s}', \check{s}'\}$*. If only one of the collateral constraints for direct finance binds,* $\lambda(\check{s}') > 0 = \lambda(\hat{s}')$*, then the marginal values must comove,* $\mu(\hat{s}') > \mu(\check{s}')$ *and* $\mu_i(\hat{s}') \geq \mu_i(\check{s}')$*.*

Proposition 7 implies that the marginal values of firm and intermediary net worth comove, for example, when the intermediary has very limited net worth and hence the collateral constraints for intermediated finance are slack for all states. They also comove if the firm hedges one of two possible states, as then the intermediary effectively must be hedging that state, too. Thus, the marginal value of intermediary net worth may be high exactly when the marginal value of firm net worth is high, too. The marginal values may however move in opposite directions, for example, if a high realization of productivity raises firm net worth substantially, which lowers the marginal product of capital and hence the marginal value of firm net worth, while it may raise loan demand substantially and hence raise the marginal value of intermediary net worth.

5 Conclusion

We develop a dynamic theory of financial intermediation and show that the capital of both the financial intermediary and corporate sector affect real economic activity, such as firm investment, financing, and the spread between intermediated and direct finance. Financial intermediaries are modeled as collateralization specialists that are better able to collateralize claims than households themselves. Financial intermediaries require capital as their ability to borrow against their collateralized loans is limited by households' ability to collateralize the assets backing the loans themselves.

The spread on intermediated finance is high when both firms and intermediaries are poorly capitalized, and in particular when intermediaries are moreover poorly capitalized relative to firms. Intermediary capital in our model accumulates more slowly than the capital of firms, and thus spreads on intermediated finance may initially rise as loan demand increases more than loanable funds as the net worth of the corporate sector increases relative to the net worth of financial intermediaries. A credit crunch, that is, a drop in intermediary net worth results in a drop in intermediated finance, a rise in spreads on intermediated loans, and a drop in real activity. The recovery can be delayed, or stall, with real activity constant at a reduced level and persistently high spreads on intermediated finance, because it takes time for intermediaries to reaccumulate sufficient net worth. In the cross section, the model predicts that more constrained firms borrow from financial intermediaries, consistent with stylized facts. In addition, the model shows that the marginal value of intermediary and firm net worth may comove. Our model may provide a useful framework for the analysis of the dynamic interaction between financial structure and economic activity.

Appendix

Appendix A: Intermediated vs. direct finance in the cross section

This appendix considers the static environment without uncertainty of Section 3 taking the spread on intermediated finance as given to show that our model has plausible implications for the choice between intermediated and direct finance in the cross section of firms. Consider the firm's problem taking the interest rate on intermediated finance R_i' as given. Each firm maximizes (22) subject to (9) through (12) given its net worth w. Severely constrained firms borrow as much as possible from intermediaries while less constrained firms borrow less from intermediaries and dividend paying firms do not borrow from intermediaries at all, consistent with the cross sectional stylized facts. These cross-sectional results are similar to the ones in Holmström and Tirole (1997).

Proposition 8 (Intermediated vs. direct finance across firms) *Suppose $R_i' > \beta^{-1}$.*[13] ***(i)*** *Firms with net worth $w \leq \underline{w}_l$ borrow as much as possible from intermediaries, firms with net worth $\underline{w}_l < w < \underline{w}_u$ borrow a positive amount from intermediaries but less than the maximal amount, and firms with net worth exceeding $\underline{w}_u$ do not borrow from intermediaries, where $0 < \underline{w}_l < \underline{w}_u$.* ***(ii)*** *Only firms with net worth exceeding $\bar{w}$ pay dividends at time 0, where $\underline{w}_u < \bar{w} < \infty$.* ***(iii)*** *Investment is increasing in w and strictly increasing for $w \leq \underline{w}_l$ and $\underline{w}_u < w < \bar{w}$.*

Intermediated finance is costlier than direct finance. Indeed, under the conditions of the proposition, intermediated finance is costlier than the shadow cost of internal finance of well capitalized firms. Thus, well capitalized firms, which pay dividends, do not borrow from financial intermediaries. In contrast, firms with net worth below some threshold ($\underline{w}_u$) have a shadow cost of internal finance which is sufficiently high that they choose to borrow a positive amount from intermediaries. For severely constrained firms, with net worth below $\underline{w}_l$, the shadow cost of internal funds is so high that they borrow as much as they can from intermediaries, that is, their collateral constraint for intermediated finance binds. Moreover, more constrained firms have lower investment and are hence smaller.

The cross-sectional capital structure implications are plausible: smaller (and more constrained) firms borrow more from financial intermediaries and have higher costs of financing, while larger (and less constrained firms) borrow from households, for example in bond markets, and have lower financing costs.

[13] We consider the case in which $R_i' > \beta^{-1}$ since, proceeding analogously as in the first part of the proof, one can show that $R_i' < \beta^{-1}$ would imply that $\lambda_i' > 0$ and thus the cross sectional financing implications would be trivial as all firms would borrow the maximal amount from intermediaries. When $R_i' = \beta^{-1}$, this would also be true without loss of generality.

Appendix B: Proofs

Proof of Proposition 1. Using (16) and the fact that $\nu_i' = 0$ (proved below in Lemma 1, part (iii)), we have $(R_i')^{-1} = \beta\mu'/\mu + \beta\lambda_i'/\mu$ and, taking expectations,

$$\frac{1}{R+\rho_i} \equiv E\left[(R_i')^{-1}\right] = \frac{1}{R+\rho} + E\left[\beta\frac{\lambda_i'}{\mu}\right]$$

and hence $\rho \geq \rho_i$ with equality iff $E[\lambda_i'] = 0$. Moreover, since $R_i' \geq R$ (proved below in Lemma 1, part (i)), $\rho_i \geq 0$. Finally, using (15), we have $1/(R+\rho) \equiv E[\beta\mu'/\mu] = 1/R - E[\beta\lambda'/\mu]$, implying that $\rho > 0$ iff $E[\lambda'] > 0$. □

Proof of Proposition 2. First, consider the intermediary's problem. The first order conditions are (4)-(6) and $\mu_i' = 1 + \eta_d'$, where $\beta_i\eta_d'$ is the multiplier on the constraint $w_i' \geq 0$. Since (3) holds with equality, the non-negativity constraints on l' and l_i' render the non-negativity constraint on w_i' redundant and hence $\mu_i' = 1$. Using (5) we have $\eta' = (R\beta_i)^{-1}\mu_i - 1 \geq (R\beta_i)^{-1} - 1 > 0$ (and $l' = 0$) and similarly using (6) $\eta_i' > 0$ as long as $R_i' < \beta_i^{-1}$. Therefore, for $l_i' > 0$ it is necessary that $R_i' \geq \beta_i^{-1}$. If $R_i' > \beta_i^{-1}$, then $\mu_i' > 1$ (and $l_i' = w_i$) while if $R_i' = \beta_i^{-1}$, $0 \leq l_i' \leq w_i$.

Now consider the representative firm's problem. The first order conditions are (4)-(6) and $\mu' = 1 + \nu_d'$, where $\beta\nu_d'$ is the multiplier on the constraint $w' \geq 0$. Proceeding as in the proof of Proposition 8 one can show that $\mu' = 1$. Suppose $\nu_i' > 0$ (and hence $b_i' = 0$). Since $k > 0$, (12) is slack and $\lambda_i' = 0$. Using (13) and (16) we have $1 \leq \mu < R_i'\beta$ which implies that $R_i' > \beta^{-1}$. But at such an interest rate on intermediated finance $l_i' = w_i > 0$, which is not an equilibrium as $b_i' = 0$. Therefore, $\nu_i' = 0$ and $R_i' \leq \beta^{-1}$. Moreover, if $R_i' < \beta^{-1}$, then $\lambda_i' = (R_i'\beta)^{-1}\mu - 1 > 0$ and hence $b_i' = (R_i')^{-1}(\theta_i - \theta)k(1-\delta) > 0$. Since $l_i' = 0$ if $R_i' < \beta_i^{-1}$, we have $R_i' \in [\beta_i^{-1}, \beta^{-1}]$ in equilibrium. The firm's investment Euler equation (19) simplifies to $1 = \beta(1/\mu)[A'f_k(k) + (1-\theta_i)(1-\delta)]/\wp_i(R_i')$. Given the interest rate on intermediated finance, the firm's problem induces a concave value function and thus μ (weakly) decreases in w, implying that k (weakly) increases.

We first show that intermediaries are well capitalized and there is a minimum spread on intermediated finance $\beta_i^{-1} - R > 0$ for all levels of firm net worth when $w_i \geq w_i^*$ and for levels of firm net worth $w \leq \underline{w}(w_i)$ when $w_i < w_i^*$. If $R_i' = \beta_i^{-1}$, a well capitalized firm invests k^* which solves (19) specialized to $1 = \beta[A'f_k(k^*) + (1-\theta_i)(1-\delta)]/\wp_i(\beta_i^{-1})$, while less well capitalized firms invests $k \leq k^*$. The intermediary can meet the required demand for intermediated finance for any level of firm net worth w if $w_i \geq w_i^* \equiv \beta_i(\theta_i - \theta)k^*(1-\delta)$. Suppose instead that $w_i < w_i^*$. In this case the intermediary is able to meet the firm's loan demand at $R_i' = \beta_i^{-1}$ only if the firm is sufficiently constrained; the constrained firm invests

$k = w/\wp_i(\beta_i^{-1})$ using (9), (11), and (12) at equality, and thus $b_i' = \beta_i(\theta_i - \theta)k(1-\delta)$; the intermediary can meet this demand as long as $w \leq \underline{w}(w_i) \equiv \wp_i(\beta_i^{-1})/[\beta_i(\theta_i - \theta)(1-\delta)]w_i$.

Suppose now that $w_i < w_i^*$ and $w > \underline{w}(w_i)$ as defined above. First, consider $w_i \in [\bar{w}_i, w_i^*)$ where $\bar{w}_i \equiv \beta(\theta_i - \theta)\bar{k}(1-\delta)$ and $1 = \beta[A'f_k(\bar{k}) + (1-\theta)(1-\delta)]/\wp$, that is, $\bar{w}_i$ is the loan demand of the well capitalized firm when the cost of intermediated finance is $R_i' = \beta^{-1}$. Note that $R_i' < \beta^{-1}$ on $(\bar{w}_i, w_i^*)$ since the intermediary has more than enough net worth to accommodate the loan demand of the well capitalized firm (and thus any constrained firm) at $R_i' = \beta^{-1}$. Thus, the firm's collateral constraint binds, that is, $w_i = (R_i')^{-1}(\theta_i - \theta)k(1-\delta)$. If the firm is poorly capitalized, $d = 0$ and (9) implies $w + w_i = \wp k$, and $R_i' = (\theta_i - \theta)(1-\delta)(w/w_i + 1)$. If the firm is well capitalized, $\mu = 1$ and $\bar{k}(w_i)$ solves $1 = \beta[A'f_k(\bar{k}(w_i)) + (1-\theta_i)(1-\delta)]/[\wp - w_i/\bar{k}(w_i)]$. Moreover, $\bar{w}(w_i) \equiv \wp\bar{k}(w_i) - w_i$ and for $w \geq \bar{w}(w_i)$ the cost of intermediated finance is constant at $\bar{R}_i'(w_i) = (\theta_i - \theta)\bar{k}(w_i)(1-\delta)/w_i$. Note that $\bar{R}_i'(w_i^*) = \beta_i^{-1}$ and $\bar{w}(w_i^*) = \wp k^* - w_i^* = \wp_i(\beta_i^{-1})k^* = \underline{w}(w_i^*)$, that is, the two boundaries coincide at w_i^*. In contrast, at $\bar{w}_i$ we have $\underline{w}(\bar{w}_i) = \wp_i(\beta_i^{-1})/[\beta_i(\theta_i - \theta)(1-\delta)]\bar{w}_i = \wp_i(\beta_i^{-1})\beta/\beta_i\bar{k} = \wp\bar{k}\beta/\beta_i - \bar{w}_i < \bar{w}(\bar{w}_i)$ and $\bar{R}_i'(\bar{w}_i) = \beta^{-1}$.

Finally, consider $w_i \in (0, \bar{w}_i)$ and $w > \underline{w}(w_i)$ as defined above. If the firm is well capitalized (16) implies $\lambda_i' = (R_i'\beta)^{-1} - 1 \geq 0$. Moreover, since $w_i < \bar{w}_i$ the intermediary cannot meet the well capitalized firm's loan demand at $R_i' = \beta^{-1}$ and thus the cost of intermediated finance is in fact β^{-1} and $\lambda_i' = 0$, that is, the collateral constraint for intermediated finance does not bind. Thus, the firm's investment Euler equation (19) simplifies to $1 = \beta[A'f_k(\bar{k}) + (1-\theta_i)(1-\delta)]/\wp_i(\beta^{-1})$ which is solved by $\bar{k}$ as defined earlier in the proof. Define $\bar{w}(w_i) \equiv \wp\bar{k} - w_i$; the firm is well capitalized for $w \geq \bar{w}(w_i)$. Suppose $w < \bar{w}(w_i)$ and hence $\mu > 1$. If the collateral constraint for intermediated finance does not bind, then (16) implies $R_i' = \beta^{-1}\mu > \beta^{-1}$ and (19) implies $R_i' = [A'f_k(k) + (1-\theta)(1-\delta)]/\wp$, while (9) yields $w + w_i = \wp k$. Observe that $k < \bar{k}$ and R_i' decreases in w. If instead the collateral constraint binds, then $R_i' = (\theta_i - \theta)k(1-\delta)/w_i$ and $w + w_i = \wp k$ (so long as $w > \underline{w}(w_i)$). Note that k and R_i' increase in w in this range. The collateral constrain is just binding at $\hat{w}(w_i) \equiv \wp\hat{k}(w_i) - w_i$ where $[A'f_k(\hat{k}(w_i)) + (1-\theta)(1-\delta)]/\wp = (\theta_i - \theta)\hat{k}(w_i)(1-\delta)/w_i$.

We now show that if the collateral constraint for intermediated finance binds at some $w < \bar{w}(w_i)$ then it binds for all $w^- < w$. Note that $d = 0$ in this range and $w + w_i = \wp k$. At w^-, either $b_i'^- < w_i$ and $R_i' = \beta_i^{-1}$ and hence $\lambda_i'^- = (\beta_i^{-1}\beta)^{-1}\mu^- - 1 > 0$ or $b_i'^- = w_i$ and $w^- + w_i = \wp k^-$, implying $k^- < k$. Suppose the collateral constraint for intermediated finance is slack at w^-. Then $R_i'^- b_i'^- < (\theta_i - \theta)k^-(1-\delta) < (\theta_i - \theta)k(1-\delta) = R_i' b_i'$ and since $b_i'^{-} = w_i$ and $b_i' \leq w_i$ by above $R_i'^{-} w_i < R_i' b_i' \leq R_i' w_i$ which implies $R_i'^{-} < R_i'$. But

$$R_i'^{-}\beta = \mu^{-} = \beta \frac{A' f_k(k^{-}) + (1-\theta_i)(1-\delta)}{\wp - (R_i'^{-})^{-1}(\theta_i - \theta)(1-\delta)} > \beta \frac{A' f_k(k) + (1-\theta_i)(1-\delta)}{\wp - (R_i')^{-1}(\theta_i - \theta)(1-\delta)} = \mu > R_i'\beta$$

or $R_i'^{-} > R_i'$, a contradiction.

Moreover, $\underline{w}(w_i) < \hat{w}(w_i) < \bar{w}(w_i)$ on $w_i \in (0, \bar{w}_i)$. Suppose, by contradiction, that $\hat{w}(w_i) \leq \underline{w}(w_i)$ and recall that $\underline{w}(w_i) + w_i = \wp k$ and $\hat{w}(w_i) + w_i = \wp \hat{k}(w_i)$, so $\hat{k}(w_i) \leq k$. But $\hat{R}_i'(w_i) = (\theta_i - \theta)\hat{k}(w_i)(1-\delta)/w_i \leq (\theta_i - \theta)k(1-\delta)/w_i = \beta_i^{-1}$. But if $\hat{R}_i'(w_i) \leq \beta_i^{-1}$, then at $\hat{w}(w_i)$ we have $\mu = \hat{R}_i'(w_i)\beta < 1$ (since the collateral constraint is slack), a contradiction. Thus, $\underline{w}(w_i) < \hat{w}(w_i)$. Suppose, again by contradiction, that $\bar{w}(w_i) \leq \hat{w}(w_i)$ and hence $\bar{k} \leq \hat{k}(w_i)$. Recall that $\hat{k}(w_i)$ solves $[A' f_k(\hat{k}(w_i)) + (1-\theta)(1-\delta)]/\wp = (\theta_i - \theta)\hat{k}(w_i)(1-\delta)/w_i$. At $\bar{w}_i$ this equation is solved by $\bar{k}$ (and $\hat{R}_i'(\bar{w}_i) = \beta^{-1}$), but since $w_i < \bar{w}_i$, $\hat{k}(w_i) < \bar{k}$, a contradiction. Moreover, as $w_i \to 0$, $\hat{k}(w_i) \to 0$ and $\hat{w}(w_i) = \wp\hat{k}(w_i) - w_i \to 0$. $\square$

Proof of Proposition 3. Consider a deterministic economy. Suppose intermediaries pay out their entire net worth at some point. From that point on, the firm's problem is as if there is no intermediary. We first characterize the solution to this problem and then show that the solution implies shadow interest rates on intermediated finance at which it would not be optimal for intermediaries to exit.

To characterize the solution in the absence of intermediaries, consider a steady state at which $\mu = \mu' \equiv \bar{\mu}$ and note that (15) implies $\bar{\lambda}' = ((R\beta)^{-1} - 1)\bar{\mu} > 0$. The investment Euler equation (19) simplifies to $1 = \beta[A' f_k(k) + (1-\theta)(1-\delta)/\wp$ which defines $\bar{k}$. The firm's steady state net worth is $\bar{w}' = A' f(\bar{k}) + (1-\theta)\bar{k}(1-\delta)$ and the firm pays out

$$\begin{aligned} \bar{d} &= \bar{w}' - \wp\bar{k} = A' f(\bar{k}) - \bar{k}[1 - (R^{-1}\theta + (1-\theta))(1-\delta)] \\ &> A' f(\bar{k}) - \beta^{-1}\bar{k}[1 - (R^{-1}\theta + \beta(1-\theta))(1-\delta)] \\ &= \int_0^{\bar{k}} [A' f_k(k) - \beta^{-1}(1 - (R^{-1}\theta + \beta(1-\theta))(1-\delta))]dk > 0. \end{aligned}$$

Therefore, $\bar{\mu} = 1$. Investment $\bar{k}$ is feasible as long as $w \geq \bar{w} = \bar{w}' - \bar{d}$. Whenever $w < \bar{w}$, $k < \bar{k}$ and hence using (19) we have $\mu/\mu' = \beta[A' f_k(k) + (1-\theta)(1-\delta)]/\wp > 1$. The shadow interest rate on intermediated finance is $R_i' = \beta^{-1}\mu/\mu' \geq \beta^{-1}$ for all values of w. But then it cannot be optimal for intermediaries to pay out all their net worth in a deterministic economy as keeping $\varepsilon > 0$ net worth for one more period improves the objective by $(\beta_i R_i' - 1)\varepsilon > 0$.

Consider now an eventually deterministic economy. From time T onward, the economy is deterministic and the conclusion obtains by above as long as the intermediary has positive net worth in all states at time T. Suppose not, that is, suppose intermediary net worth is zero for some state. As before the discounted marginal value on an infinitesimal amount of intermediary net worth at time T lent out for one period is at least $\beta_i R_i' \geq \beta_i \beta^{-1} > 1$ since $R_i' \geq \beta^{-1}$. Lending for τ periods thus guarantees a discounted marginal value of $(\beta_i \beta)^\tau$. As $\tau \to \infty$, the marginal value grows without bound. (Note that since we consider an infinitesimal amount, the collateral constraint cannot be biding for any finite τ.) The expected marginal value of this lending policy at time 0 is at least $(\beta_i R)^T$ times the marginal value at time T and hence grows without bound as $\tau \to \infty$.

But the marginal value of intermediary net worth at time 0 is finite as either the intermediary pays dividends and the marginal value is one, or the intermediary saves into at least one state at R_i' and thus $\mu_i = R_i' \beta \mu_i'$ and R_i' is bounded above by (12) and otherwise $R_i' = R$. Furthermore, μ_i' is bounded by a similar argument going forward until dividends are paid at which point the marginal value is one. But then it cannot be an equilibrium for intermediaries to pay out all their net worth. □

Proof of Lemma 1. Part (i): If $R_i' < R$, then using (15) and (16) we have $0 < (R - R_i')\beta\mu' \leq R_i' \beta \lambda_i'$ and thus $b_i' > 0$. But (5) and (6) imply that $0 < (R - R_i')\beta\mu_i' \leq R_i' \beta \eta_i'$ and thus $l_i' = 0$, which is not an equilibrium.

Part (ii): Given $\nu_i' = 0$ (see part (iii)), (15) and (16) imply that $\lambda' = (R_i'/R - 1)\mu' + R_i'/R\lambda_i' \geq \lambda_i'$.

Part (iii): First, suppose to the contrary that $\nu_i' > 0$. Then $\lambda_i' = 0$ as $b_i' = 0 < (R_i')^{-1}(\theta_i - \theta)k(1-\delta)$ implies that (12) is slack. Using (16) and (15) we have $\beta\mu' R_i' > \mu \geq \beta\mu' R$ and thus $R_i' > R$. Equations (5) and (6) imply that $R\eta' - R_i'\eta_i' = (R_i' - R)\mu_i' > 0$ and thus $\eta' > 0$ and $l' = 0$. But if $w_i' > 0$, which is always true under the conditions of Proposition 3, we have $l_i' = (R_i')^{-1} w_i' > 0 = b_i'$, which is not an equilibrium. If instead $w_i' = 0$, then $l_i' = 0$ and we can set $R_i' = (\beta\mu'/\mu)^{-1}$ and $\eta_i' = 0$ w.l.o.g.

Part (iv): Suppose to the contrary that $\eta_i' > 0$ (and hence $l_i' = 0$). Since intermediaries never pay out all their net worth in a deterministic economy, equation (3) implies $0 < w_i' \leq Rl'$ and hence $\eta' = 0$. But then (5) and (6) imply $\beta_i \mu_i'/\mu_i R = 1 > \beta_i \mu_i'/\mu_i R_i'$ or $R > R_i'$ contradicting the result of part (i). Thus, $\eta_i' = 0$ and $\mu_i' = (\beta_i R_i')^{-1}\mu_i$.

Part (v): Suppose $\lambda' = 0$. Then (15) reduces to $1 = \beta\mu'/\mu R$ and thus $1 \leq \mu = \beta R \mu' < \mu'$ and $d' = 0$. By part (ii), $\lambda_i' = 0$ and using (16) we have $R_i' = R$, $\mu_i' = (\beta R)^{-1}\mu_i > 1$, and $d_i' = 0$. The investment k^{**} solves $R = [A' f_k(k^{**}) + (1-\theta_i)(1-\delta)]/\wp_i(R)$ or $R - 1 + \delta = A' f_k(k^{**})$; this is the first best investment when dividends are discounted at R and it can never be optimal to invest more than that. To see this use (19) and note $[A' f_k(k) + (1-\theta_i)(1-\delta)]/\wp_i(R_i') = \mu/(\beta\mu') \geq R = [A' f_k(k^{**}) + (1-\theta_i)(1-\delta)]/\wp_i(R)$, that is, $f_k(k) \geq f_k(k^{**})$. Note that the firm's net worth next period, using (10) and (19), is

$$\begin{aligned} w' &= A'f(k^{**}) + (1-\theta_i)(1-\delta)k^{**} - [Rb' - \theta(1-\delta)k^{**}] - [Rb'_i - (\theta_i - \theta)(1-\delta)k^{**}] \\ &> R\wp_i(R)k^{**} - [Rb' - \theta(1-\delta)k^{**}] - [Rb'_i - (\theta_i - \theta)(1-\delta)k^{**}] = R[k^{**} - b' - b'_i] \\ &= Rw_{ex}. \end{aligned}$$

Note that $d' = 0$, $d'_i = 0$, $k' \leq k^{**}$, and $w' > w_{ex}$, and from (9) next period, $k' = w' + b'' + b''_i$. If $R''_i > R$, then $b''_i = w'_i$ and $b'' = R^{-1}\theta(1-\delta)k'$. Therefore, $\wp k' = w' + w'_i$, but using (9) we have $\wp k^{**} \leq k^{**} - b' = w_{ex} + b'_i < w' + w'_i = \wp k'$, a contradiction. If $R''_i = R$, then $b'' + b''_i = k' - w' < k^{**} - w_{ex} = b' + b'_i$, that is, the firm is paying down debt, and $w'' > w'$ and $w''_i > w'_i$. But then w and w_i grow without bound unless the firm or the intermediary eventually pay a dividend. But since μ and μ_i are strictly increasing as long as $R'_i = R$, if either pays a dividend at some future date, then $\mu < 1$ or $\mu_i < 1$ currently, a contradiction. □

Proof of Proposition 5. First, note that $k^* > 0$ due to the Inada condition and hence $w'^* \geq A'f(k^*) + k^*(1-\theta_i)(1-\delta) > 0$. Moreover, $d^* > 0$ since otherwise the value would be zero which would be dominated by paying out all net worth. Hence, $\mu^* = \mu'^* = 1$. By Proposition 3 intermediary net worth is positive and hence $d_i^* > 0$ (arguing as above), which implies $\mu_i^* = \mu_i'^* = 1$. But then $\eta'^* = (R\beta_i)^{-1} - 1 > 0$ and $l_i'^* > 0$ (and $\eta_i'^* = 0$), since otherwise intermediary net worth would be 0 next period. Therefore, $R_i'^* = \beta_i^{-1}$, and thus $\lambda_i'^* = (\beta_i^{-1}\beta)^{-1} - 1 > 0$, that is, the firm's collateral constraint for intermediated finance binds. Moreover, k^* solves $1 = \beta[A'f_k(k^*) + (1-\theta_i)(1-\delta)]/\wp_i(\beta_i^{-1})$ and d'^*, b'^*, $b_i'^*$, and w'^* are determined by (9)-(12) at equality. Specifically, $d^* = A'f(k^*) + k^*(1-\theta_i)(1-\delta) - \wp_i(\beta_i^{-1})k^* > 0$ and $b_i'^* = \beta_i(\theta_i - \theta)k^*(1-\delta)$. The net worth of the firm after dividends is $w^* = \wp_i(\beta_i^{-1})k^*$. Finally, $l_i'^* = b_i'^* = w_i^*$ and $d_i^* = (\beta_i^{-1} - 1)w_i^*$. □

Proof of Proposition 6. Consider first region D and take $w \geq \bar{w}(w_i)$ (to be defined below) and $d > 0$ forever ($\mu = \mu' = 1$). The investment Euler equation then implies $1 = \beta[A'f_k(k) + (1-\theta_i)(1-\delta)]/\wp_i(R_i)$. If the collateral constraint for intermediated finance (12) does not bind, then $\mu = R'_i\beta\mu'$, that is, $R'_i = \beta^{-1}$, and investment is constant at $\bar{k}$ which solves $1 = \beta[A'f_k(\bar{k}) + (1-\theta_i)(1-\delta)]/\wp_i(\beta^{-1})$ or, equivalently, $1 = \beta[A'f_k(\bar{k}) + (1-\theta)(1-\delta)]/\wp$. Define $\bar{w}(w_i) \equiv \wp\bar{k} - w_i$ and $\bar{w}_i = \beta(\theta_i - \theta)\bar{k}(1-\delta)$. At $\bar{w}_i$, (12) is just binding. For $w_i \in (0, \bar{w}_i)$, (12) is slack. Moreover, $w'_i = \beta^{-1}w_i$ and, if $w'_i \in (0, \bar{w}_i)$, the ex dividend net worth is $w_{ex} = \bar{w}(w_i)$ both in the current and next period, and we have immediately $w'_{ex}/w'_i < w_{ex}/w_i$. Further, using (10) and (19) we have

$$w' = A'f(\bar{k}) + (1-\theta)\bar{k}(1-\delta) - R'_i w_i > [A'f_k(\bar{k}) + (1-\theta)(1-\delta)]\bar{k} - R'_i w_i = R'_i \bar{w}(w_i).$$

But $w'_{ex} = \bar{w}(w'_i) < \bar{w}(w_i)w'_i/w_i = R'_i w_{ex}$, so $d' = w' - w'_{ex} > 0$. For $w_i \in [\bar{w}_i, w^*_i)$, (12) binds and $k(w_i)$ solves $1 = \beta[A'f_k(k(w_i)) + (1-\theta_i)(1-\delta)]/[\wp - w_i/k(w_i)]$ and $R'_i = (\theta_i - \theta)k(w_i)/w_i(1-\delta)$. Note that the last two equations imply that $k(w_i) \geq \bar{k}$, $w_i/k(w_i) \geq \bar{w}_i/\bar{k}$, and $R'_i \leq \beta^{-1}$ in this region. As before, define $\bar{w}(w_i) = \wp k(w_i) - w_i$ and note that the ex dividend net worth is $w_{ex} = \bar{w}(w_i)$. Suppose $w^+_i > w_i$ then $k(w^+_i) > k(w_i)$, $k(w^+_i)/w^+_i < k(w_i)/w_i$, and $w^+_{ex}/w^+_i = \wp k(w^+_i)/w^+_i - 1 < w_{ex}/w_i$. Moreover, $w'_i = R'_i w_i > w_i$ and hence k (strictly) increases and R'_i (strictly) decreases in this region. Proceeding as before,

$$\begin{aligned} w' &= A'f(k(w_i)) + (1-\theta_i)k(w_i)(1-\delta) > [A'f_k(k(w_i)) + (1-\theta_i)(1-\delta)]k(w_i) \\ &\geq R'_i\beta[A'f_k(k(w_i)) + (1-\theta_i)(1-\delta)]k(w_i) = R'_i\bar{w}(w_i). \end{aligned}$$

But $w'_{ex} = \bar{w}(w'_i) < \bar{w}(w_i)w'_i/w_i = R'_i w_{ex}$, so $d' = w' - w'_{ex} > 0$. Finally, if $w_i \geq w^*_i$ and $w \geq \bar{w}(w_i) = w^*$, the steady state of Proposition 5 is reached.

We now show that the above policies are optimal for both the firm and the intermediary given the interest rate process in region D and hence constitute an equilibrium. Since $R'_i > \beta_i^{-1}$ before the steady state is reached, the intermediary lends its entire net worth to the firm, $l'_i = w_i$, and does not pay dividends until the steady state is reached. Hence, the intermediary's policy is optimal. To see that the firm's policy is optimal in region D, suppose that the firm follows the optimal policy from the next period onward but sets $\tilde{d} = 0$ in the current period. If the firm invests the additional amount, then $\tilde{k} = (w_i + w)/\wp > k$ and $\tilde{w}' > w'$ (and therefore $\tilde{\mu}' = 1$). The investment Euler equation requires $1 = \beta/\tilde{\mu}[A'f_k(\tilde{k}) + (1-\theta_i)(1-\delta)]/\wp_i(R'_i)$, but since $f_k(\tilde{k}) < f_k(k)$ and k satisfies the investment Euler equation at $\mu = \mu' = 1$, this implies $\tilde{\mu} < 1$, a contradiction. Suppose the firm instead invests the same amount $\tilde{k} = k$ but borrows less $\tilde{b}'_i < b'_i$. Then $\tilde{w}' > w'$, $\tilde{\mu}' = 1$, and from (19) $\tilde{\mu} = 1$. If $R'_i < \beta^{-1}$, then (12) is binding, a contradiction. If $R'_i = \beta^{-1}$, then the firm is indifferent between paying dividends in the current period or in the next period. But in equilibrium $b'_i = w_i$ and hence $\tilde{d} = d > 0$ for the representative firm. By induction starting at the steady state and working backwards, the firm's policy is optimal in region D. Further, we show in Lemmata 3 and 4 that the equilibrium in region D is the unique equilibrium converging to the steady state.

Consider now region ND with $w_i \leq w^*_i$ (as Lemma 2 shows) and $w < \bar{w}(w_i)$ as defined in the characterization of region D above and $d = 0$. Denote the firm's ex dividend net worth by $w_{ex} \leq w$. There are 3 cases to consider: $w_{ex}/w_i > \bar{w}/\bar{w}_i$, $w_{ex}/w_i \in [w^*/w^*_i, \bar{w}/\bar{w}_i]$, and $w_{ex}/w_i < w^*/w^*_i$.

First, if $w_{ex}/w_i > \bar{w}/\bar{w}_i$, then $w_{ex}+w_i < \bar{w}(w_i)+w_i = \bar{w}+\bar{w}_i$ and $k \leq (w_{ex}+w_i)/\wp < (\bar{w}+\bar{w}_i)/\wp = \bar{k}$. Note that since $b_i' \leq w_i - d_i \leq w_i$, we have $w_{ex}/b_i' \geq w_{ex}/w_i > \bar{w}/\bar{w}_i$. If (12) binds, then $R_i' = (\theta_i-\theta)(1-\delta)(w_{ex}/b_i'+1)/\wp > (\theta_i-\theta)(1-\delta)(\bar{w}/\bar{w}_i+1)/\wp = \beta^{-1}$. If (12) does not bind, then $R_i' = [A'f_k(k)+(1-\theta)(1-\delta)]/\wp > [A'f_k(\bar{k})+(1-\theta)(1-\delta)]/\wp = \beta^{-1}$. In either case, $R_i' > \beta^{-1}$, and hence $d=0$, $d_i=0$, and $b_i' = w_i$.

Second, consider $w_{ex}/w_i \in [w^*/w_i^*, \bar{w}/\bar{w}_i]$. If $w_{ex}/b_i' > \bar{w}/\bar{w}_i$ we are in the first region and hence $d_i = 0$ and $b_i' = w_i$, a contradiction. Hence, w.l.o.g. $w_{ex}/b_i' \in [w^*/w_i^*, \bar{w}/\bar{w}_i]$. Take $\tilde{w}_i$ such that $w_{ex}/b_i' = \bar{w}(\tilde{w}_i)/\tilde{w}_i$. Note that (12) binds at $\tilde{w}_i$ and $\bar{w}(\tilde{w}_i)$, and thus $b_i' + w_{ex} < \tilde{w}_i + \bar{w}(\tilde{w}_i)$ and moreover $k < \hat{k}(\tilde{w}_i)$. If (12) does not bind, then

$$\begin{aligned}\hat{R}_i'(\tilde{w}_i) &= (\theta_i-\theta)(1-\delta)(\bar{w}(\tilde{w}_i)/\tilde{w}_i+1)/\wp > (\theta_i-\theta)(1-\delta)(w_{ex}/b_i'+1)/\wp > R_i' \\ &= [A'f_k(k)+(1-\theta)(1-\delta)]/\wp > [A'f_k(\hat{k}(\tilde{w}))+(1-\theta)(1-\delta)]/\wp.\end{aligned}$$

But since (12) binds at $\tilde{w}_i$ and $\bar{w}(\tilde{w}_i)$, $\hat{R}_i'(\tilde{w}_i) < [A'f_k(\hat{k}(\tilde{w}))+(1-\theta)(1-\delta)]/\wp$, a contradiction. Therefore, (12) binds and $R_i' = \hat{R}_i'(\tilde{w}_i)$. From (19), $\beta\mu'/\mu[A'f_k(k)+(1-\theta_i)(1-\delta)]/\wp_i(R_i') = 1 = \beta[A'f_k(\hat{k}(\tilde{w}_i))+(1-\theta_i)(1-\delta)]/\wp_i(\hat{R}_i'(\tilde{w}_i))$ and, since $k < \hat{k}(\tilde{w}_i)$, $\mu > \mu' \geq 1$, that is, $d=0$. Further, if $w_{ex}/w_i \in (w^*/w_i^*, \bar{w}/\bar{w}_i]$, then $R_i' \in (\beta_i^{-1}, \beta^{-1}]$, and thus $d_i = 0$ and $b_i' = w_i$. If $w_{ex}/w_i = w^*/w_i^*$, then either $d_i > 0$ or $b_i' < w_i$ yields $R_i' > \beta_i^{-1}$ and therefore $d_i = 0$ and $b_i' = w_i$ at such w_{ex} and w_i as well.

Third, consider $w_{ex}/w_i < w^*/w_i^*$. As before, w.l.o.g. $w_{ex}/b_i' < w^*/w_i^*$. Then from (12), $R_i' \leq (\theta_i-\theta)(1-\delta)(w_{ex}/b_i'+1)/\wp < (\theta_i-\theta)(1-\delta)(w^*/w_i^*+1)/\wp = \beta_i^{-1}$, that is, $R_i' < \beta_i^{-1}$. From (19), $\beta\mu'/\mu[A'f_k(k)+(1-\theta_i)(1-\delta)]/\wp_i(R_i') = 1 = \beta[A'f_k(k^*)+(1-\theta_i)(1-\delta)]/\wp_i(\beta_i^{-1})$ and, since $k < k^*$ and $R_i' < \beta_i^{-1}$, $\mu > \mu' \geq 1$, that is, $d=0$. Moreover, (12) binds, since otherwise $\beta_i^{-1} > R_i' = [A'f_k(k)+(1-\theta)(1-\delta)]/\wp > [A'f_k(k^*)+(1-\theta)(1-\delta)]/\wp$, but since in the steady state (12) binds $\beta_i^{-1} < [A'f_k(k^*)+(1-\theta)(1-\delta)]/\wp$, a contradiction.

Thus, we conclude that $d=0$, (property (i) in the statement of the proposition), $d_i = 0$ (except possibly in the first period (see Lemma 2), that R_i' satisfies the equation in property (ii) of the proposition), and that $b_i' = w_i$ and $k = (w+w_i)/\wp$ if $R_i' > R$ and $k = w/\wp_i(R)$ if $R_i' = R$ (property (iii)). Moreover, using (10) and (19) we have

$$\begin{aligned} w' &= A'f(k)+(1-\theta_i)(1-\delta)k - [R_i'b_i' - (\theta_i-\theta)(1-\delta)k] \\ &> R_i'\wp_i(R_i')k - [R_i'b_i' - (\theta_i-\theta)(1-\delta)k] \geq R_i'\wp k - R_i'b_i' = R_i'w,\end{aligned}$$

which, together with the fact that $w_i' = R_i'w_i$, implies that $w'/w_i' > w/w_i$ (property (iv)). Note that the equilibrium is thus unique in region ND as well. □

Proof of Lemma 2. We first show that $d_i > 0$ when $w_i > w_i^*$. If $w \geq w^*$, the stationary state is reached and the result is immediate. Suppose hence that $w < w^*$. Suppose instead that $d_i = 0$. We claim that $R_i' < \beta_i^{-1}$ for such w_i and w. Either $R_i' = R$ and hence the claim is obviously true or $R_i' > R$, but then $b_i' = w_i$. Using (12) and (9) we have $R_i' \leq (\theta_i - \theta)(1-\delta)k/b_i' \leq (\theta_i - \theta)(1-\delta)(w/w_i + 1) < (\theta_i - \theta)(1-\delta)(w^*/w_i^* + 1) = \beta_i^{-1}$, that is, $R \leq R_i' < \beta_i^{-1}$. But as long as $d_i = 0$, $w_i' = R_i' w_i \geq R w_i > w_i$, that is, intermediary net worth keeps rising. If eventually firm net worth exceeds w^*, then the steady state is reached and $\mu_i' = 1$ from then onward. But then $\mu_i = \beta_i R_i' \mu_i' = \beta_i R_i' < 1$, which is not possible. The intermediary must pay a dividend in the first period, because if it pays a dividend at any pont after that, an analogous argument would again imply that $\mu_i < 1$ in the first period, which is not possible. Similarly, if $w < w^*$ forever, then $w > w_i^*$ forever and the firm must eventually pay a dividend in this region, as never paying a dividend cannot be optimal. But by the same argument again then the dividend must be paid in the first period.

To see that at most an initial dividend is paid and no further dividends are paid until the steady state is reached, note that in equilibrium once $R_i' > \beta_i^{-1}$, then this is the case until the steady state is reached. But as long as $R_i' > \beta_i^{-1}$, the intermediary does not pay a dividend (and this is true w.l.o.g. also at a point where $R_i' = \beta_i^{-1}$ before the steady state is reached). Before this region is reached, $R_i' < \beta_i^{-1}$, but then the intermediary would not postpone a dividend in this region, as other wise again $\mu_i = \beta_i R_i' \mu_i' = \beta_i R_i' < 1$, which is not possible. □

Lemma 3 *Consider an equilibrium with $R_i' \in [\beta_i^{-1}, \beta^{-1}]$ and $\mu = \mu' = 1$ and assume the equilibrium is unique from the next period onward. Consider another equilibrium interest rate $\tilde{R}_i'$, then $\tilde{k} \lesseqgtr k$ and $\tilde{R}_i' \lesseqgtr R_i'$ is impossible.*

Proof of Lemma 3. Using (19) at the two different equilibria, if $\tilde{k} \lesseqgtr k$ and $\tilde{R}_i' \lesseqgtr R_i'$, then

$$\frac{\tilde{\mu}}{\tilde{\mu}'} = \beta \frac{A' f_k(\tilde{k}) + (1-\theta_i)(1-\delta)}{\wp_i(\tilde{R}_i')} \gtrless \beta \frac{A' f_k(k) + (1-\theta_i)(1-\delta)}{\wp_i(R_i')} = 1 \tag{23}$$

If $\tilde{k} < k$ and $\tilde{R}_i' < R_i' = \beta_i^{-1}$, then by (23) $\tilde{\mu} > \tilde{\mu}'$. Thus, $\tilde{\mu} > \tilde{\mu}' \tilde{R}_i' \beta_i$ implying that (12) must be binding. But then the firm must pay a dividend and $1 = \tilde{\mu} > \tilde{\mu}'$, a contradiction.

If $\tilde{k} > k$ and $\tilde{R}_i' > R_i'$ and the collateral constraint binds at the original equilibrium, then $\tilde{w}' \geq A' f(\tilde{k}) + (1-\theta)i)(1-\delta)\tilde{k} > A' f(k) + (1-\theta_i)(1-\delta)k = w'$. Since $\tilde{w}' > w'$, $\mu' = 1$, and the equilibrium is unique , $\tilde{\mu}' = 1$. By (23), $\tilde{\mu} < \tilde{\mu}' = 1$, a contradiction.

If $\tilde{k} > k$ and $\tilde{R}_i' > R_i'$ and the collateral constraint does not bind at the original equilibrium, the $R_i' = \beta^{-1}$ (using (16)). But then $\tilde{\mu}/\tilde{\mu}' \geq \tilde{R}_i' \beta > 1$ while (23) implies $\tilde{\mu}/\tilde{\mu}' < 1$, a contradiction. □

Lemma 4 *The equilibrium in region D is the unique equilibrium converging to the steady state.*

Proof of Lemma 4. The proof is by induction. First, note that if $w \geq w^*$ and $w_i \geq w_i^*$, then the unique steady state is reached. Consider an equilibrium interest rate R_i' in region D and suppose the equilibrium is unique from the next period on. Suppose $R_i' \in [\beta_i^{-1}, \beta^{-1})$ and consider another equilibrium with $\tilde{R}_i'$. If the collateral constraint (12) binds at this equilibrium, then $\tilde{R}_i' = (\theta_i - \theta)(1-\delta)\tilde{k}/w_i \gtrless (\theta_i - \theta)(1-\delta)k/w_i = R_i'$, which is impossible by Lemma (3). If the collateral constraint (12) does not bind at this equilibrium and $\tilde{k} < k$, then $\tilde{R}_i' < (\theta_i - \theta)(1-\delta)\tilde{k}/w_i < (\theta_i - \theta)(1-\delta)k/w_i = R_i'$, which is also impossible by Lemma (3). If the collateral constraint (12) does not bind at this equilibrium and $\tilde{k} > k$, by Lemma (3) $\tilde{R}_i' < R_i$. But then by (16) $\tilde{\mu}/\tilde{\mu}' = \beta\tilde{R}_i' < \beta R_i' < 1$. Since $\tilde{k} > k$ and the collateral constraint binds at R_i', $\tilde{w}' > w'$ implying $\tilde{\mu}' = 1$ and by above inequality $\tilde{\mu} < 1$, a contradiction. Thus for $R_i' \in [\beta_i^{-1}, \beta^{-1})$ the equilibrium is unique. Suppose $R_i' = \beta^{-1}$. By Lemma (3), we need only consider the two cases $\tilde{k} \gtrless k$ and $\tilde{R}_i' \lesseqgtr R_i' = \beta^{-1}$. If $\tilde{k} < k$ and $\tilde{R}_i' > \beta^{-1}$, (16) implies that $\tilde{\mu} > 1$ and hence the firm does not pay a dividend. But then the firm must be borrowing less from intermediaries, which cannot be an equilibrium as $l_i' = w_i$ at this interest rate. If $\tilde{k} > k$ and $\tilde{R}_i' < R_i' = \beta^{-1}$, and if (12) binds at $\tilde{R}_i'$, $\tilde{R}_i' = (\theta_i - \theta)(1-\delta)\tilde{k}/w_i > (\theta_i - \theta)(1-\delta)k/w_i \geq R_i'$, a contradiction; if (12) instead does not bind at $\tilde{R}_i'$, $\tilde{\mu}/\tilde{\mu}' = \beta\tilde{R}_i' < 1$. Since $\tilde{k} > k$ and $\tilde{R}_i'\tilde{b}_i' \leq R_i' w_i$, $\tilde{w}' > w'$ implying $\tilde{\mu}' = 1$ and by above inequality $\tilde{\mu} < 1$, a contradiction. Therefore the equilibrium in region D is unique. □

Proof of Proposition 7. Part (i): By assumption the expected productivity in the first period equals the deterministic productivity from time 1 onward (denoted $\bar{A}'$ here), that is, $E[A'] = \bar{A}'$. Define the first best level of capital k_{fb} by $r + \delta = \bar{A}' f_k(k_{fb})$. Using the definition of the user cost of capital the investment Euler equation (19) for the deterministic case can be written as

$$r + \delta + \frac{\rho}{R+\rho}(1-\theta_i)(1-\delta) + \frac{\rho_i}{R+\rho_i}(\theta_i - \theta)(1-\delta) = R\beta\bar{A}' f_k(k^*) < \bar{A}' f_k(k^*)$$

and thus $k^* < k_{fb}$. Now suppose that $\lambda(s') = 0$, $\forall s' \in S$. Part (ii) of Lemma 1 then implies that $\lambda_i(s') = 0$, $\forall s' \in S$, and (15) and (16) simplify to $\mu = R\beta\mu'$ and $\mu = R_i'\beta\mu'$, implying that $R_i' = R$, $\forall s' \in S$, and that $d' = 0$, $\forall s' \in S$, as otherwise $\mu < 1$. Moreover, (6) simplifies to $\mu_i = R\beta_i\mu_i'$ and thus $d_i' = 0$, $\forall s' \in S$, as well since otherwise $\mu_i < 1$. Investment Euler equation (19) reduces to $r + \delta = \bar{A}' f_k(k_{fb})$, that is, investment must be k_{fb}. We now show that this implies that the sum of the net worth of the intermediary and the firm exceeds their steady state (cum dividend) net worth in at least one state, which in turn implies that at least one of them pays a dividend, a contradiction. To see this note that $w' = A'f(k_{fb}) + k_{fb}(1-\delta) - Rb' - R_i'b_i'$ and $w_i' = Rl' + R_i'l_i' \geq R_i'l_i' = R_i'b_i'$ and thus

$$w'+w_i' \geq A'f(k_{fb})+k_{fb}(1-\delta)-Rb' \geq A'f(k_{fb})+(1-\theta)k_{fb}(1-\delta) > A'f(k^*)+(1-\theta)k^*(1-\delta)$$

whereas $w'^* + w_i'^* = \bar{A}'f(k^*) + (1-\theta)k^*(1-\delta)$. For $A' > \bar{A}'$, $w' + w_i' > w'^* + w_i'^*$, and either the intermediary or the firm (or both) must pay a dividend, a contradiction.

Part (ii): If $\lambda_i(s') = 0$, $\forall s' \in S$, then $(\beta\mu'/\mu)^{-1} = R_i' = (\beta_i\mu_i'/\mu_i)^{-1}$ where the first equality uses (16) and the second equality uses (6) and the fact that part (iv) of Lemma 1 holds for an eventually deterministic economy.

Part (iii): Since $\lambda(\hat{s}') = 0$, $\lambda_i(\hat{s}') = 0$ by part (ii) of Lemma 1 and $R_i(\hat{s}') = R$. From (15), $\mu(\hat{s}') = \mu(\check{s}') + \lambda(\check{s}') > \mu(\check{s}')$. Using (6), $(\beta_i\mu_i(\hat{s}')/\mu_i)^{-1} = R \leq R_i(\check{s}') = (\beta_i\mu_i(\check{s}')/\mu_i)^{-1}$ and thus $\mu_i(\hat{s}') \geq \mu_i(\check{s}')$. □

Proof of Proposition 8. The first order conditions are (13)-(16) and $\mu' = 1 + \nu_d'$ where $\beta\nu_d'$ is the multiplier on the constraint $w' \geq 0$. By the Inada condition, (14) implies that $k > 0$ and using (10) at equality and (11) and (12) we have $d' \geq A'f(k)+k(1-\theta_i)(1-\delta) > 0$ and $\mu' = 1$. But (13) and (15) imply $1 \leq \mu = R\beta + R\beta\lambda'$ and thus $\lambda' > 0$ since $R\beta < 1$ by assumption; that is, all firms raise as much financing as possible from households.

Suppose the firm pays dividends at time 0. Then $\mu = \mu' = 1$ and (16) implies $0 > 1 - R_i'\beta = R_i'\beta\lambda_i' - R_i'\beta\nu_i'$ and thus $\nu_i' = 1 - (R_i'\beta)^{-1} > 0$, $b_i' = 0$, and $\lambda_i' = 0$; thus, the firm does not use intermediated finance. Note that the problem of maximizing (22) subject to (9) through (12) has a (weakly) concave objective and a convex constraint set and hence induces a (weakly) concave value function. Thus, μ is (weakly) decreasing in w and let $\bar{w}$ be the lowest value of net worth for which $\mu = 1$; by the Inada condition, such a $\bar{w} < +\infty$ exists. At $\bar{w}$, $d = 0$, $\bar{w} = \bar{k}\wp$ (using (9)), and $\bar{k}$ solves $1 = \beta[A'f_k(\bar{k}) + (1-\theta)(1-\delta)]/\wp$ (using (14)). For $w \geq \bar{w}$, $d = w - \bar{w}$ while the rest of the optimal policy is unchanged.

Suppose $\lambda_i' = 0$ and $\nu_i' = 0$. Then $\mu = R_i'\beta > 1$. Moreover, rearranging (14) we have $1 = \beta/(R_i'\beta)[A'f_k(\underline{k}) + (1-\theta)(1-\delta)]/\wp$ which defines $\underline{k} < \bar{k}$. Define $\underline{w}_u$ such that investment is $\underline{k}$ and $b_i' = 0$; then $\underline{w}_u = \underline{k}\wp$. Similarly, define $\underline{w}_l$ such that investment is $\underline{k}$ and $b_i' = (R_i')^{-1}(\theta_i - \theta)\underline{k}(1-\delta)$; then $\underline{w}_l = \underline{k}[\wp - (R_i')^{-1}(\theta_i - \theta)(1-\delta)]$. Note that $\underline{w}_l < \underline{w}_u < \bar{w}$. So firms below $\underline{w}_l$ raise as much financing as possible from intermediaries (since $\mu > R_i'\beta$ by concavity and hence $\lambda_i' > 0$). Firms with net worth between $\underline{w}_l$ and $\underline{w}_u$ pay down intermediary financing linearly. Firms with net worth above $\underline{w}_u$ do not borrow from intermediaries and scale up until $\bar{k}$ is reached at $\bar{w}$, at which point firms initiate dividends. □

References

Bernanke, Ben, and Mark Gertler, 1989, Agency costs, net worth and business fluctuations, *American Economic Review* 79, 14-31.

Boyd, John H., and Edward C. Prescott, 1986, Financial intermediary coalitions, *Journal of Economic Theory* 38, 211-232.

Brunnermeier, Markus K., and Yuliy Sannikov, 2010, A macroeconomic model with a financial sector, Working paper, Princeton University.

Caballero, Ricardo J., and Arvind Krishnamurthy, 2001, International and domestic collateral constraints in a model of emerging market crises, *Journal of Monetary Economics* 48, 513-548.

Diamond, Douglas W., 1984, Financial intermediation and delegated monitoring, *Review of Economic Studies* 51, 393-414.

Diamond, Douglas W., 2007, Legal systems, bank finance, and debt maturity, Working paper, University of Chicago.

Diamond, Douglas W., and Philip H. Dybvig, 1983, Bank runs, deposit insurance, and liquidity, *Journal of Political Economy* 91, 401-419.

Diamond, Douglas W., and Raghuram G. Rajan, 2000, A theory of bank capital, *Journal of Finance* 55, 2431-2465.

Gertler, Mark, and Nobuhiro Kiyotaki, 2010, Financial intermediation and credit policy in business cycle analysis, Working paper, NYU and Princeton.

Hahn, F.H., 1973, On the foundations of monetary theory, in: M. Parkin and A.R. Nobay, eds., Essays in Modern Economics, Barnes and Noble, New York, NY, 230-242.

Holmström, Bengt, and Jean Tirole, 1997, Financial intermediation, loanable funds, and the real sector, *Quarterly Journal of Economics* 112, 663-91.

Jorgenson, Dale W., 1963, Capital theory and investment behavior, *American Economic Review (Papers and Proceedings)* 53, 247-259.

Kehoe, Timothy J., and David K. Levine, 1993, Debt-constrained asset markets, *Review of Economic Studies* 60, 865-888.

Kehoe, Timothy J., and David K. Levine, 2006, Bankruptcy and collateral in debt constrained markets, Working paper, Federal Reserve Bank of Minneapolis.

Kiyotaki, Nobuhiro, and John Moore, 1997a, Credit cycles, *Journal of Political Economy* 105, 211-248.

Kiyotaki, Nobuhiro, and John Moore, 1997b, Credit chains, Working paper, University of Minnesota and LSE.

Ramakrishnan, Ram T., and Anjan V. Thakor, 1984, Information reliability and a theory of financial intermediation, *Review of Economic Studies* 51, 415-432.

Rampini, Adriano A., and S. Viswanathan, 2007, Collateral, financial intermediation, and the distribution of debt capacity, Working paper, Duke University.

Rampini, Adriano A., and S. Viswanathan, 2010, Collateral, risk management, and the distribution of debt capacity, *Journal of Finance* 65, 2293-2322.

Rampini, Adriano A., and S. Viswanathan, 2012, Collateral and capital structure, Working paper, Duke University.

Townsend, Robert M., 1978, Intermediation with costly bilateral exchange, *Review of Economic Studies* 45, 417-425.

Williamson, Stephen D., 1986, Costly monitoring, financial intermediation, and equilibrium credit rationing, *Journal of Monetary Economics* 18, 159-179.

Figure 1: Capital, Collateral Value, and Financing

This figure shows, on the left, the extent to which one unit of capital can be collateralized by households (fraction θ, interpreted as structures) and intermediaries (fraction θ_i, interpreted to include equipment), in the middle, the collateral value next period after depreciation, and on the right, the maximal amount that households and intermediaries can finance, as well as the minimum amount of internal funds required.

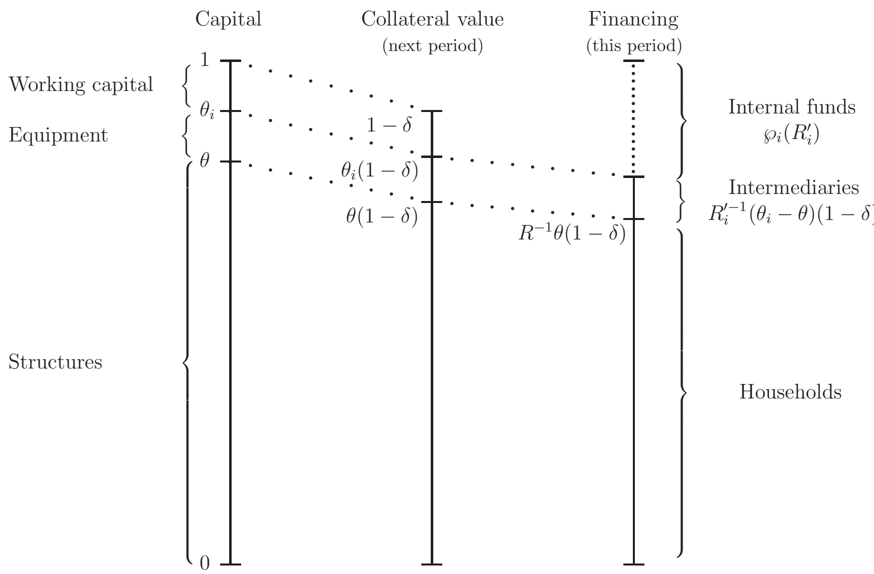


Figure 2: Role of Firm and Financial Intermediary Net Worth

Interest rate on intermediated finance $R_i' - 1$ (percent) as a function of firm (w) and intermediary net worth (w_i). The parameter values are: $\beta = 0.90$, $R = 1.05$, $\beta_i = 0.94$, $\delta = 0.10$, $\theta = 0.80$, $\theta_i = 0.90$, $A' = 0.20$, and $f(k) = k^\alpha$ with $\alpha = 0.333$.

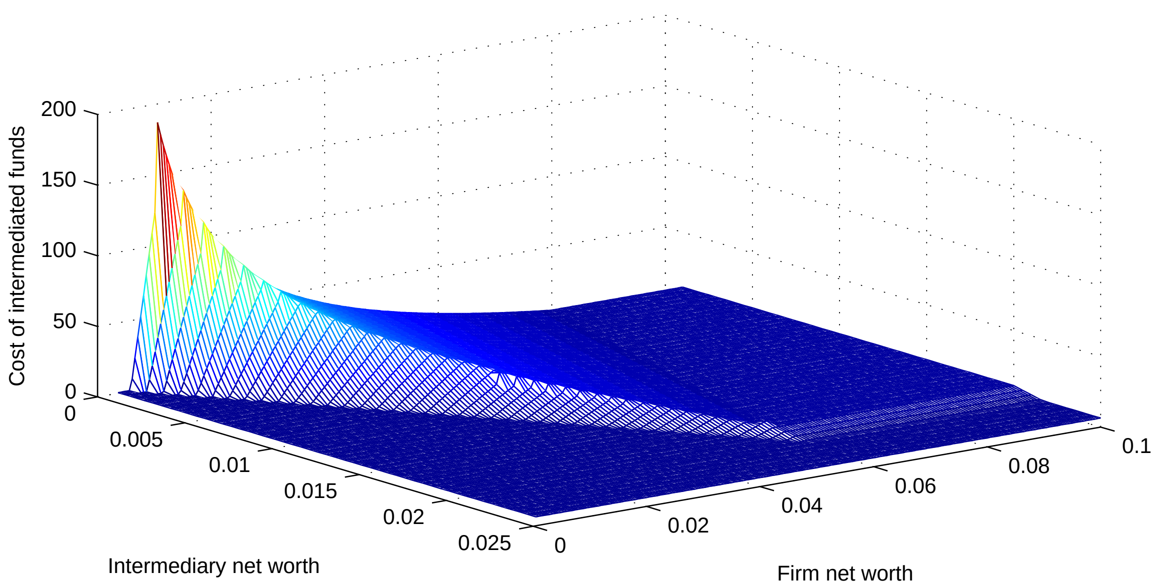


Figure 3: Role of Firm and Financial Intermediary Net Worth

Contour of area where spread exceeds $\beta_i^{-1} - R$: w_i^* (solid) and $\underline{w}(w_i)$ (solid); $\hat{w}(w_i)$ (dashed); contour of area where spread equals $\beta^{-1} - R$: $\bar{w}_i$ (dash dotted) and $\bar{w}(w_i)$ (dash dotted). The parameter values are as in Figure 2.

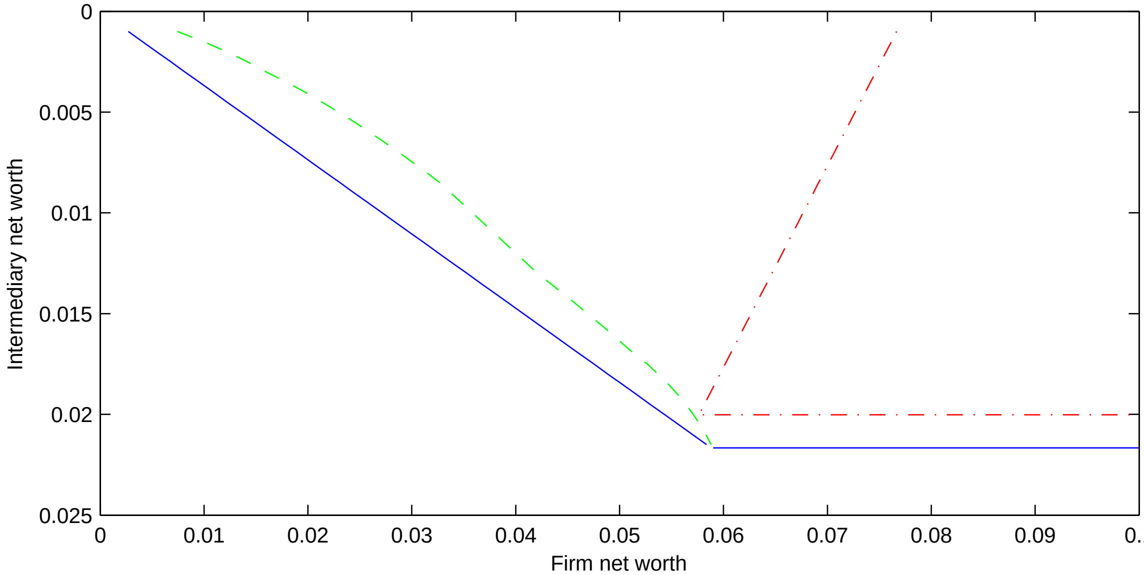


Figure 4: Role of Firm and Financial Intermediary Net Worth

Interest rate on intermediated finance $R_i' - 1$ (percent) as a function of firm (w) for different levels of intermediary net worth (w_i). The parameter values are as in Figure 2.

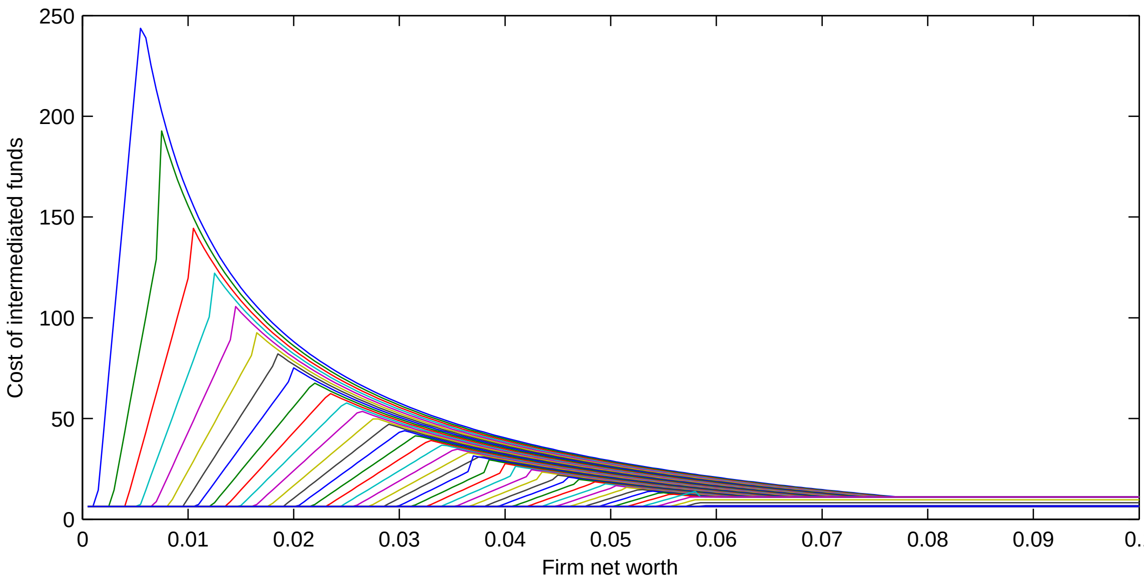


Figure 5: Dynamics of Firm and Financial Intermediary Net Worth

Contours of the regions describing the deterministic dynamics of firm and financial intermediary net worth (see Proposition 6). Region ND, in which firms pay no dividends, is to the left of the solid line and Region D, in which firms pay positive dividends, is to the right of (and including) the solid line. The point where the solid line reaches the dotted line is the deterministic steady state (w^*, w_i^*). The kink in the solid line is the point $(\bar{w}, \bar{w}_i)$ where $R_i' = \beta^{-1}$ and the collateral constraint just binds. The solid line segment between these two points is $\bar{w}(w_i) = \wp k(w_i) - w_i$ (with $R_i' \in (\beta_i^{-1}, \beta^{-1})$). The solid line segment sloping down is $\bar{w}(w_i) = \wp \bar{k} - w_i$ (with $R_i' = \beta^{-1}$). Region ND is dividend by two dash dotted lines: below the dash dotted line through $(\bar{w}, \bar{w}_i)$ $R_i' > \beta^{-1}$; between the two dash dotted lines $R_i' \in (\beta_i^{-1}, \beta^{-1})$; and above the dash dotted line through (w^*, w_i^*) $R_i' < \beta_i^{-1}$. The parameter values are as in Figure 2.

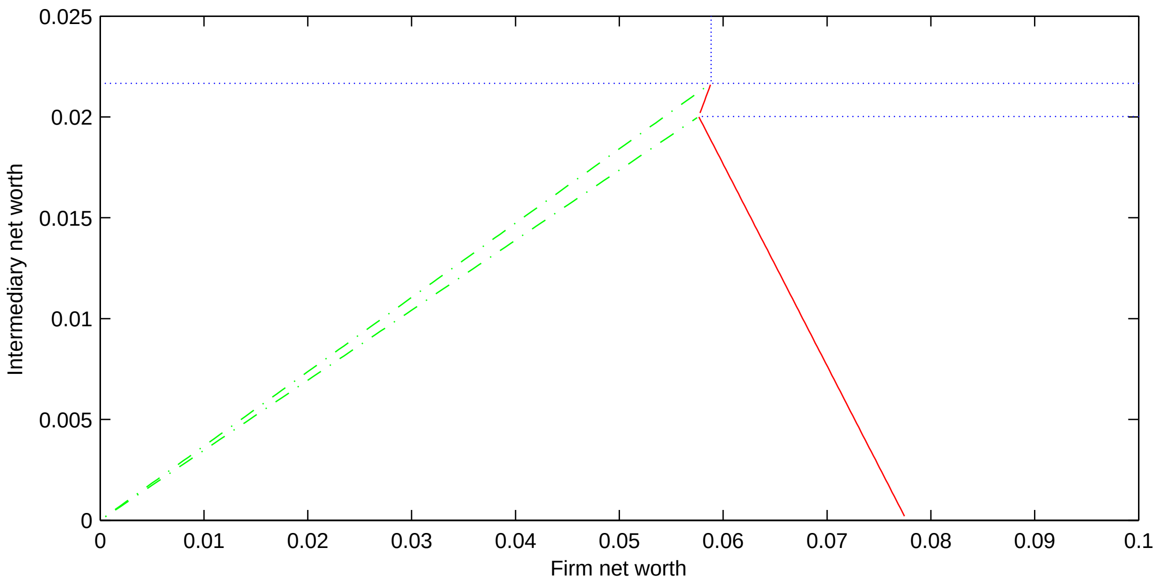
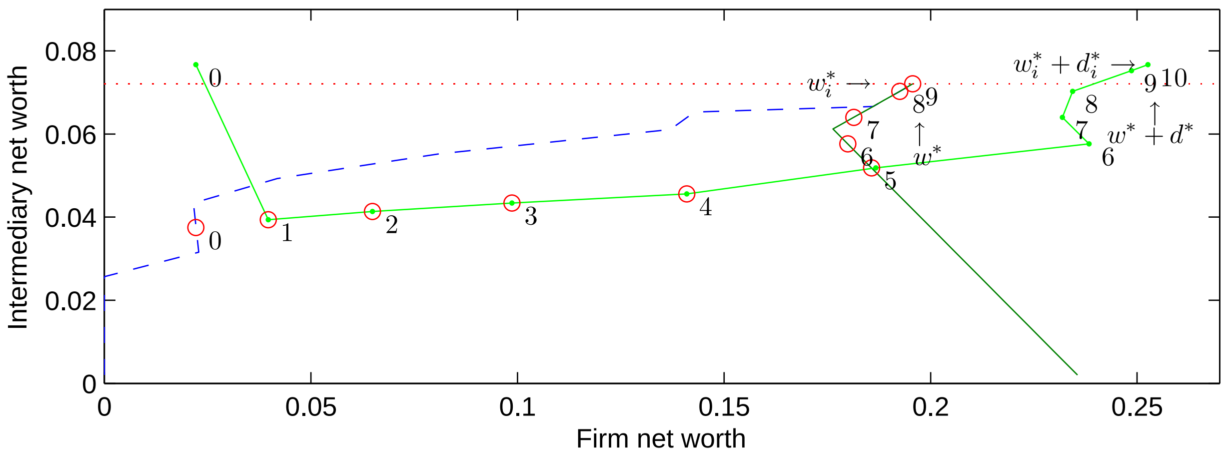


Figure 6: Dynamics of Firm and Financial Intermediary Net Worth

This figures illustrates the deterministic dynamics starting from initial values of net worth $w = 0.0222$ and $w_i = w_i'^*$. Panel A traces out the path of firm and intermediary net worth in w vs. w_i space with the contours as in Figure 5. Panel B shows the evolution of the interest rate on intermediated finance (Panel B1), firm net worth (dashed) and intermediary net worth (solid) (cum dividends (higher) and ex dividend (lower)) (Panel B2), intermediated lending to firms (solid) and households (dashed) (Panel B3), and investment (Panel B4). The parameter values are as in Figure 2 except that $\alpha = 0.8$.

Panel A: Joint evolution of firm and intermediary net worth



Panel B: Interest rates, net worth, lending, and investment over time

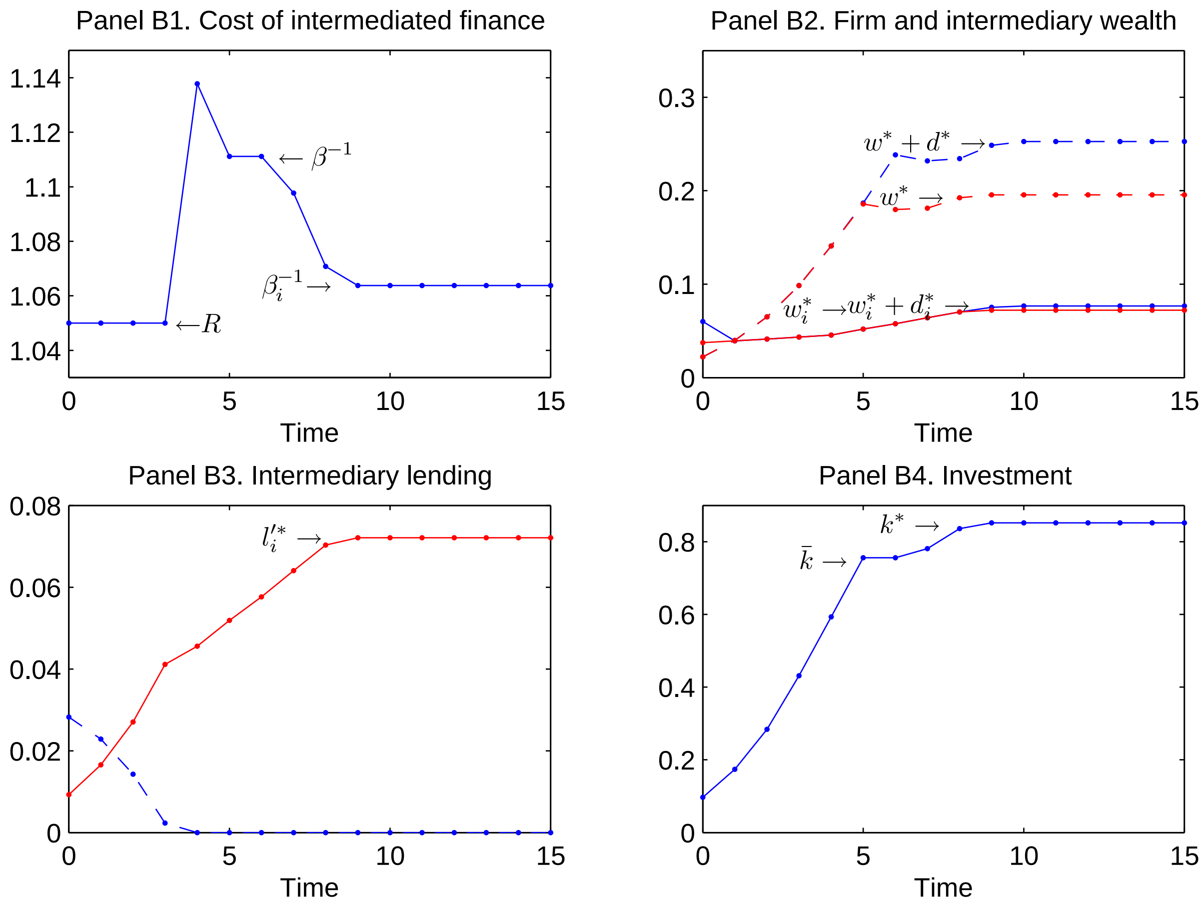
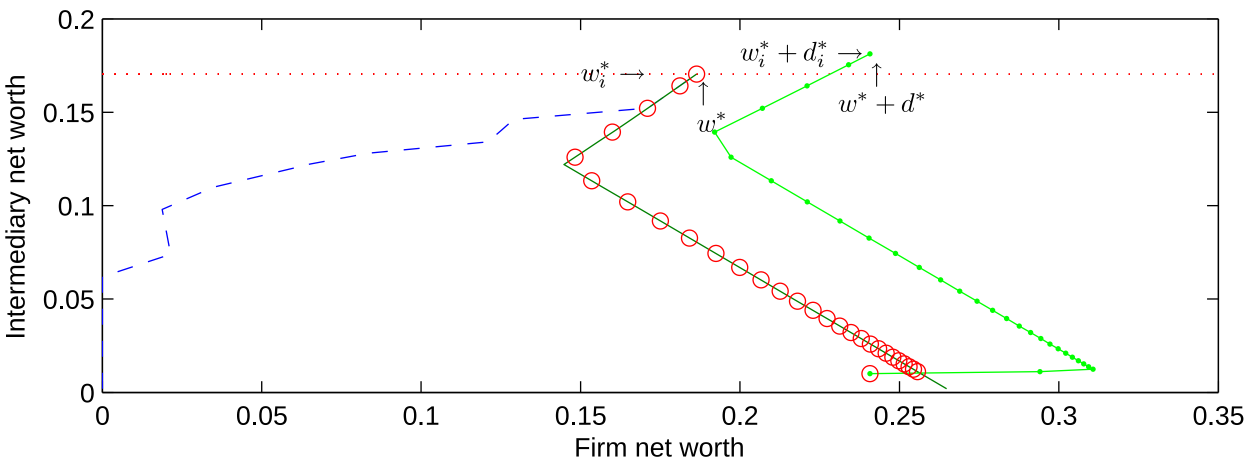


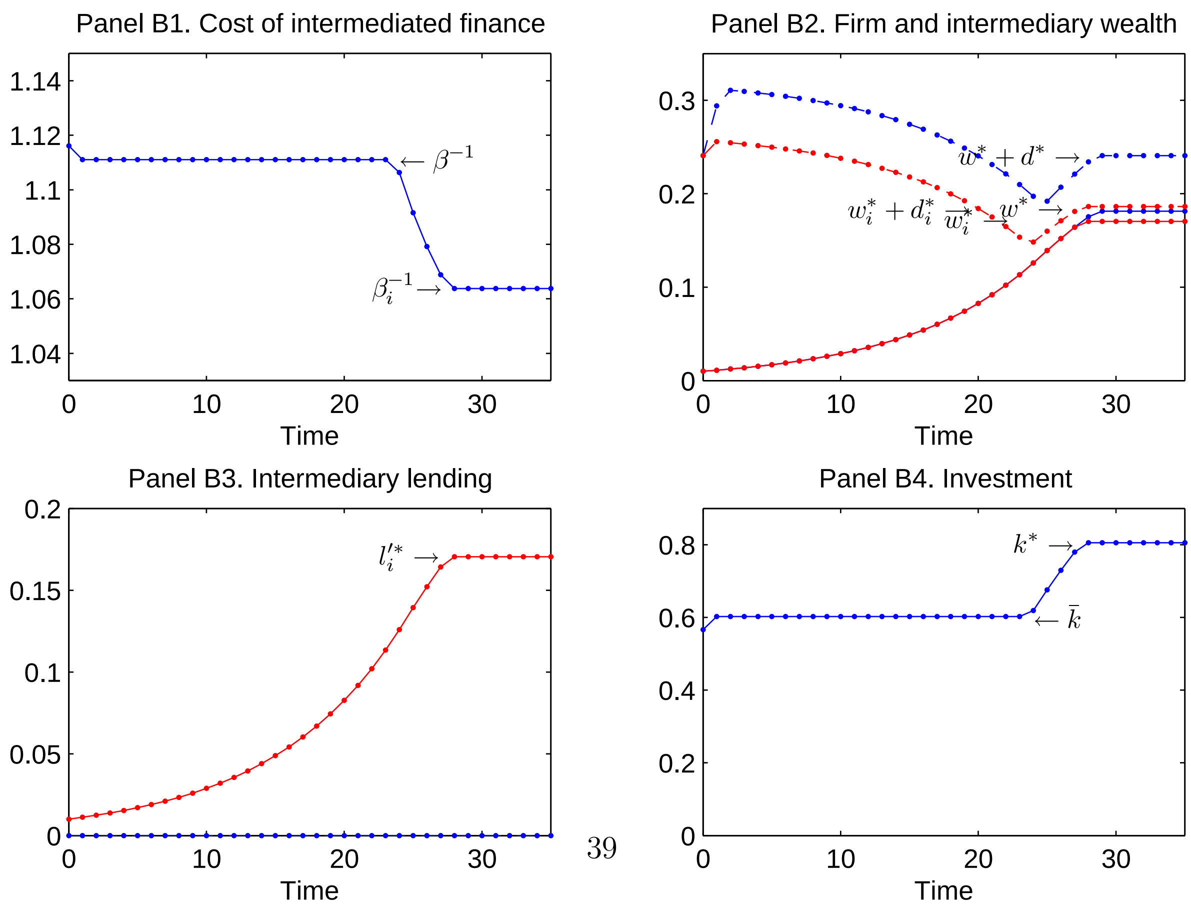
Figure 7: Dynamics of a Credit Crunch

This figures illustrates the deterministic dynamics after a credit crunch starting from initial values of net worth $w = w'^*$ and $w_i = 0.01$. Panel A traces out the path of firm and intermediary net worth in w vs. w_i space with the contours as in Figure 5. Panel B shows the evolution of the interest rate on intermediated finance (Panel B1), firm net worth (dashed) and intermediary net worth (solid) (cum dividends (higher) and ex dividend (lower)) (Panel B2), intermediated lending to firms (solid) and households (dashed) (Panel B3), and investment (Panel B4). The parameter values are as in Figure 6 except that $\theta = 0.65$.

Panel A: Joint evolution of firm and intermediary net worth



Panel B: Interest rates, net worth, lending, and investment over time



FINANCIAL INTEGRATION, HOUSING AND ECONOMIC VOLATILITY*

Elena Loutskina

University of Virginia, Darden School of Business

&

Philip E. Strahan

Boston College, Wharton Financial Institutions Center & NBER

April 2012

Abstract

The Financial Crisis and the Great Recession illustrate the sensitivity of the economy to a housing bust. This paper shows that financial integration, fostered by expansion of nationwide branch networks, amplified housing-price volatility and increased the economy's sensitivity to local housing-price shocks. We exploit variation in credit-supply subsidies across local markets from the Government-Sponsored Enterprises to measure housing price changes unrelated to fundamentals. Using this instrument, we find that a 1% rise in housing prices causes a 0.25% increase in economic growth. This effect is larger in localities more financially integrated with other markets through bank ownership ties. Financial integration thus raised the effect of collateral shocks on the economy, thereby increasing economic volatility.

*We thank seminar participants at Babson College, Boston College, George Washington University, NBER (and our discussant, Antoinette Schoar), University of Virginia-Darden, and Virginia Tech.

I. INTRODUCTION

The recent 'Great Recession', many argue, had its origins in the boom and bust in housing, and the knock-on effects of the resulting financial crisis (Brunnermeier, 2008). Some argue that the length and depth of this recession stems from the slow recovery of housing and the associated debt overhang for consumers (Mian and Sufi, 2011). In this paper, we study links from housing to the overall economy in the years leading up to the crash (1994 to 2006). During this period, local housing prices became more volatile as regions such as the Sun Belt experienced dramatic booms. Figure 1 plots the mean absolute growth shock of local housing prices from 1975 to 2006. Volatility trends down during the 1970s and 1980s. Starting in the 1990s, however, volatility stops falling and then begins to rise. This trend break coincides with changes in the financial and banking systems in the US, which have become increasingly well integrated as deregulation allowed banks to form nationwide branch networks and as securitization allowed mortgage credit to flow easily across markets. We show that shocks to local housing demand were amplified by financial integration because capital could flow freely across connected markets. Financial integration also strengthened the link from housing to the overall economy.

Financial integration may dampen or amplify economic shocks. Morgan, Rime and Strahan (2004) – MRS hereafter – show theoretically that integration's effect on volatility depends on the sources and magnitudes of shocks hitting the local economy. With integration, local economies become more insulated from shocks to the supply of local finance (e.g. local bank capital). During the 1980s and early 1990s, these shocks were a major source of business-cycle instability (Bernanke and Lown, 1991). The number of bank and S&L failures during the 1980s averages more than 150 per year (Kroszner and Strahan, 2008), and the collapse of the

S&L industry amplified downturns in areas such as Texas and California. Integration makes local economies less sensitive to these financial disturbances because capital can flow in from external sources and thus allow investment to continue, even if local lenders are distressed. MRS show empirically that state-level banking integration fostered by deregulation during the 1970s and 80s lowered volatility of local economies in these years.

MRS's theoretical model, however, also shows that integration, by allowing financial capital to flow away from depressed areas and into booming ones, can amplify local cycles. For example, if collateral values rise sharply in a locality, borrower debt capacity and demand for credit increases; integration helps bring financial resources from abroad to satisfy higher credit demand. The influx of credit from external sources raises growth above what would have been possible in a stand-alone, or dis-integrated, financial system. These flows correspondingly reduce collateral values in areas with relatively weak credit demand because these markets face capital outflows. Thus, capital flows generated by credit demand shocks will reduce co-movements in collateral values across financially integrated markets.

Beyond its effects on capital flows, integration is also associated with lower investment by lenders in private information about local business conditions, borrower credit quality and housing-price fundamentals (Loutskina and Strahan, 2011; Romero-Cortes, 2011). As a result of securitization, for example, residential mortgage credit supply responds more now to changes in the market value of collateral than in the past because lenders condition their credit decisions more on public signals (e.g. borrower FICO scores and loan-to-value ratios) and less on private information (Rajan, Seru and Vig, 2010). Both of these forces – more 'flighty' capital and more reliance on public information – may increase collateral volatility and raise the sensitivity of local cycles to variation in collateral values. Consistent with these ideas, we find that financial integration during our sample raises the volatility of housing prices, that shocks in the housing sector have a quantitatively substantial causal impact on local economies, and that the transmission of these housing-price shocks increases with financial integration.

The analysis proceeds in three steps. First, we document a positive relationship between financial integration and the magnitude of local house-price shocks. To do so, we measure financial integration at the level of the Central Business Statistical Area (CBSA), the US Census Bureau's definition of a city. The measure (*In-CBSA ratio)* is based on the ownership of bank branches across CBSAs, equal to the fraction of local deposits owned by a banking company also owning branches in other CBSA markets. So, a CBSA in which all of its branches are owned by banks with branches in other CBSAs would have *In-CBSA ratio* = 100%.

We find that the volatility of shocks to CBSA-level housing price growth increases with financial integration. The effect increases in magnitude when we use variation across states in restrictions on interstate branching as an instrument for financial integration (Rice and Strahan, 2010). Thus, there is a robust difference in local house-price volatility between more- and less-integrated local markets. This result reverses that of MRS, who use data from the 1970s and 1980s, when shocks to the financial sector were an important source of business-cycle variation.[1] Our results, however, are consistent with the theoretical argument that, in the absence of shocks to financial institutions, integration amplifies the impact of collateral shocks. To test this mechanism, we compare shocks for all unique pairs of local markets. If integration increases capital flightiness in response to collateral values shocks, then integration between pairs of markets ought to reduce the correlation between shocks across markets. Using housing price

[1] Like MRS, we have also tested whether the amount of deposits in external markets, as a second integration measure, affects volatility. This second integration measure is also positively related to volatility in some specifications, although its magnitude is smaller and less significant than our primary integration measure.

changes to proxy for such demand shocks, we find this result. Markets that are *more* integrated with each other have *less* similar changes in housing prices, controlling for aggregate shocks (time dummies), for pair-wise CBSA fixed effects and for the similarity of industry composition. Again, we find that the effects increase in magnitude when we instrument for integration using a pair-wise combination of each area's regulatory stance toward interstate branching.[2]

In the second part of the analysis, we build an instrument for house-price appreciation that exploits the importance of the Government-Sponsored Enterprises (GSEs) – Fannie Mae and Freddie Mac – in housing finance. Fannie and Freddie subsidize mortgage credit, but only for mortgages that fall below the jumbo-loan threshold (Loutskina and Strahan, 2009). Borrowers with housing demand near the jumbo-loan threshold stand to benefit from an increase in the threshold, leading to an increase in housing demand and housing prices (Adelino, Schoar and Severino, 2011). While the jumbo-loan cutoff changes uniformly across CBSAs, its effects vary across markets. For example, in Los Angeles - where about 5.3% of mortgages were made to borrowers within 5% of the jumbo-loan cutoff - the change in cut-off would have a bigger impact than in Wichita, Kansas - where this fraction was about 0.5%. Since there is both cross-sectional and time-series variation in the amount of such demand (e.g. LA v. Wichita), we generate a set of instruments based on the product of the sensitivity to changes in the jumbo-loan cutoff in market i during year $t-1$ times the change in the cutoff itself between years $t-1$ and t. The instruments depend only on the distribution of mortgage credit during the preceding year and the change in the jumbo-loan cutoff during the current year, which is the same across all local markets and depends mechanically on lags of increases in nationwide prices. Furthermore, we exploit the elasticity of the housing supply across different geographies to better capture the

[2] Kalemni, Papaionnou and Peydro (2010) find similar effects following financial integration across 20 developed economies.

response of housing prices to changes in demand (Saiz, 2010). Thus, it is plausible to assume that these instruments pick up variation in changes in housing demand exogenous to overall economic fundamentals in the local area.

We find that these instruments are powerful. Local housing prices appreciate faster in markets where credit on jumbo borrowers was more constrained in the prior year, based on the distribution of borrowers around the jumbo cutoff. This effect is stronger in markets with relatively inelastic housing supply because prices are more sensitive to changes in demand where the physical supply of housing is limited by geographic barriers.

Armed with exogenous variation in housing prices, the third part of the analysis shows that housing prices have a strong causal impact on local economic growth in employment and output. In our base model, a 1% increase in housing prices causes an increase in local GDP growth of about 0.25% and an increase in non-construction, non-finance employment growth of about 0.15%. The latter effect implies that higher prices spill over to sectors not directly affected by housing. We then show that the effects of house-price shocks are stronger in local markets with high levels of financial integration than in markets with low integration. In local areas one-standard deviation above the mean level of financial integration, a 1% housing price shock leads to a 0.30% increase in GDP growth. Taken together – higher housing price volatility and increased sensitivity to house-price shocks – the results imply that financial integration has increased economic volatility, both by amplifying variation in collateral values (house prices) and by strengthening links from collateral to the overall economy.

Our paper contributes to three strands of the literature. First, the effect of financial integration on economic volatility has been explored both across US states and also in the context of liberalization of international capital markets (e.g., Morgan, Rime and Strahan (2004), Demyanyk, Ostergaard and Sorenson (2007), Kalemni, Papaionnou and Peydro (2010)). We find that integration can amplify shocks and de-sychronize asset markets in an environment of strong credit demand and a profitable financial sector. In other settings, where financial shocks are important, integration can increase synchronization because credit supply shocks propagate across connected markets (e.g. Peek and Rosengren, 2000). Second, conventional explanations for the US housing boom blame loose lending practices as a key driver of price appreciation (e.g., Mian and Sufi (2009), Keys et al (2010), Demyanyk and Van Hemert (2010), Loutskina and Strahan (2011)). Yet these studies do little to explain why booms were concentrated in places like Florida, Arizona and California. Financial integration can help rationalize large regional booms by allowing capital to flow into areas with strong credit demand.

Third, many have argued that the so-called 'Great Recession' has its root in the crash of housing prices beginning in the middle of 2006. Our results are consistent with this explanation but also suggest that the economic boom was itself fueled by house-price appreciation. The findings extend the work of Mian and Sufi (2009 and 2011), who show that household debt and consumption were strongly correlated with house-price appreciation during the boom. Conversely, declines in consumer spending and financial distress across local markets during the bust are also associated with declines in housing equity. Unlike Mian and Sufi (2011), however, we go a step further and estimate the total effect of housing price shocks on the economy, and we condition this estimate on aspects of the financial system. Shocks to housing have had a large effect on the overall economy, especially in markets that are well integrated nationally.

In the next section we briefly review the forces leading to increased integration over time. In Section III, we describe our integration measures in detail, and document their link to local volatility. Section IV then estimates the relationship between shocks to housing prices and local growth. Here, we first establish a first-stage model that relates changes in credit-supply subsidies from the GSEs to house-price appreciation. We then use this model to generate an instrument for housing price changes to estimate its causal impact on the economy as a whole. Section V concludes.

II. FORCES OF CHANGE LEADING TO FINANCIAL INTEGRATION

Deregulation integrates the banking system

Into the 1970s, most lending occurred through insured depository institutions, and technological, legal and regulatory barriers prevented integration across geographical and product markets. Over time, these barriers have eroded. The process began during the 1970s, when only 12 states allowed unrestricted statewide branching and another 16 prohibited branching entirely. Between 1970 and 1994, 38 states eased their restrictions on in-state branching. States also prohibited ownership of their banks by out-of-state bank holding companies. These barriers to integration began to fall when Maine passed a 1978 law allowing entry by out-of-state BHCs if, in return, banks from Maine were allowed to enter those states. Other states followed suit, and state deregulation of intra-state banking was nearly complete by 1992 (Jayaratne and Strahan, 1996; Jayaratne and Strahan, 1998).

The transition to full *inter*state banking and branching was fostered by passage of the Interstate Banking and Branching Efficiency Act of 1994 (IBBEA), which effectively permitted bank holding companies to enter other states without permission and allowed banks to operate branches across state lines (Rice and Strahan, 2010). With these legal changes, banks now operate across many states and localities, which allows financial resources to flow more easily across geographical markets through banks' internal capital markets (Houston, James and Marcus, 1997).

Despite the passage of IBBEA, states continue to exercise authority under this law to restrict or limit interstate branch entry. While IBBEA opened the door to nationwide branching, it allowed states to influence the manner in which it was implemented. States, for example, had the option to opt into interstate branching immediately after passage of IBBEA or to wait until the default trigger date of June 1, 1997. Moreover, states that opposed entry by out-of-state banks could use provisions of IBBEA to erect barriers to some forms of out-of-state entry, to raise the cost of entry, and to distort the means of entry. IBBEA allowed states to employ various means to erect these barriers. States could set regulations on interstate branching with regard to four important provisions: (1) the minimum age of the target institution, (2) whether or not to permit *de novo* interstate branching, (3) whether or not to permit acquisition of individual branches rather than whole banks, and (4) how tightly to control the percentage of deposits in insured depository institutions controlled by any single bank or bank holding company.

Following Rice and Strahan (2010), we use these four state powers to build a simple index of interstate branching restrictions that exhibits variation both across states and over time. The index equals zero for states that are most open to out-of-state entry. We add one to the index when a state adds any of the four barriers just described. Specifically, we add one to the index: if a state imposes a minimum age on target institutions of interstate acquirers of 3 or more years; if a state does not permit *de novo* interstate branching; if a state does not permit the acquisition of individual branches by an out-of-state bank; and if a state imposes a deposit cap less that 30%. So, the index ranges from zero to four. For most states that adopt branching in 1997 (i.e. states not choosing to opt in early), we set the index at 4 for 1994-1996; in subsequent year we set the index based on each state's policy choices. For example, Illinois adopted interstate branching in 1997 but set a minimum age of 5 years for acquisitions, did not permit *de novo* branching by out-of-state banks, and did not permit single-branch purchases. In 2004, however, Illinois relaxed its policies across each of these three dimensions. Thus, for Illinois we set the branching index at 4 for 1994-1996; we reduce the index to 3 in 1997-2004; and we reduce it further to 0 in 2005-2006.[3] We use this index below as our policy instrument for financial integration.

Securitization integrates housing finance

The move toward integration in mortgage lending was also spurred by the activities of the Government-Sponsored Enterprises (GSEs) - The Federal National Mortgage Association (Fannie Mae) and the Federal Home Loan Mortgage Corporation (Freddie Mac). By the 1990s, both Fannie Mae and Freddie Mac had become heavy buyers of mortgages from all types of lenders, with the aim of holding some of those loans and securitizing the rest. Together they have played the dominant role in fostering the development of the mortgage secondary market. As shown by Frame and White (2005), the GSEs combined market share has grown rapidly since the early 1980s. In 1990 about 25% of the $2.9 trillion in outstanding mortgages were either purchased and held or purchased and securitized by the two major GSEs. By 2003, this market share had increased to 47%.[4] This market share fell after 2004 in the wake of the accounting scandals and the growth of subprime mortgages by private lenders, and then increased significantly since 2006 in response to the credit crisis. GSE access to implicit government

---

[3] Rice and Strahan (2010) report a table detailing each state's policy choices and timing. We do not reproduce that table here for the sake of brevity.

[4] GNMA provides a very important source of mortgage finance to low-income borrowers, holding or securitizing about 10% of all mortgages outstanding.

support allows them to borrow at rates below those available to private banks, and to offer credit guarantees on better terms than competitors without such implicit support.[5]

As shown in Loutskina and Strahan (2010), the GSEs enhance mortgage liquidity, reduce the cost of borrowing, and increase mortgage acceptance rates conditional on borrower credit quality. The GSEs buy and hold some mortgages, and they also often securitize them. When the GSEs buy mortgages, they bear both credit and interest rate risk. When GSEs securitize mortgages, they either buy them and issue mortgage-backed securities (MBS), or they just sell credit protection to the original lender. In the first case, the originating bank retains no stake in the mortgage. In the second case, the bank continues to fund the mortgage and bear the interest rate risk, but obtains the option to sell the mortgage off as an MBS (because of the credit protection). In all cases, the GSEs enhance liquidity and thus foster integration of credit markets.

The GSEs operate under a special charter, however, that limits the size and risk of mortgages that they may purchase or securitize. These limitations were designed to ensure that the GSEs meet the legislative goal of promoting access to mortgage credit for low and moderate-income households. The GSEs may only purchase non-jumbo mortgages, defined in 2006 as those below $417,000 for loans secured by single-family homes. The loan limit increases each year by the percentage change in the national average of single-family housing prices during the prior year, based on a survey of major lenders by the Federal Housing Finance Board. The limit is 50% higher in Alaska and Hawaii. Because the loan limit changes mechanically and only as a function of *national* housing prices, local housing supply or demand conditions have no effect on

[5] Passmore, Sherlund and Burgess (2005) argue that most (but not all) of the benefits of GSE subsidies accrue to their shareholders rather than mortgage borrowers. To take advantage their low borrowing costs, during the 1990s the GSEs increasingly opted to hold, rather than securitize, many of the mortgages that they buy. Policymakers became concerned about the resulting expansion of interest rate risk at the GSEs (Greenspan, 2004), although the 2008 crisis resulted more from the credit guarantees offered by the agencies than from exposure to their retained mortgage portfolio.

the jumbo loan cutoff. We exploit this fact in developing our instrument for housing price growth below.

Both the moves to allow geographical expansion of banks within and across states, as well as the expansion of GSEs and securitization have benefited both lenders and borrowers and fostered capital flows across regions. We do not study the effects of securitization explicitly, however, because its effects are common across all markets and thus absorbed by time effects in our models. Instead, we focus on the extension of ownership connections among banks, which vary both across CBSAs and over time. This dimension of integration matters not only for information-intensive relationship lending to small and medium-sized businesses, but also for credit supplied to segments of the mortgage market where securitization is costly, including jumbo mortgages, second-lien mortgages, bridge loans and non-prime mortgages.

III. FINANCIAL INTEGRATION AND HOUSE-PRICE VOLATILITY

In this section we test how financial integration affects the volatility of local housing prices, and how the synchronicity (or interrelatedness) of housing price changes between market pairs varies with pair-wise measures of financial integration. In our first set of models, we build a panel dataset based on house-price volatility and financial integration at the level of the Central Business Statistical Area (CBSA) over the 1994 to 2006 period (unit of analysis = CBSA-year). In the second set of models, we build a richer panel by creating all CBSA-year pairs, again over the 1994 to 2006 period (unit of analysis = CBSA-pair-year). We test whether the correlation or similarity of housing prices shocks between pairs of markets changes as the two markets become more financially integrated with each other.

To start, we measure the volatility of the housing prices using the absolute deviation of housing price growth in a CBSA-year from the conditional mean, after removing time and CBSA fixed effects. Specifically, we estimate the following regression:

$$Ln\ Housing\ Price_{i,t} - Ln\ Housing\ Price_{i,t-1} = \alpha_t + \gamma_i + growth\text{-}shock_{i,t}\ . \qquad (1)$$

Data for housing price growth rates are constructed from the Federal Housing Finance Association's (FHFA) CBSA-level house price index. The residual $growth\text{-}shock_{i,t}$ captures how much housing price growth rates differ in each CBSA and year compared to average housing price growth in this year across all geographies. The absolute value of this residual reflects housing price fluctuations specific to a given geography:

$$Vol_{i,t}=|growth\text{-}shock_{i,t}|.$$

The CBSA-year regressions test how integration affects housing-price volatility, as follows:

$$Vol_{i,t} = \alpha_t + \gamma_i + \beta^1 Integration_{i,t} + Other\ Controls + \varepsilon_{i,t}\ , \qquad (2)$$

where $Integration_{i,t}$ equals our measures of the extent to which financial activity in a CBSA-year is connected to financial activity in other CBSAs (*In-CBSA ratio*, defined below).

The pair-wise regressions have the following structure:

$$Interrelatedness_{i,j,t} = \alpha_t + \gamma_{i,j} + \beta^2 Integration_{i,j,t} + Other\ Controls + \varepsilon_{i,j,t}\ , \qquad (3a)$$

where $Interrelatedness_{i,j,t}$ equals the negative of the absolute value of the difference in housing-price growth shocks between two CBSAs in a given year:

$$Interrelatedness_{i,j,t} = -\left|\, growth\text{-}shock_{i,t} - growth\text{-}shock_{j,t} \right|. \qquad (3b)$$

So, an increase in $Interrelatedness_{i,j,t}$ measures a decline in the difference in growth shocks between two CBSAs. In Equation (3a), $Integration_{i,j,t}$ measures the pair-wise connectedness of two CBSA markets in a given year (*Common-CBSA ratio*, defined below).

As noted in the introduction, financial integration may raise volatility either because integrated lenders condition their credit decisions more on prices and less on other dimensions of credit risk (e.g. specialized knowledge about the local economy), or because capital flows more easily toward high-demand markets and away from low-demand markets. Both channels imply $\beta^1 > 0$ in Equation (2). By looking at integration's effects on pair-wise markets, we can isolate the capital flows channel. Imagine two CBSA markets – 'A' and 'B' – that are well integrated. A shock to prices in 'A' (and thus to credit demand there) will draw financial resources away from 'B', thus accommodating the credit demand and raising prices in A and lowering them in B. This second capital flight channel thus suggests that financial integration ought to make house-price changes become less correlated as integration between two markets increases, so $\beta^2 < 0$ in equation (3a).[6]

Measuring Financial Integration by CBSA-year

Our measure of financial integration is built from the distribution and ownership of bank branches and deposits across local markets. The measure is based on information on total deposits, location and ownership of all bank branches from the Federal Deposit Insurance Corporation's (FDIC) *Summary of Deposits*, available online annually from 1994 forward.[7] We

[6] House price variation driven by local credit supply shocks will tend to attenuate this effect.

[7] See http://www2.fdic.gov/sod/.

construct the *In-CBSA ratio,* equal to the fraction of all deposits in a CBSA that are owned by a holding company which also owns deposits in one or more other CBSAs.[8]

Variation in the *In-CBSA* $ratio_{i,t}$ depends on bank entry decisions into market *i* in year *t*, which in turn may reflect risk management or diversification motivations of potential entrants. Since the intrinsic volatility of a particular market may play a role in this entry decision, the relationships observed in the fixed effects OLS estimate of Equation (2) could be biased by reverse causality. For example, if out-of-state banks prefer to enter safe markets, the coefficient on financial integration would tend to be biased downward in OLS. To eliminate this potential source of bias, we also estimate Eq. (2) using an instrumental variable model, where the instrument for the *In-CBSA* ratio equals the index of restrictions on interstate branching described in Section II. This index ranges from zero to four, where four represents the highest level of barriers to entry by out-of-state banks.

Measuring Integration by CBSA-year pairs

To measure integration between pairs of CBSAs, we build the *Common CBSA Ratio*. For each CBSA pair, we sum up all deposits with a common ownership link, add these across the two markets, and then divide by the total amount of deposits in the two CBSAs. Higher values of *Common CBSA Ratio* indicate a greater degree of shared financial resources – greater integration – between CBSAs. We also estimate our model with a dummy-variable version of *Common CBSA Ratio,* equal to one when there are any commonly owned deposits and zero otherwise. This second approach is arguably more robust than the first, and its coefficient is also somewhat easier to interpret. As already mentioned, since bank entry decisions may be

[8] We define a banking company as the highest entity within a bank holding company for banks owned by holding companies, or for the bank itself for stand-alone banks.

endogenously driven by economic conditions in local markets, we use an instrument for the *Common CBSA Ratio*, again based on the state-level branching restrictions index. In this case, since each observation represents a pair of CBSAs, the instrument equals the sum of the branching restrictions index in the states where the two CBSAs are located. Hence, we again report both the fixed effects OLS model as well as the IV model.[9]

Table 1 reports summary statistics for our volatility and integration measures. Panel A reports the CBSA-year level means and standard deviations for house-price volatility and the four integration measures; Panel (B) reports these statistics for the two pair-wise interrelatedness measures and the pair-wise integration measure. The *In-CBSA ratio* average 81.4% (Panel A), indicating that in the typical CBSA-year the majority of deposits are owned by banking companies with deposits elsewhere. This variable has substantial variation, with a standard deviation of 15.3%. The average house-price growth shock equals 4.56%, suggesting substantial CBSA-specific shocks to local markets after removing trends in overall housing price appreciation. The pair-wise data tell a similar story, with an average difference in growth residuals between pairs of CBSAs of 4.07 percentage points. Almost 40% of market pairs have some ownership links, with an average *Common CBSA ratio* of 8.28%.

Volatility increases with integration

Table 2 reports our estimation of Equation (2), linking financial integration to total house-price growth volatility, along with the first-stage model for the *In-CBSA Ratio*. All models include time fixed effects to take out aggregate trends as well as the national business cycle. In addition, we control in all models for the share of employment across the following

[9] We include the pair-wise fixed effects even in the IV model. Since the instrument depends on branching in two areas rather than one, a change in the branching index in *either* locality's state generates within-CBSA variation over time. Thus, we get strong identification in the first-stage model, even including the pair-wise fixed effects.

different industry segments: construction, mining and logging; finance; education and health services; manufacturing; trade, transportation and utilities; information technology; professional and business services and other services.[10] In some models, we also incorporate CBSA-level fixed effects to capture time invariant market-level characteristics that may be correlated with volatility. In every case, we cluster data at the CBSA level to build our standard errors.

The results strongly suggest, first, that financial integration is greater in CBSAs located in states with fewer restrictions on interstate branching (Table 2, column 1). An increase in the branching index from 0 to 4 – from least to most restrictive – comes with a decline in the *In-CBSA ratio* of about 5%, which is large relative to the variation in this variable (σ = 15.3% - see Table 1). The branching restrictions index has strong explanatory power in the first stage as well, with a t-statistic above 3. The model without CBSA fixed effects easily passes the Kleibergen and Papp (2006) test for weak instruments, which is designed for models with clustered errors. Since we have weak identification in the model with CBSA effects, we omit the IV model with fixed effects.

Second, financial integration is associated with greater volatility of housing prices (columns 2-4). *In-CBSA ratio* has a positive and significant effect on volatility in OLS without the CBSA effects (column 2) and a slightly smaller coefficient with the CBSA effects (column 3); in both OLS models, however, the economic magnitudes are small. As noted, however, endogenous entry by banks may bias the coefficient on integration downward (that is, toward

[10] The industry share variables are built off the industry employment numbers provided by the Bureau of Labor Statistics. The employment data is provided at detailed industry level. We aggregate the data at the level of 9 different industries: (i) construction, mining and logging; (ii) manufacturing including durable and non-durable goods manufacturing; (iii) trade, transportation, and utilities; (iv) information; (v) financial activities; (vi) professional and business services; (vii) education and health services; (viii) leisure and hospitality; and (ix) Other services. For each industry, we compute the percentage contribution to the CBSA level employment. The employment in the government sector is the omitted variables.

zero), and this notion is supported by the IV model, where the coefficient rises in magnitude substantially. In this model (column 4), a standard deviation increase in the *In-CBSA ratio* would increase house-price growth volatility by 0.4%, a substantial increase relative to the dispersion in house-price volatility (σ = 2.8% - see Table 1).

Interrelatedness across markets falls with integration

Table 3 reports the estimation of Equation (3a), along with the first stage model linking integration between pairs of CBSA markets (*Common CBSA ratio*) to the sum of the branching restrictions index in the two states. In these pair-wise models, the dependent variable equals the negative of the absolute value of the difference in house-price growth shocks in a given year (recall Equation (3b) above). As noted, all of the models include time fixed effects and a separate fixed effect for every unique pair of CBSAs – a total of 65,508 unique fixed effects. These fixed effects remove factors such as geographical distance that may affect the similarity of housing markets between two CBSAs. We also include a variable capturing the 'distance' or similarity of the industry mix between pairs, equal to the sum of squared difference in industry shares (i.e. the Euclidean distance). This pair-wise factor will capture variation over time in the differences in industry mix between markets. We also group our data into clusters for each CBSA to build standard errors. So, although the models are built from nearly one million observations, there are just 362 independent clusters.

Table 3 reports the results for specifications using the continuous measure of integration (*Common CBSA ratio* = the fraction of commonly owned deposits), and using a dummy variable equal to one for markets that have some degree of commonly owned bank deposits. The latter model is somewhat easier to interpret and also may be more robust to outliers. Columns (1) and

(2) report the first stage models, where for both the continuous and dummy variable approaches we have very strong identification (t-stat > 10 for the branching restrictions instrument). For example, increasing the degree to which a CBSA pair are restricted from cross ownership from 0 (most open) to 8 (least open) would come with a 16% increase in the probability that the two CBSAs have some common ownership in deposits (column 2).

Consistent with Table 2, we find that markets that are more integrated with each other have less commonality in growth shocks, and we also find that magnitudes increase when we instrument for integration with branching restrictiveness (columns 5 & 6). For example, the indicator variable model suggests that markets that share bank deposits have house-growth shocks that are 4.4% less similar, which is large relative to the overall variation of these differential shocks (σ = 4.13% - see Table 1). The results support the idea that capital flows affect collateral values. In markets that are financially connected, markets with high credit demand (e.g. high house prices) can draw on financial capital from markets with lower demand, thereby reducing the correlatedness of collateral values between the two markets. In markets that share financial resources, housing price growth rates become *less similar*. This result is strong evidence that financial integration amplifies credit-demand shocks; capital flowing between these markets lowers the similarity in shocks to the value of collateral.

IV. THE EFFECTS OF HOUSING PRICES ON ECONOMIC GROWTH

In this section we ask two questions. First, did the increase in housing-price volatility lead to greater business-cycle instability? Second, did financial integration strengthen the link from housing prices to overall economic performance, thus further raising overall volatility? The first question is motivated by the trend toward greater housing price volatility (recall Figure 1).

The second question is suggested by theories of financial integration, which imply that more mobile financial capital should strengthen the link from shocks to credit demand – e.g. housing prices, or more generally, the value of collateral – and economic output.

To answer these questions, we trace out the causal impact of shocks to housing prices on overall economic output by CBSA-year $(Y_{i,t})$, measured by personal income growth, employment growth, employment growth without sectors directly affected by housing (construction and finance) and GDP growth. Specifically, we estimate panel regressions with the following structure:

$$Y_{i,t} = \alpha^{y}_{t} + \gamma^{y}_{i} + \beta^{y}_{1}\ \textit{House-Price Growth}_{i,t} + \textit{Other Control Variables} + \varepsilon_{i,t} \qquad (4a)$$

and

$$Y_{i,t} = \alpha^{y}_{t} + \gamma^{y}_{i} + \beta^{y}_{1}\ \textit{House-Price Growth}_{i,t} + \beta^{y}_{2}\ \textit{Financial Integration}_{i,t} \qquad (4b)$$
$$+ \beta^{y}_{3}\textit{Financial Integration}_{i,t} * \textit{House-Price Growth}_{i,t} + \textit{Other Control Variables} + \varepsilon_{i,t}\ .$$

We estimate Equations (4a) and (4b) for our CBSA-year panel dataset from 1994 to 2006, including both year and CBSA fixed effects. The year effects remove trends as well as the national business cycles, while the CBSA effects take out long-run differences in average economic growth rates.

To test how financial integration affects links from house price shocks (or, more generally, collateral shocks), we interact *House-Price Growth* with *In-CBSA ratio*, using the branching restrictions index as the instrument for *In-CBSA ratio*, as in Table 2. If changes in housing prices raise borrower debt capacity and, in turn, raises consumer demand and firm investment, then $\beta^{y}_{1} > 0$ (4a)*;* if financial integration, by allowing capital to flow in from external markets, strengthens this effect, then $\beta^{y}{}_{3} > 0$ in (4b). In order to estimate the overall impact of housing on the economy, we first estimate Equation (4a) without financial integration, and then estimate models with the interaction term in (4b).

As additional controls variables, we include the share of employment across industry sectors as before; three measures of the strength and health of the local banking sector: the average capital-asset ratio, the log asset size of banks operating in the CBSA, and the average growth rate of assets of local banks; and, in some specifications, one lag of the dependent variable.[11]

GSE Housing-Finance Subsidies as a Source of Instruments for Housing Price Growth

Shocks to the overall economy will both affect and be affected by the value of housing, as well as the value of real estate and collateral more generally. Our aim is to trace out the causal impact of shocks to housing on the overall economy; hence, we need instruments that move housing prices (and so are sufficiently powerful) but otherwise remain unrelated to fundamental drivers of economic growth (and so meet the exclusion restriction for valid instruments). We use subsidies in housing-finance from the GSEs to build such instruments.[12] Potential home buyers receive a financing subsidy through the activities of the GSEs, who stand ready to buy mortgages that fall below the jumbo-loan cutoff and meet a set of credit-worthiness underwriting criteria. The cut-off is binding on borrowers, as is evident from the histogram of loan applications and loan approval rates presented in Figures 2A and 2B (adapted from

[11] Industry shares are from the Bureau of Labor Statistics. Bank characteristics are taken from the Bank *Call Reports*; CBSA-level averages equal the weighted average of banks operating in the CBSA based on the share of deposits held in a given CBSA by each bank.

[12] Adelino, Schoar and Severino (2011) use a similar strategy at the transaction level to trace out how GSE subsidies affect the price per square foot of housing.

Loutskina and Strahan (2009)). The large spike in loan applications and approval rates just below the jumbo cut-off indicate that the funding is both more abundant and cheaper below the jumbo loan cut-off. The cutoff is the same everywhere (except Alaska and Hawaii), and it increases annually based on a mechanical formula linked to past changes in national housing prices. The increase in the jumbo-loan cutoff thus raises the subsidy to some potential home buyers, but the increase, crucially, is not dependent on conditions in the local area (CBSA).

We exploit the idea that the impact of this increased subsidy varies across local housing markets. For example, in a market where all home prices fall below the jumbo-loan cutoff in *t-1*, home buyers there would receive no incremental benefit from an increase in the cutoff in year *t*; all potential homebuyers would already be subsidized. In contrast, in markets with substantial demand near the jumbo-loan threshold, potential homebuyers would benefit greatly when the cutoff rises.

We use two strategies to measure differences across markets in the impact of changes in the jumbo-loan cutoff on housing demand. Detailed data for all mortgage applications to lenders above $50 million in assets are collected annually under the Home Mortgage Disclosure Act (HMDA). The HMDA data include loan size, whether or not a loan was accepted, some information on borrower credit characteristics, and the location of the property down to the Zip code level. Using these data, we estimate the fraction of loan applications in CBSA *i* and year *t-1* that are above the jumbo cutoff then, but would fall below that cutoff in the subsequent year (year *t)* as a consequence of the increase in the cutoff between the two years. This ratio captures the percentage of borrowers that would benefit from the change in the cut-off though getting access to more readily available and/or cheaper credit.

This first instrument is incomplete because it ignores borrower self-selection into the area just below the cut-off (recall Figure 2A). A large fraction of home buyers reduce their borrowing to fall below the cut-off in year *t-1*, but many would also benefit from an increase in the jumbo-loan cutoff. For example, often home buyers will increase their equity investment in a property to be able to finance their borrowed funds in the subsidized, non-jumbo segment. Others will split their borrowing into a senior (non-jumbo) mortgage to gain the subsidy, and finance the remainder with a second-lien mortgage from a portfolio lender (i.e. a lender who holds the mortgage) plus equity. Thus, many mortgage applicants below – but not too far below – the jumbo-loan cutoff would also benefit from its increase. To capture this portion of demand, we build an instrument equal to the total fraction of applications within 5% of the jumbo-loan cutoff (on either side) in year *t-1*, multiplied by the percentage change in the cutoff between years *t-1* and *t*.

For each instrument, we also add an interaction with a measure of housing-supply elasticity built for 263 CBSAs based on physical impediments to expansion in the housing stock, such as waterways, mountains, etc.[13] Saiz (2010) shows that cities with high supply elasticity have both slower increases in housing prices over time and faster population growth, compared to low-elasticity cities. These results make sense because low barriers to the expansion of housing implies that increased demand from population growth can be accommodated without increasing the cost of housing (e.g. land is not scarce in these areas). In our setting, we expect prices to respond more to the demand shocks associated with changes in the jumbo-loan cutoff in markets with low housing-supply elasticity than in markets with high elasticity.

[13] We use the elasticity estimates available online at: http://real.wharton.upenn.edu/~saiz/ and then convert them to the new definitions of CBSA using the zip-code overlap.

Figure 3 illustrates our identification strategy graphically for two extreme cases: a local market where most of the demand for housing is already subsidized by the GSEs and with very high supply elasticity (e.g. Wichita, where supply elasticity equals 5.5 and only 0.5% of total mortgage applications lie within 5 percentage points of the jumbo-loan cutoff), versus a market with a large mass of demand near the jumbo-loan cutoff and with low supply elasticity (e.g. Los Angeles, where supply elasticity equals 0.63 and about 5.4% of total mortgage applications lie within 5 percentage points of the jumbo-loan cutoff). An increase in the GSE jumbo-loan cutoff shifts housing demand only slightly in Wichita but substantially in Los Angeles. Because supply responds elastically in Wichita, prices barely rise. In LA, however, prices rise sharply, both because demand shifts further from the increased subsidy and because supply responds very little. Thus we trace a shock in a supply of funding to the housing price changes accounting for both CBSA-specific demand shifts and the CBSA-specific supply conditions.

The first-stage model then takes the following form:

$$\begin{aligned} \textit{House-Price Growth}_{i,t} &= \alpha^{HP}_{t} + \gamma^{HP}_{i} + + \textit{Other control variables} + \\ &+ \beta_1^{HP}\textit{Share-New-NJ}_{i,t-1} + \beta_2^{HP}\textit{Share-New-NJ}_{i,t-1} \times \textit{Saiz-Elasticity}_i + \\ &+ \beta_3^{HP}\textit{Share-Near-NJ}_{i,t-1} + \beta_4^{HP}\textit{Share-Near-NJ}_{i,t-1} \times \textit{Saiz-Elasticity}_i + \varepsilon_{i,t}\,, \end{aligned} \tag{5}$$

where $\textit{Share-New-NJ}_{i,t-1}$ equals the fraction of jumbo applications in CBSA *i* and year *t-1* that will fall below the jumbo-loan cutoff next year (year *t*); $\textit{Share-Near-NJ}_{i,t-1}$ equals the share of applications within +/- 5% of the cutoff in year *t-1* times the percentage change in the cutoff between *t* and *t-1*. We expect housing prices to grow fastest in markets with a large mass of demand that would benefit from an increase in the jumbo cutoff; thus, we expect: $\beta_1^{HP} > 0$, *and* $\beta_3^{HP} > 0$. Since house prices should react less if supply is elastic, we expect the interaction terms to offset, meaning $\beta_2^{HP} < 0$, $\beta_4^{HP} < 0$. We estimate Equation (5) with year and CBSA fixed effects, and we cluster the standard errors at the level of the CBSA. (Note that the direct effect of the Saiz elasticity measure, which is constant over time, is absorbed by the CBSA fixed effects.)

Results

Table 4 reports summary statistics for our instruments, for housing price growth and for personal income, employment and GDP growth during the 1994-2006 period. We obtain the CBSA-year level data on employment (and employment by segment) from the Bureau of Labor Statistics; the personal income data from the Bureau of Economic Analysis; and the local geography GDP from Moody's Analytics.[14]

The analysis begins in 1994 because the financial integration data, based on deposits, become available starting in 1994, and because HMDA data become available only in 1992. We end the analysis in 2006 for two reasons. First, we do not want our estimates to be driven by the Financial Crisis and the ensuing Great Recession. Second, our identification strategy relies on the consistent and mechanical increase in the jumbo-loan cutoff over time. This cutoff was raised aggressively in high-priced markets, however, in response to pressure to support housing prices after the Financial Crisis. Moreover, the level of the cutoff has been maintained across other markets even as housing prices have dropped sharply. The instrumental variables are thus both less powerful after 2006 as well as becoming potentially set in response to local conditions.

Table 5 reports the first-stage equation (Eq. (5)) linking the instruments to house-price appreciation, along with the time and CBSA fixed effects, industry share and banking sector

[14] The CBSA-year level GDP estimates are also available from Bureau of Economic Analysis (BEA) but only starting in 2001. We cross-reference the Moody's Analytics data with BEA and find the correlation of 98.7% between two data series.

control variables. We report the models with and without CBSA fixed effects, and for each of these models we report each instrument with its interaction with the Saiz supply-elasticity variable separately and combined with the other instrument. All of the sets of instruments are powerful, with statistically significant effects on both the direct effect and the interaction with Saiz elasticity. Moreover, the signs and magnitudes of the coefficients are economically sensible individually. For example, a standard deviation increase in *Share-Near-NJ* leads to an increase in housing price growth of 2.7% (a little more than one-half of a standard deviation – see Table 4). Each instrument is also more positive in markets with low supply elasticity. Sign patterns are difficult to interpret in the final regressions (columns 3 and 6), with both instruments and interaction terms, because the instruments are highly correlated (ρ=0.92). Finally, the model with all four instruments passes the test for under-identification and weak instruments under CBSA-clustered errors easily (Kleibergen and Paap, 2006).

Table 6 reports IV estimates linking the exogenous component of housing price appreciation to economic outcomes (Equation 4a). We estimate all models with time and CBSA fixed effects and with time-varying industry share variables (coefficient suppressed), and time-varying measures of banking system characteristics (coefficients suppressed). Table 6 reports a total of eight specifications - with and without the lagged dependent variable, times four different measures of output: personal income growth (columns 1 & 2), total employment growth (columns 3 & 4), the growth of total employment excluding employment in financial firms and construction (columns 5 & 6), and GDP growth (columns 7 & 8). By including the lag of the dependent variable, we can alleviate the concern that the instruments, which depend on last year's distribution of home buyers, pick up conditions in the local economy from the prior year. Employment without construction and finance allows us to test whether any effects that we observe spillover beyond segments not directly tied to housing finance. Panel A of Table 6 uses all four identifying instruments: *Share New Non-Jumbo, Share New Non-Jumbo*Saiz Elasticity, Share Near Non-Jumbo*Change in Cutoff, Share Near Non-Jumbo*Change in Cutoff*Saiz Elasticity*; Panel B uses just two identifying instruments: *Share New Non-Jumbo*Saiz Elasticity* and *Share Near Non-Jumbo*Change in Cutoff*Saiz Elasticity*.

The coefficient estimates are statistically and economically significant across all specifications, ranging from 0.14 to 0.26. An exogenous 1% increase in housing prices (stemming from a credit supply increase) thus causes the local economy to expand by 0.14 to 0.26 percentage points faster than otherwise. The coefficients on total employment growth are smaller than GDP growth, which makes sense because GDP includes all sources of production from local sources (i.e. it includes returns to capital as well as labor).[15] Moreover, the coefficient on employment growth without segments directly tied to housing suggests that spillovers from higher collateral values raise output beyond the housing sector. Coefficients on personal income growth tend to be somewhat smaller because some of the variation depends on sources of income not tied specifically to the local area. Comparing Panels A and B, we find similar estimates regardless of the choice of instruments.[16]

Table 7 reports our last test, where we introduce an interaction between housing price growth and financial integration (Equation 4b). For this model, we add the branching restrictions index and its interaction with all of the other instruments and model *Housing price growth, In-*

[15] We have also estimated these models separately for the early (1994-2000) and late (2001-2006) portions of our sample. We find that housing is positively and statistically significantly related to economic outcomes in both samples, with somewhat larger magnitudes in the first half of the sample.

[16] We have explored other ways to build instruments to check for model robustness. For example, we have estimated models in which we eliminate the time-variation in the share near non-jumbo by using its average value at the beginning of our sample. These results lead to somewhat larger coefficients on the house-price growth variable with a higher level of statistical significance than those reported in Table 6.

CBSA ratio and their interaction as jointly endogenous.[17] Since three endogenous variables makes identification more challenging compared to the models in Table 6, we use all of our instruments to maximize power.

The results suggest that house price shocks have a greater impact on economic outcomes in financially integrated markets. Across all four specifications, housing price growth and financial integration are jointly significant at better than 1%. Moreover, the interaction term suggests that better integration has an economically important effect on the size of the causal impact of housing prices on economic output. For example, at the mean of the *In-CBSA ratio* (0.81), a 1% increase in housing prices would generate an increase in GDP growth of 0.15% (0.15 = -0.70+0.81*1.044); in markets one-standard deviation *above* the mean level of integration (0.81+0.15), the same 1% housing-price shock would lead to an increase of 0.30% (0.30 = -0.70+0.96*1.044). The interaction effect of integration on housing is statistically significant across all four models, with a magnitude that varies from 1.0 to 1.4. Because credit supply can respond more elastically to increases in collateral values when local markets are better integrated, an increase in housing prices generates a larger positive spillover in integrated markets. In these areas, the higher demand for credit can draw financial resources in from other sectors.

V. CONCLUSION

The Financial Crisis and subsequent Great Recession of 2007-2011 have emphasized for everyone the importance of a strong housing market to the economy. Housing prices not only increased sharply during the 2000s, but they also became more volatile across local markets. In

[17] The branching index will also help identify housing growth, as Favara and Imbs (2010) show that housing prices grew faster in states more open to interstate banking due to greater availability of credit.

fact, a first-order characteristic of the housing boom was its concentrated effects on Sun Belt areas like Florida, California, Nevada and Arizona. We show that this regional volatility can be explained in part by better financial integration. We then demonstrate a causal link from housing to the overall economy, using variation in the impact of credit-supply subsidies from the GSEs to construct an instrument for housing price changes that is unrelated to economic conditions in the local economy. Our estimates suggest that a 1% rise in housing prices increase growth by about 0.25%. This effect is larger in localities that are better integrated with other markets through bank ownership ties. The results suggest that financial integration raises the effect of collateral shocks on the economy, thereby increasing economic volatility.

REFERENCES

Ashcraft, A. B. and T. Schuermann, 2008, "Understanding the securitization of sub-prime mortgage credit," http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1071189&rec=1&srcabs=1093137.

Adelino, M, A. Schoar and F. Severino, 2011, "Credit Supply and House Price: Evidence from Mortgage Market Segmentation," mimeo.

Brunnermeier, M. K., 2008, "Deciphering the 2007-08 liquidity and credit crisis," *Journal of Economic Perspectives*.

Demyanyk, Y. and C. Ostergaard and B. Sorenson, 2007, "US Banking Deregulation, Small Business and Interstate Insurance of Personal Income, *Journal of Finance*.

Demyanyk, Y. and O. Van Hemert, 2010, "Understanding the subprime mortgage crisis," *Review of Financial Studies*.

Demsetz, R. S., and P. E. Strahan, 1997, "Diversification, size, and risk at bank holding companies," *Journal of Money, Credit, and Banking* 29, 300-313.

Favara, G. and J. Imbs, 2010, "Credit Supply and the Price of Housing," http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1706711.

Frame, W. S. and L. J. White, 2005, "Fussing and fuming about Fannie and Freddie: How much smoke, how much fire?" *Journal of Economic Perspectives*, Spring 2005.

Greenspan, A., 2004, Testimony before the Committee on Banking, Housing, and Urban Affairs, U.S. Senate, February 24, 2004.

Houston, Joel, Christopher James, and David Marcus, 1997, "Capital Market Frictions and the Role of Internal Capital Markets in Banking," *Journal of Financial Economics* 46, 135-164.

Jayaratne, J., and P. E. Strahan, 1996, "The finance-growth nexus: Evidence from bank branch deregulation," *Quarterly Journal of Economics* 101, 639-70.

Jayaratne, Jith and Philip E. Strahan, 1998, "Entry Restrictions, Industry Evolution and Dynamic Efficiency: Evidence from Commercial Banking," *Journal of Law and Economics* 41(1), 239-74.

Kalemli-Ozcan, S., Papaionnou E. and Peydro J., 2010, "Financial regulation, financial globablization and the synchronization of economic activity," Working Paper no. 1221, ECB.

Keys, B. J., Mukerjee, T. K., Seru, A. and V. Vig, 2010, "Did securitization lead to lax screening? Evidence from sub-prime loans," *Quarterly Journal of Economics*, 125(1), 2010.

Kleibergen, F., Paap, R., 2006, "Generalized reduced rank tests using the singular value decomposition," *Journal of Econometrics.*

Kroszner, R. S. and P. E. Strahan, 2008, "Regulation and deregulation of the U.S. banking industry: Causes, consequences and implications for the future," forthcoming in *Economic Regulation and its Reform: What Have We Learned?* Edited by Nancy Rose, NBER.

Loutskina, E., 2010, "Does securitization affect bank lending? Evidence from bank responses to funding shocks," *Journal of Financial Economics,* forthcoming.

Loutskina, E., and P. E. Strahan, 2009, "Securitization and the declining impact of bank financial condition on loan supply: Evidence from mortgage originations," *Journal of Finance* 64(2), 861-922.

Loutskina, E., and P. E. Strahan, 2011, "Informed and Uninformed Investment in Housing: The downside of diversification," *Review of Financial Studies* 24(5), 1447-80.

Mian A. and A. Sufi, 2009, "The consequences of mortgage credit expansion: Evidence from the 2007 mortgage crisis," *Quarterly Journal of Economics*, November 2009, 124(4), 1449-1496.

Mian A. and A. Sufi, 2011, "House Prices, Home Equity Based Borrowing and the US Household Leverage Crisis," *American Economic Review* 101, 2132-56.

Morgan, D., Rime, B. and Strahan, P.E., 2004, "Bank Integration and State Business Cycles," Quarterly Journal of Economics 119(4), 1555-85.

Passmore, W., S. M. Sherlund, and G. Burgess, 2005, "The effect of housing Government-Sponsored Enterprises on mortgage rates," *Real Estate Economics* 33(3), 427-63.

Peek, J. and E. Rosengren, 2000, "Collateral Damage: The effect of the Japanese Banking Crisis on Real Activity in the United States," *American Economic Review* 90(1), 34-45

Rajan, U., A. Seru and V. Vig, 2010, The failure of models that predict failure: Distance, incentives and default," http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1296982

Rice, T. and P. E. Strahan, 2010, "Does Credit Competition Affect Small-Firm Finance? 2010, *Journal of Finance* 65(3).

Romero-Cortes, K., 2011, ""Did Local Lenders Forecast the Bust? Evidence from the Real Estate Market," Working paper, Boston College.

Saiz, Albert, 2010, "The Geographic Determinants of House Prices," *Quarterly Journal of Economics* 110, 379-406.

Stiroh, K. and P.E. Strahan, 2003, "Competitive Dynamics of Competition: Evidence from US Banking," *Journal of Money, Credit and Banking* 35(5), 801-28.

Stock J. H. and M. Yogo, 2005, "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, edited by D. W. K. Andrew and J. Stock, Cambridge University Press, 80-108.

Figure 1: Volatility of the Housing Prices

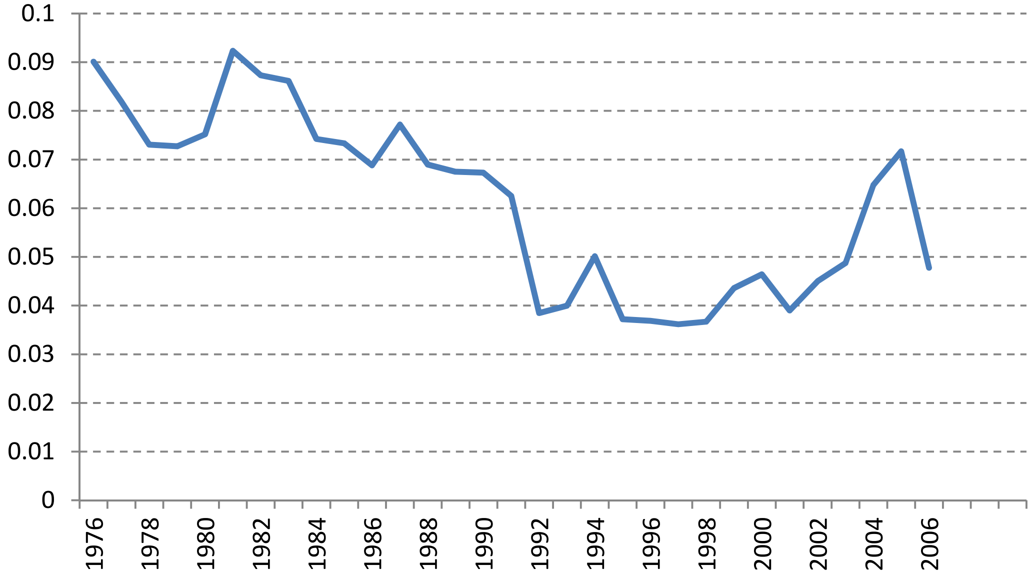


Figure 2A: Histogram of Loan Applications 1994-2006.

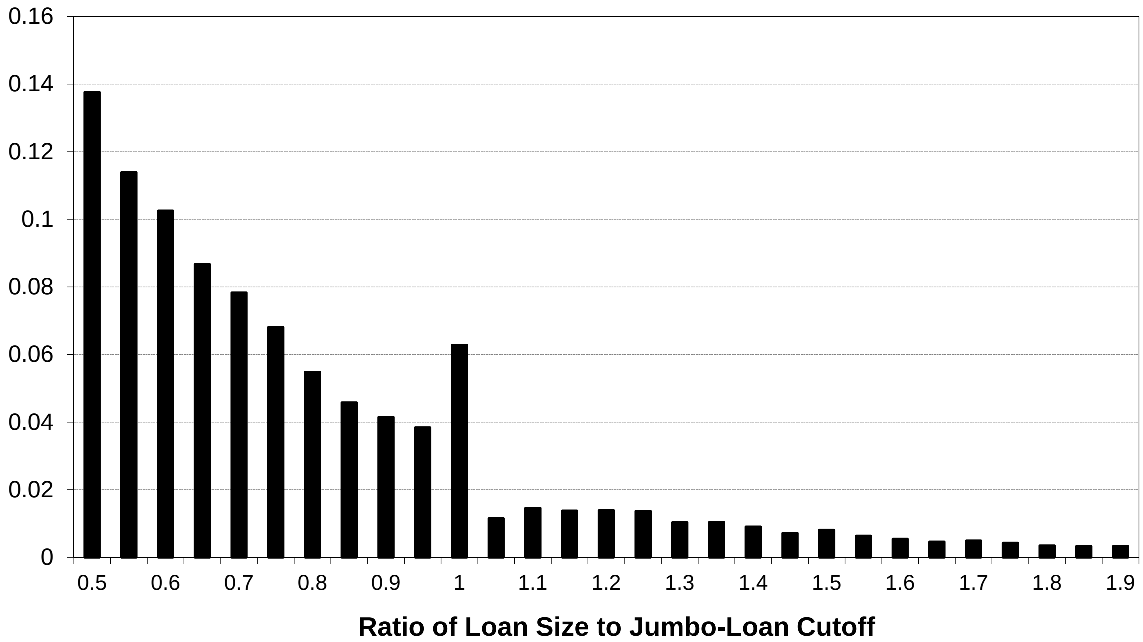


Figure 2B: Share of Approved Loan Applications 1994-2006.

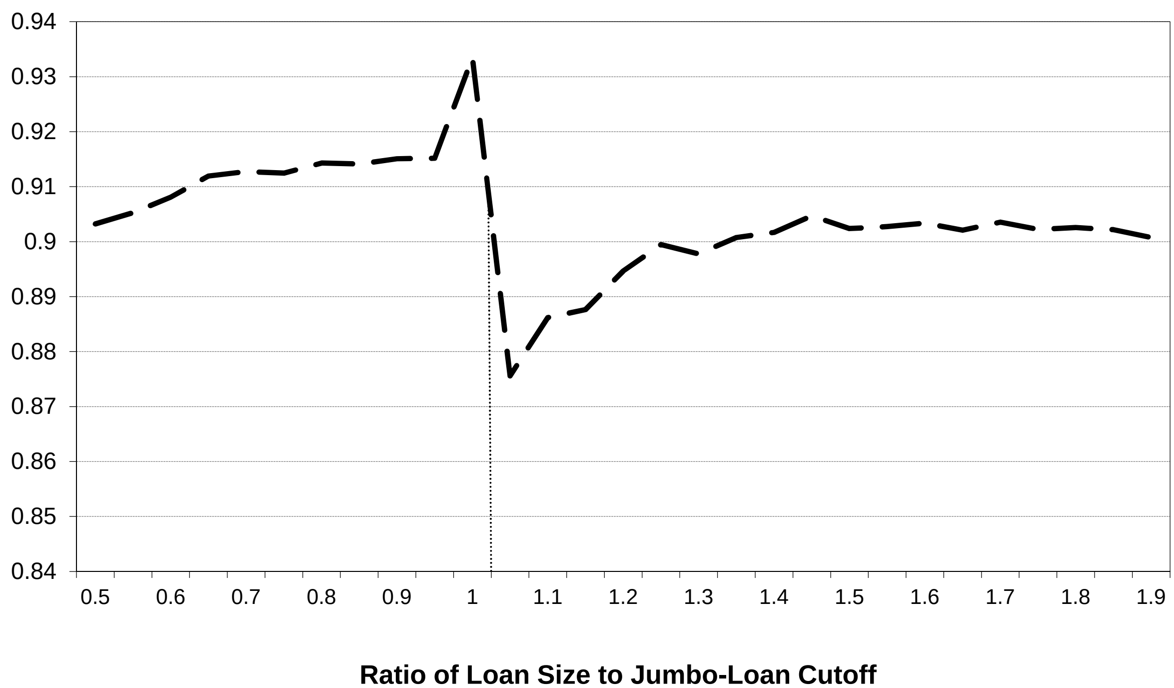
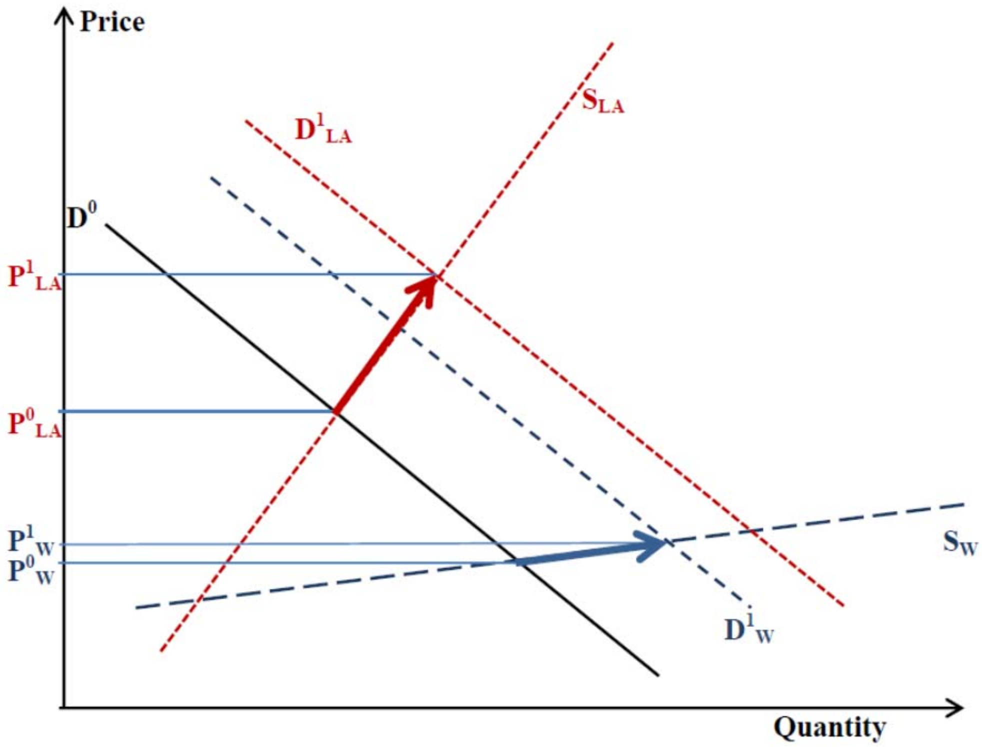


Figure 3: Responses of Different Markets to Changes in GSE Loan Cut-off



The graph illustrates the responses of two hypothetical markets to changes in the GSE loan cut-off. The subscript LA represents Los-Angeles CA and subscript W represents Wichita KS. Two markets are characterized by different elasticity of housing supply (S_{LA} and S_W) as well as different shifts in the demand curves caused by the same change in the loan cut-off (D^1_{LA} and D^1_W). The graph illustrates the corresponding changes in the housing prices.

Table 1: Summary Statistics for Measures of Integration and Housing Price Growth

Panel A reports summary statistics for two measures of financial integration that vary across CBSA-years. The In-CBSA ratio equals the fraction of deposits in CBSA-year that are owned by banking companies with deposits in other CBSAs; the Out-CBSA ratio equals the total deposits in other CBSA owned by banking companies opering in the CBSA divded by total CBSA deposits. Panel B reports summary statistics at the level of CBSA-pair-years, where the measure of integration equals the sum of deposits with common ownership in a pair of CBSAs divided by total deposits in the two CBSAs.

Panel A: CBSA-Year Panel	***Mean***	***StDev***
In-CBSA Ratio	81.4%	15.3%
Housing Price Growth	5.05%	4.55%
Absolute Value of Housing Price Growth Residual	4.56%	2.77%
Panel B: CBSA-Pair-Year Panel		
% of shared deposits	8.28%	14.38%
% of shared deposits when positive	22.32%	16.03%
Indicator for CBSA pair with positve shared deposits	36.38%	N/A
- Absolute Value of Differential Growth Shock	-4.07%	4.13%

Table 2: Housing Price Volatility and Financial Integration

This table reports regressions of housing price volatility on measures of financial integration. The dependent variable is constructed as follows: first, we regress housing price growth on a CBSA fixed effect and year fixed effect and save the residual. We use the absolute value of this growth residual as teh dependent variable. Each model includes time effects. We report the OLS models with and without CBSA level fixed effects. The IV model is only well identified without the CBSA fixed effects. Standard errors are clustered by CBSA.

Dependent Variable:	In-CBSA Ratio	Absolute Value of Residual House-Price Growth		
	First-Stage	OLS		IV
	(1)	(2)	(3)	(4)
Branch Restriction Index	-0.0133***	-	-	-
	(3.02)	-	-	-
In-CBSA Ratio	-	0.00832**	0.00554	0.0307**
	-	(2.48)	(0.63)	(2.18)
Share of employment in construction, mining and logging	0.503**	-0.0199	-0.275***	-0.0298
	(2.27)	(-0.859)	(-3.857)	(-1.164)
Share of employment in financial sector	-0.898**	0.0735**	0.575***	0.0941**
	(2.22)	(2.30)	(3.46)	(2.34)
Share of employment in education and health services	-0.181	0.0319***	0.169*	0.0351***
	(1.45)	(3.70)	(1.94)	(3.68)
Share of employment in manufacturing	0.0544	0.0135**	-0.128*	0.0123*
	(0.52)	(2.03)	(-1.747)	(1.67)
Share of employment in trade, transportation, and utilities	0.0721	-0.00244	-0.0116	-0.00254
	(0.43)	(-0.255)	(-0.170)	(-0.239)
Share of employment in information	-0.164	-0.0098	-0.303	0.00295
	(0.21)	(-0.176)	(-1.216)	(0.05)
Share of employment in professional and business services	0.624***	-0.0236	-0.0725	-0.0387
	(3.27)	(-1.349)	(-0.821)	(-1.539)
Share of employment in leisure and hospitality	0.425***	-0.0143	0.272**	-0.0226
	(2.85)	(-1.157)	(2.42)	(-1.395)
Share of employment in other services	0.272	-0.0613	-0.372	-0.0731
	(0.50)	(-1.437)	(-1.540)	(-1.520)
Sum of Squared employment shares	-0.0381	0.00802	-0.0302	0.00713
	(0.13)	(0.47)	(-0.213)	(0.37)
Time Effects	yes	yes	yes	yes
CBSA Effects	no	no	yes	no
Number of Observations	4,397	4,397	4,397	4,397
R^2	10.0%	14.6%	26.9%	13.4%

*** p<0.01, ** p<0.05, * p<0.1

Table 3: Housing Price Interrelatedness Between Market Pairs and Financial Integration

This table reports regressions of the negative of the absolute value of the difference in housing price shocks between pairs of CBSA markets on measures of financial integration between the two market pairs. The dependent variable is constructed as follows: first, we regress housing price growth on a CBSA fixed effect and year fixed effect and save the residual. We use the absolute value of this growth residual as the growth shock in market i, year t. Each model includes time effects and CBSA-pair fixed effects. Standard errors are clustered by CBSA.

Dependent Variable:	Interrelatedness	Interelatedness Indicator	- Absolute Value of Differential Growth Shock			
	First-Stage		*OLS*		*IV*	
	(1)	(2)	(3)	(4)	(5)	(6)
Branch Restriction Index	-0.00432***	-0.0195***	-	-	-	-
	(10.41)	(10.65)	-	-	-	-
Interelatedness	-	-	-0.0245***	-	-0.200***	-
	-	-	(8.17)	-	(4.92)	-
Interelatedness Indicator	-	-	-	-0.00260***	-	-0.0442***
	-	-	-	(4.07)	-	(4.61)
Distance between Employment Shares	-0.00635	-0.0295	-0.0144**	-0.0143**	-0.0147**	-0.0147**
	(0.54)	(0.57)	(2.10)	(2.08)	(2.17)	(2.15)
Time Effects	yes	yes	yes	yes	yes	yes
CBSA-Pair Fixed Effects	yes	yes	yes	yes	yes	yes
Number of Observations	707,256	707,256	707,256	707,256	707,256	707,256
R^2	18.2%	20.2%	23.0%	23.0%	16.0%	14.0%

*** p<0.01, ** p<0.05, * p<0.1

Table 4: Summary Statistics for Economic Growth, Housing Price Growth and Instrument for Housing Price Growth

This table reports summary statistics for housing price growth, four measures of local economic growth, and two instruments built reflecting the distribution of mortgage credit around the jumbo-mortgage cutoff.

	Mean	***StDev***
Housing Price Growth	5.41%	4.63%
Personal Income Growth	5.21%	2.55%
Employment Growth	1.46%	2.39%
Employment Growth, without construction and finance	1.14%	2.62%
CBSA level GDP growth	5.39%	3.04%
Share of New Non-Jumbo borrowers	0.357%	0.788%
Share Near the Jumbo Cutoff * Change in Cutoff	0.092%	0.145%
Saiz Measure of Housing Supply Elasticity	2.595	1.422

Table 5: Regressions relating Housing Price Growth to Distribution of Mortgage Credit around the Jumbo-Loan Cutoff

This table reports regressions of housing price growth by CBSA-Year on the share of borrowers in year t-1 that will become non-jumbo in year t (share new non-jumbo), and the total fraction of borrowers within +/- 5% of the jumbo-loan cutoff in year t-1 times the change in the jumbo loan cutoff between t-1 and t. All regressions include time and CBSA fixed effects, along with measures of industry structure and the health of the local banking system. Column 6 includes all instruments and acts at the first-stage for the subequent IV models (Tables 6 and 7). Standard errors are clustered by CBSA.

Dependent Variable:	*Housing Price Growth*					
	(1)	(2)	(3)	(4)	(5)	(6)
Share of New Non-Jumbo borrowers	0.25	-	-3.374***	0.168**	-	-2.003***
	(1.11)	-	(6.31)	(2.08)	-	(4.30)
Share of New Non-Jumbo borrowers	-0.209**	-	0.845**	-0.243***	-	0.401
* Saiz Elasticity of housing supply	(2.02)	-	(2.55)	(2.77)	-	(1.22)
Share Near the Jumbo Cutoff * Change in Cutoff	-	4.687***	22.91***	-	1.835**	5.376**
	-	(3.97)	(7.48)	-	(1.97)	(2.62)
Share Near the Jumbo Cutoff * Change in Cutoff	-	-2.013**	-6.594***	-	-1.032***	-3.907*
* Saiz Elasticity of housing supply	-	(2.05)	(3.46)	-	(2.73)	(1.84)
Saiz Elasticity of housing supply	-0.00447***	-0.00342***	-0.00225***	-	-	-
	(4.09)	(3.47)	(2.64)	-	-	-
Time fixed effects	yes	yes	yes	yes	yes	yes
Industry structure	yes	yes	yes	yes	yes	yes
Banking Sector Controls	yes	yes	yes	yes	yes	yes
CBSA dummmies	no	no	no	yes	yes	yes
Observations	2,783	2,783	2,783	2,783	2,783	2,783
R-squared	0.316	0.322	0.347	0.524	0.516	0.525

*** p<0.01, ** p<0.05, * p<0.1

Table 6: IV Regressions relating Local Economic Growth to Housing Price Growth

This table reports IV regressions of economic growth on housing price growth by CBSA-Year; first stage results appear in Table 5. All regressions include time and CBSA fixed effects, along with measures of industry structure and the health of the local banking system. Standard errors are clustered by CBSA.

Panel A: This panel uses *share new non-jumbo* , *share new non-jumbo * Saiz elasticity* , *share new NJ * change in cutoff* , and *share new NJ * Saiz elasticity* as identifying instruments.

	Personal Income Growth		*Total Employment Growth*		*Employment Growth w/o Construction or Finance*		*GDP Growth*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
House-Price Growth	0.186***	0.137***	0.222***	0.209***	0.168***	0.152***	0.259***	0.245***
	(4.25)	(3.52)	(5.83)	(5.76)	(5.12)	(4.77)	(4.66)	(4.39)
Lagged Dependent variable	-	(0.00)	-	-0.121**	-	-0.159***	-	0.0784*
	-	(0.05)	-	(2.53)	-	(2.92)	-	(1.90)
Time fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Industry structure	yes	yes	yes	yes	yes	yes	yes	yes
Banking Sector Controls	yes	yes	yes	yes	yes	yes	yes	yes
CBSA dummmies	yes	yes	yes	yes	yes	yes	yes	yes
Observations	2,783	2,783	2,783	2,783	2,783	2,783	2,783	2,783
R-squared	0.384	0.392	0.235	0.254	0.298	0.32	0.156	0.165

Panel B: This panel uses *share new non-jumbo * Saiz elasticity* and *share new NJ * Saiz elasticity* as identifying instruments.

	Personal Income Growth		*Total Employment Growth*		*Employment Growth w/o*		*GDP Growth*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
House-Price Growth	0.211***	0.176***	0.227***	0.225***	0.157***	0.149***	0.240***	0.234***
	(4.50)	(3.83)	(5.38)	(5.37)	(3.94)	(3.66)	(4.07)	(3.98)
Lagged Dependent variable	-	(0.02)	-	-0.127***	-	-0.158***	-	0.0804*
	-	(0.54)	-	(2.73)	-	(2.96)	-	(1.92)
Time fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Industry structure	yes	yes	yes	yes	yes	yes	yes	yes
Banking Sector Controls	yes	yes	yes	yes	yes	yes	yes	yes
CBSA dummmies	yes	yes	yes	yes	yes	yes	yes	yes
Observations	2,783	2,783	2,783	2,783	2,783	2,783	2,783	2,783
R-squared	0.376	0.387	0.232	0.245	0.302	0.321	0.162	0.169

*** p<0.01, ** p<0.05, * p<0.1

Table 7: IV Regressions relating Local Economic Growth to Housing Price Growth , with Financial Integration Interaction

This table reports IV regressions of economic growth on housing price growth, financial integration (In CBSA ratio) and their interaction, by CBSA-Year. All three of these are treated as endogenous variables, with instruments from Table 5 plus the branching restrictions index. All regressions include time and CBSA fixed effects, along with measures of industry structure and the health of the local banking system. Standard errors are clustered by CBSA.

	Personal Income Growth	*Total Employment Growth*	*Employment growth w/o Construction or Finance*	*GDP Growth*
	(1)	(2)	(3)	(4)
House-Price Growth	-0.74	-1.10	-0.82	-0.70
	(0.59)	(0.44)	(0.65)	(0.35)
House-Price Growth *In CBSA Ratio	1.014*	1.426**	1.055*	1.044*
	(1.75)	(2.12)	(1.77)	(1.69)
In CBSA Ratio	0.06	0.13	0.157*	0.212*
	(0.99)	(1.53)	(1.75)	(1.76)
Time fixed effects	yes	yes	yes	yes
Industry structure	yes	yes	yes	yes
Banking Sector Controls	yes	yes	yes	yes
CBSA dummmies	yes	yes	yes	yes
Chi^2-test for joint sig. of three endogenous variables	19.69	22.86	12.28	18.25
Observations	2,783	2,783	2,783	2,783
R-squared	0.547	0.553	0.426	0.44

*** p<0.01, ** p<0.05, * p<0.1

What Fuels the Boom Drives the Bust: Regulation and the Mortgage Crisis

Jihad C. Dagher * **Ning Fu** †‡

December 3, 2011

Abstract We show that the lightly regulated non-bank mortgage originators contributed disproportionately to the recent boom-bust housing cycle. Using comprehensive data on mortgage originations, which we aggregate at the county level, we first establish that the market share of these independent non-bank lenders increased in virtually all US counties during the boom. We then exploit the heterogeneity in the market share of independent lenders across counties as of 2005 and show that higher market participation by these lenders is associated with increased foreclosure filing rates at the onset of the housing downturn. We carefully control for counties' economic, demographic, and housing market characteristics using both parametric and semi-nonparametric methods. We show that this relation between the pre-crisis market share of independents and the rise in foreclosure is more pronounced in less regulated states. The macroeconomic consequences of our findings are significant: we show that the market share of these lenders as of 2005 is also a strong predictor of the severity of the housing downturn and subsequent rise in unemployment. Overall our findings lend support to the view that more stringent regulation could have averted some of the volatility on the housing market during the recent boom-bust episode.

Key words: regulation, independent mortgage companies, foreclosures, county performance, credit supply, matching estimator.
JEL Classification Numbers: G01, G21, E51

*International Monetary Fund. Email: jdagher@IMF.org
†Harvard University. Email: Ning_Fu@hks13.harvard.edu
‡We thank George Akerlof, Nathaniel Arnold, Stijn Claessens, Enrica Detragiache, Giovanni Favara, Benjamin Keys, Luis Marques, Anna Scherbina, and seminar participants at the IMF Research Department and IMF Institute for useful comments. All views expressed are those of the authors alone and do not represent the views of the IMF.

1 Introduction

The global financial crisis had its roots in the U.S. housing market. Following a period of unprecedented boom in mortgage lending, the U.S. housing market entered a downturn phase during 2006, a year that saw a sharp increase in mortgage delinquency. These problems later spilled into the financial sector by weakening the balance sheets of financial institutions. The far reaching consequences of this housing bust have prompted a growing body of research that seeks to gain a better understanding of the drivers of this housing cycle.

There is now substantial evidence that the unprecedented housing boom was fueled by deteriorating lending standards which led to a worsening in the risk profile of the marginal borrower (Dell'Ariccia et al, 2008; Mian and Sufi, 2009a; Purnanandam, 2010). This evident deterioration in lending standards has led to widespread calls for changes in the regulatory and supervisory systems under which mortgage lenders operate. That enhanced regulation and supervision could have averted bad lending remains, however, a theoretical premise with little empirical work to validate such link. Nevertheless, the Dodd-Frank Act which led to the most significant overhaul of the United States financial regulatory system since the Great Depression was at least partially motivated by that premise.

In this paper we show that the less regulated mortgage lenders contributed disproportionally to the boom in mortgage originations and that their lending was associated with a sharper increase in foreclosures.

Depending on their status, mortgage lenders in the U.S. operated, prior to the crisis, under different regulatory structures with differing degrees of oversight particularly between banks and non-bank mortgage originators. Banks were more regulated under federal banking laws and especially more tightly supervised by federal agencies (see e.g. Belsky and Richardson; 2010). They are subject to a range of federal examinations such as fair lending, the Community Reinvestment Act (CRA), and safety and soundness assessment. They must comply with CRA provisions such as reporting requirements and merger review. Depository institutions that are insured by the Federal Deposit Insurance Corporation (FDIC) must also in addition meet a minimum risk-based capital and reserve requirements. Federal agencies were also required to regularly examine the compliance of the banks they regulate with applicable laws related to their mortgage lending such as the CRA, Truth In Lending Act (TILA), and fair lending laws (see e.g. Immergluck, 2009). Independent non-bank mortgage lenders (henceforth independents), on the other hand, escaped most of these federal regulations and were instead lightly regulated and supervised at the state level (see e.g. Belskey and Retsinas, 2008; Treasury Blueprint, 2008; Immergluck, 2009).[1] A major trade organization representing these independents lenders, the Mortgage Bankers Association, has also called for establishing a federal regulator to develop a uniform national mortgage standards and regulate independent mortgage lenders (see Belsky and Richardson, 2010).

Using comprehensive data on mortgage originations we distinguish between these two types of lenders and first show that the mortgage boom was to an important extent fueled by an expansion of independents. While independent lenders accounted for around one-third of mortgage lending in 2003, they contributed to more than 60% of the increase in mortgage lending between 2003 and 2005. We show that this expansion of independents was more pronounced in areas experiencing higher growth in house prices, a variable that we instrument for using housing supply elasticity (see e.g. Mian and Sufi, 2009b).

We then exploit the heterogeneity in the market share of independents across counties and show that their presence is a strong predictor of the rise in foreclosures.[2] This relation holds after controlling for economic and demographic differences between counties. We

[1] Treasury Blueprint (p81): "Treasury recommends subjecting participants in the mortgage origination process that are not employees of federally regulated depository institutions (or their subsidiaries) to uniform minimum licensing standards. [footnote: Federally regulated mortgage lenders and their employees are subject to an extensive scheme of federal supervision of their lending practices and compliance with applicable laws and regulation]". Immergluck (p66): "Banks and thrifts are subject to regular examination for compliance with not just CRA but also fair lending laws and the Truth in Lending Act. Mortgage companies have generally not been subject to routine examination for compliance with any of these laws on a regular basis. Federal regulatory have large cadres of well-trained examiners to conduct these regular examinations. Meanwhile, mortgage companies are typically regulated by state mortgage banking agencies in the states in which they conduct business. Suffice it to say that, in most states, the capacity of state mortgage regulators is generally not as great as that of the federal regulatory agencies".

[2] By "market presence" we refer to the extent of the market share of a lender, i.e., the percentage of loan volume originated by the lender, and not to its physical location or the location of its branches.

also control for measures of credit and house price growth during the boom and find that the market share of independents remains a significant predictor of foreclosures. The recent literature on the mortgage crisis underlined the role of the increased reliance on an originate-to-distribute model, or in other words, the rise in securitization rates, in the deterioration of lending standards (see, e.g., Keys et al, 2009; Purnanandam, 2010). While independents securitized a significantly larger share of their originations we find that the market share of independents explains to a great extent the relation between the securitization share and the rise in foreclosures, and not the other way around. These results suggest that the type of lender, alone, is an important determinant of mortgage defaults. We focus our empirical exercise on the early rise in foreclosures prior to the liquidity crunch and thus minimize the possibility that our results be contaminated by these factors (see e.g., Ivashina and Scharfstein, 2010). We ensure that the relation between the market share of independent and the rise in foreclosures is not captured by changes in the house prices by instrumenting for the latter. In fact, the early in rise in foreclosures preceded the fall in house prices. We interpret these findings as a strong indication that the expansion of independents came at the cost of fast deteriorating lending standards. This interpretation is compatible with the findings from the recent literature that suggest that the expansion in mortgage credit was to a large extent fueled by the willingness of lenders to extend credit to a riskier category of borrowers (see e.g. Dell'Ariccia et al, 2008; Mian and Sufi, 2009a).

The housing downturn was characterized by a significant contraction in mortgage credit and in house prices, and a subsequent increase in unemployment starting in 2008 which is one of the hallmarks of the Great Recession. We examine these variables as useful measures of the severity of the crisis on the regional level. We show that our key variable, the market share of independents as of 2005, is also a strong predictor of the contraction in credit and house prices, and the rise in unemployment.

A salient feature of our methodological approach is the use of matching techniques to supplement the traditional parametric regression analysis. We use these semi-nonparametric methods to ensure better control for the covariates thus minimizing the impact of possible confounding factors. These methods also help us ensure that our results are not dependent on a linear specification. A standard approach in the matching literature is to compare the mean of the dependent variable between a treatment sample and a matched control sample. We follow this approach and use the Abadie-Imbens bias-adjusted matching estimator (see Abadie and Imbens, 2002). In addition to this step, we repeat our linear regressions on the subsample of matched counties, hence effectively using the matching as a nonparametric preprocessing of the data (see e.g. Ho et al, 2010). In the benchmark exercise, we match U.S. counties with no restriction on the state, but we also show results from intra-state matching which lead to similar findings.

Compared to banks, a conspicuous characteristic of independents is their lack of regulation and supervision. It is thus natural to attribute, with some confidence, differences between the outcomes of their lending to their heterogeneous regulation and supervisory structure, as in Keys et al (2009). We nevertheless pursue and test several alternative hypotheses. More specifically, we test whether our findings could be captured by either differences in mortgage lender competition across counties, or by the geographical diversification of lenders, and we find that none of these factors can capture the effect of independents on foreclosures. In the benchmark regressions we only control for one measure of securitization, specifically the share of private securitization defined as in Mian and Sufi (2009a). As robustness, we also use more comprehensive measures by including other forms of securitization and find that this does not affect our results. One might argue that an important difference between independents and banks is that the latter are depository institutions. We therefore exploit the heterogeneity in the ratio of core deposits to assets across banks by merging HMDA data with data on banks and thrifts' balance sheets to construct and control for a weighted measure of the core deposit ratio of lenders in a county, and show that our results remain robust.

To further explore the regulation argument, we examine whether the relation between the share of independents and foreclosures is more severe in less regulated states. The premise is that any state regulation that constrains risky lending is likely to have a more important impact on the lending standards of the otherwise less constrained lenders, i.e. independents, as banks are more tightly regulated and supervised by federal regulators. To this end, we exploit two different datasets on state regulation, one pertaining to anti-predatory laws and the other to broker laws. We find evidence that the impact of independents on foreclosures was smaller in states that tightened their regulation prior to and during the boom.

A growing number of papers examine the boom-bust episode in the US housing market (Demyanyk and Van Hemert 2009; Doms, Furlong, and Krainer 2007; Gabriel and Rosenthal 2007; Gerardi, Shapiro, and Willen 2007; Dell'Ariccia et al., 2008; Mayer and Pence 2008; Keys et al. 2010, Mian and Sufi 2009a, 2009b, 2010; Purnanandam 2010). Our paper differs from this literature in that we distinguish between banks and independent lenders to understand the role of regulation, an issue that has received less attention from the literature so far. In that respect, our paper is most related to Keys et al (2009) that compare the performance of subprime securitized loans originated by banks and independents around a FICO threshold that induces an exogenous increase in securitization. They find that the moral hazard problem associated with securitization is more severe for banks. Our focus is instead on the aggregate effect which could be driven by loan performance over all FICO scores for both securitized and non-securitized loans. Few studies have looked at mortgage credit at the county level, Mian and Sufi (2010) and Favara and Imbs (2010) are important exceptions. Our paper is related to Mian and Sufi (2010) in that they study the impact of the increase in leverage on county performance during the crisis; we also show results with a similar flavor as we control for the growth in mortgage credit during the boom.

The rest of the paper is organized as follows. Section 2 describes the data used in this paper and presents summary statistics. Section 3 explores the expansion of independents during the boom. Section 4 presents our key finding on the relation between the market share of independents and county outcomes during the downturn using both parametric and semi-non-parametric methods. Section 5 addresses alternative hypotheses and further explores the role of regulation using data on state regulation. Section 6 concludes.

2 Data and Summary Statistics

2.1 Data

We construct our dataset by merging data from several sources. The data appendix provides comprehensive information on the data used, and a detailed description of the steps involved in the construction of the dataset. In what follows we summarize the main steps.

Our mortgage related data come from a comprehensive sample of mortgage applications and originations between 2003 and 2008 that were collected by the Federal Reserve under the provision of the Home Mortgage Disclosure Act (HMDA). Under this provision, the vast majority of mortgage lenders are required to report.[3] The HMDA data include information on the year of the application (the data is available on an annual basis), the amount of the loan, the lender's decision, and the income of the applicant. The data also provide useful information on the lender such as the name of the institution, its type, and its regulating agency. We thus can distinguish between depository institutions and their affiliates (banks, thrifts, credit unions and mortgage companies affiliated to them) and independent non-bank mortgage originators. We restrict our attention to mortgage applications that are considered as: home purchase, conventional, one-to-four-family, and owner-occupied. We also limit our study to mortgage originations in counties situated in a Metropolitan Statistical Area (MSA) for which data is available on house price growth and on the housing supply elasticity. This leaves us with 773 counties, which account for around 80% of total HMDA mortgage originations in 2005.[4] After imposing these restrictions, our 2003-2008 sample period consists

[3]See Data Appendix for more information about these requirements and the coverage of HMDA.

[4]Restricting our sample to these counties allows us to control for variables that are otherwise not available for other counties such as measures of house price growth and of the housing supply elasticity. Focusing on

of around 28 million applications which we aggregate at county level. We do so to construct variables that capture the volume of mortgage originations in each county during a given year as well as the share of mortgage origination by lender type. We also use these data to create various measures of the share of securitization witihin a county, Herfindhal index measures, and measures of geographical diversification of lenders (for the diversification measure see Loutskina and Strahan, 2011). HMDA data also provide the median income of the census tract of the property, which we take advantage of to compute the shares of census tracts in a county that fall within a given income bracket, for six income brackets.

To further control for demographic information and local economic conditions we also supplement our dataset with county characteristics from an extensive county level database consolidated by the Inter-University Consortium for Political and Social Research (ICPSE). We also make use of the Federal Housing Finance Agency (FHFA) data on house prices which are available at the MSA level. We also make use of TransUnion Trend Data to control for the average consumer credit score and the percentage of low consumer credit score in a county.

To control for geographical characteristics that could affect house price growth in a region we supplement our dataset with a land topology-based measure of housing supply elasticity constructed by Saiz (2010). Glaeser, Gyourkou, and Saiz (2008) show that areas with very high elasticity of housing supply are unlikely to experience large house price growth.

Our foreclosure data come from Realty Trac Foreclosure Market Trend Reports data.[5] Realty Trac provides comprehensive county coverage of foreclosure filings within a quarter. The reports are available starting from the second quarter of 2005. We thus use the second quarter of the years 2005, 2006 and 2007. By using data on the second quarter for each year, we are able to get a measure of the increase in quarterly foreclosure filings prior to the liquidity crisis and the official start of the recession in the U.S., thus ensuring our results are

the larger counties also helps minimize any noise in the data that could be brought by the inclusion of areas with a small population.

[5]A recent paper by Mian et al. (2011) also makes use of the same source to compute a measure of foreclosure rates.

not driven by these factors.

We make use of data on state regulation of mortgage brokers available from Pahl (2007), and a dataset on state level anti-predatory lending laws constructed by Bostic et al. (2008). We use these data to further explore the regulation aspect. We also supplement our data with information on the ratio of core deposit to total assets of all depository institutions which we obtain from the Reports of Condition and Income (Call Report) and from the Statistics on Depository Institutions (SDI), both available from the FDIC.

2.2 Summary statistics

We provide summary statistics from both the disaggregated loan level data and the aggregated county level data.
Table 1 provides summary statistics for the loans originated by banks and independents. The table shows statistics on originated loans in 2003, 2005, and 2007, on the loan amount, the applicant's income, and the loan to income ratio. In the upper table we show statistics from the full sample. Looking at the column titled N, the number of loans, we find that the number of originated loans has increased between 2003 and 2005, and then decreased between 2005 and 2007 for both banks and independents.[6] Note that 2005 was the peak year in loan originations as shown in Figure 6. However, the extent of the boom and bust was substantially larger for independents. Notably, while in 2003 independents made around 31% of loans, they contributed to more than 60% of the increase in mortgage originations between 2003 and 2005 and the decrease between 2005 and 2007. The upper table from the full sample shows that on average banks made loans to higher income applicants. The last column shows the p-value from a t-test of the difference in means. Much of this difference however is due to the fact that banks were significantly more active on the jumbo loan market.[7] Figure 1 shows

[6]We focus on the N values for the loan amount as there are around 4% of loans in our sample without information on applicant income. HMDA requires lenders to report income when this information was relied upon in making the credit decision.

[7]A jumbo mortgage is a mortgage loan in an amount above conventional conforming loan limits. This standard is set by the two government-sponsored enterprises Fannie Mae and Freddie Mac, and sets the limit

histograms of the applicant income of originated loans for both banks and independents. We see that the distribution is in fact similar across both subsamples with some exceptions, the most notable of which is a fatter right tail for banks. In the lower table we exclude jumbo loans and find that the differences in loan income and applicant amount narrows between banks and independents, although it remains significant except for the difference in the applicant income in 2005. As for the loan to income ratio, we find both in the full sample as well as in the non-jumbo loan sample that independents gave higher LTI loans in 2003 and 2007 but lower LTI loans in 2005.

Our analysis is carried at the county level and Table 2 summarizes the main variables. We rely on HMDA to construct our variables on mortgage volume and mortgage growth rates. In the first line of Table 2, we see that in the average county, mortgage credit grew by around 30% between 2003 and 2005. It then contracted by more than 80% between 2005 and 2007. The share of loans originated by independents varies substantially across countries as we can see in Figure 2. This distribution is relatively symmetric and the mean and median market share were around 23% in 2003. This market share has increased by 4% in 2005, due to the faster expansion of independents. The share of private securitization was, in mean and median, around 0.13. We also include broader measures of securitization in our empirical exercise (See Data Appendix). The foreclosure rate measures the percentage of properties with new filings during the quarter. On average, new foreclosures were filed for 0.1% of properties, during 2005Q2. The measure shows significant variation however with a standard deviation around 0.11. New foreclosure filings doubled between 2005Q2 and 2006Q2 and nearly tripled between 2005Q2 and 2007Q2. House prices were increasing rapidly between 2003 and 2005 with an average growth rate of 27% and a median of 19%. The growth rate substantially declined between 2005 and 2007. House prices entered their downturn trend only later in 2007 and early 2008 as can be seen in Figure 6.

on the maximum value of any individual mortgage they will purchase from a lender. The loan amount cutoff for 2005 is $359,650.

3 Mortgage credit expansion: 2003-2005

In this section we show that independent lenders contributed disproportionally to the mortgage boom. We first start with some motivating facts before presenting a simple empirical exercise to quantify differences between the expansion of banks and that of independents.

The year 2005 constituted the peak of a mortgage boom that started in early 2000s and substantially accelerated to register unprecedented levels of mortgage growth between 2003 and 2005. Figure 6 plots the log of total new mortgage originations in the U.S. illustrating the rise and fall of the mortgage market between 2003 and 2008. We focus on the differences between the contribution of independents and that of banks to the boom between 2003 and 2005. The number of originated loans in Table 1 strongly indicates that independents had a disproportional contribution as we discussed. Figure 3 plots a scatter of the market share of independents in 2005 against their market share in 2003 across counties. This figure is very telling as it shows that this expansion in the market share of independents took place in the vast majority of U.S. counties.

We quantify this difference between independents and banks by running simple regressions of the change in mortgage volume on a constant. Table 3 shows the outcome of these regressions. In the first column, we regress the change in total mortgage volume, by both banks and independents, on a constant. This constant is a measure of the average credit growth between 2003 and 2005, which is estimated at around 33%. In the second and third columns we show similar regressions where the endogenous variable is the change in mortgage credit by banks and independents respectively. They suggest that, on average, credit growth by independents was around 23% higher than that of banks. In the fourth column, the endogenous variable is the change in the county market share of independents. The result indicates that on average, the market share of independents grew by around 4%. We also look at whether the expansion of independents can be characterized as being inward or outward expansion. We thus regress, in the fifth column, the change in the market share of independents on a constant and on the lagged market share in 2003. The results suggest that independents gained market shares in new areas where they had lower presence in the past.

We next pursue the question of whether independents expanded more into areas that experienced higher house price growth. The premise is that an environment of high returns on housing is conducive to increased willingness by independents, due to lighter regulation, to lend to a segment of high risk applicants. Indeed, a major empirical challenger is to circumvent endogeneity. The expansion of independents, through its effect on the supply of mortgage credit, is likely to have contibuted to the rise in house prices. We address this issue by instrumenting for house price growth by the regions's housing supply elasticity. This instrument which is taken from a dataset constructed by Saiz (2010), is based on geographical characteristics of a Metropolitan Statistical Area (MSA) and thus exogenous to changes in mortgage credit. One would expect this variable to be negatively correlated with house prices growth between 2003 and 2005 since house prices are more likelyt o be more responsive to changes in the demand for housing (and the supply of mortgage credit) in areas where the supply of housing is low, i.e., the supply of housing more constrained due to geographical features of the area such as the proximity to water. This makes the housing supply elasticity a potentially good instrument for house price growth between 2003 and 2005.[8] In the sixth column, we show that a simple regression of the change in the market share of independents on housing supply elasticity, controlling for the market share in 2003, yileds a negative and significant coefficient suggesting that independents expanded mroe in areas that have on average a lower elasticity in housing supply.

We explore the association betwen house price growth and the change in the market share of independents in Table 4. In the first two columns we regress the growth rate of lending by banks and independents, respectively, between 2003 and 2005, on the growth rate of the housing price in the previous year, 2002. We find that on average, following an increase in house prices independents increased their lending by more. Ideally, however, we want to test

[8]This variable is also used as an instrument for house price growth between 2002 and 2006 in Mian and Sufi (2009).

whether independents expand more aggressively to areas that are experiencing a housing boom. To circumvent the previsouly mentionned endogeneity problem, we instrument for house price growth between 2003 and 2004 using the housing supply elatsicity measure. In column three we show results for the first stage regression of the house price growth between 2003 and 2005 on the housing supply elasticity. We find that the instrument is strongly correlated with the endogenous regressor. In colums four and five we show the second stage regressions where the dependent variable is banks' and independents' credit growth between 2003 and 2005, respectively. While there is a positive relation between house price growth and bank lending growth, the coefficient is small and far from significant. When the dependent variable is the growth in independents' lending, on the other hand, the coefficient becomes larger in magnitude and significant at the 10% level. Therefore, these results do suggest that independents expanded relatively faster in areas that are experiencing a house price boom.

4 The Rise in Foreclosures and the Role of Independents

In this section we exploit the geographical heterogeneity of lenders and show that, controlling for county characteristics, the market share of independents is a strong predictor of the early rise in foreclosures. We also show that it predicts the subsequent contraction in credit and house prices, as well as the rise in unemployment. We begin with some motivating facts before describing our empirical methodology. We leave the interpretation and the discussion of the results to the end.

4.1 Motivating Facts

It is now well established that the housing boom was fueled by a shift in mortgage supply as a result of deteriorating lending standards that led to a worsening in the risk profile of the marginal borrower, and to the subsequent rise in foreclosures (e.g. Mian and Sufi, 2009a). In light of these findings from the literature, the patterns documented in Section 3, alone, are suggestive of a faster deterioration in the lending standards of independents. It is indeed possible that due to their lack of regulation and supervision, independents were able to expand rapidly and rip the benefits from a booming housing sector while minimizing their perceived risk through the heavy reliance on an originate-to-distribute (OTD) model. This interpretation resonate well with some of the calls that were raised during the crisis for tighter regulation on the "shadow banking" sector, including independent mortgage lenders. Nevertheless, this remains an interpretation without direct evidence that lending by independents was associated with worse outcomes. We thus look at whether counties where independents channeled a larger share of mortgage loans fared worse during the crisis. We focus in particular on the rise in foreclosure as it is a direct result of the deterioration in lending standards, and since mortgage defaults were the first sign of mortgage trouble and were at the root of the subsequent housing downturn.[9] Figure 4 shows the spike in foreclosures which started as early as in 2006.

Figure 5 shows a scatter of the increase in foreclosure filings in a county between 2005Q2 and 2007Q2 against the market share of independents in 2005. The graph from the full sample (left) is suggestive of a strong positive relationship between these two variables. A further inspection shows that this relation is robust to the exclusion of counties with the very highest shares of independents (right). Indeed, this relation could be also driven in part by confounding county characteristics that are correlated with the presence of independents. This calls for an empirical model to control for these factors. We note, however, that the pre-crisis market share of independents is far from being fully explained by economic and demographic characteristics of the counties alone, nor by factors directly related to the housing boom. Independent lenders grew in prominence during the 80s and 90s, when they gained significant market shares in some regions in the U.S., mainly in some areas in the Southwest and some pockets in the South, Midwest, and on the East Coast. In some of

[9]See e.g. Demyanyk (2010) and Mayer et al. (2009).

these regions they became the main lenders or one of the largest in market share. While their expansion during the boom has increased their market share in several regions, both new regions and regions in which they are well established, the increase in market share during that period was only around 4%, and a large share of their market share as of 2005 is explained by their historical presence or by proximity to areas of strong presence.[10] While some of these areas can be characterized as having a lower average income and lower housing supply elasticity, the sample of counties with high market share of independents is a heterogeneous one, as is the sample of counties with low presence of independents. In the matching exercise, we are in fact able to match counties of similar economic and demographic similarities but with heterogeneous market shares of lenders. This heterogeneity allows us to control for factors that could be correlated with both the presence of independents and the rise in foreclosure. We also note that one of the interesting features of the rise in foreclosures between 2006 and 2008 is that it took place in areas with historically low foreclosures, thus it was not explained by a region's per-capita income or credit risk.[11]

We also look at three useful indicators of the severity of the crisis at the regional level: the contraction in credit and in house prices, and the rise in unemployment. Figure 7 shows scatters of the growth rates of credit and house prices, and the change in unemployment, between 2005 and 2008 against the share of independents in of 2005. The figure suggests that counties with higher market shares also tended to have worse outcomes during the crisis, and as explained in the footnotes of Figure 7, the fitted lined show a statistically significant relation. We show the change between 2005 and 2008 for ease of comparison, however, and as can be seen in Figure 6, aggregate credit contracted prior to the decline in house prices, and unemployment only started increasing in 2008. While it is impossible to avoid the effects of the recession and the credit crunch when studying the relation between the market share

[10]We are able to supplement our Appendix with some maps and further analysis on this issue if the referee finds that a substantiation on this issue would be useful.

[11]A notable example is the Southwest and particularly some areas in California that saw skyrocketing foreclosures despite a historically low average foreclosure rate. The Southern states are important examples of historically high foreclosure rate areas, and low average income, that many of which did not experience as sharp of an increase in foreclosures as other states did.

of independents and unemployment, due to its late rise in 2008, we will focus our empirical analysis on the 2005-2007 period when studying the impact on credit and house prices to minimize these effects.

4.2 Empirical methodology

We exploit the heterogeneity in the market share of independents across counties to study the impact of their market participation on foreclosure outcomes during the housing downturn. We study the change in foreclosure using quarterly foreclosure data from the second quarter of 2005, 2006, and 2007. The advantage of using quarterly data is that it allows us to track changes in foreclosure prior to the liquidity crunch and the official start of the recession in Q3 and Q4 of 2007, respectively. The challenge in studying this question is that the market share of independents could be correlated with county characteristics that affect our outcome variables. We carefully address this concern by controlling for a host of economic and demographic county characteristics. We seek to disentangle the impact of lender type from that of the county to understand whether two hypothetical identical counties would have experienced different economic outcomes due to a difference in the type of lenders that dominated their mortgage markets. One might also be concerned that a relation between our key variable, the market share of independents, and the rise in foreclosures could be affected by housing shocks that are correlated with both the market share of independents and the rise in foreclosures. While this is unlikely partly because house prices only started to decline in late 2007 and early 2008, we also aim to address this concern by instrumenting for house prices.

We also study the impact of our key variable, the market share of independents, on mortgage credit, house prices and unemployment during the downturn. Our aim from such exercise is to examine whether the market share of independents is also a strong predictor of severity of the housing downturn.

4.2.1 Parametric approach

Our first methodology consists of using standard regression analysis to study the determinants of the rise in foreclosures between 2005 and 2007, focusing in particular on the impact of the market participation of independents. Our benchmark regression is a simple ordinary least squares of the following form:

$$\Delta_{05Q2-07Q2}Forc_i = \beta_0 + \beta_1 Independent_{i,05} + \beta_1 X_{i,05} + \beta_2 \Delta_{03-05} Z_i + \beta_3 Securitization_{i,05} + \epsilon_i \tag{1}$$

where $\Delta_{05-07}Forc_i$ is change in *new* foreclosure filing rates between 2005Q2 and 2007Q2, in county$i$. $Independent_{i,05}$ is a measure of the market share of independent lenders in the base year 2005, the peak year in mortgage lending, and $X_{i,05}$ summarizes county-specific controls from or prior to 2005. In these county specific controls we include various information on economic and demographic variables in each county. To control for economic characteristics we include measures of per-capita income and unemployment in 2005, per-capita income growth during the boom between 2003 and 2005, categorical variables capturing the average consumer credit score and percentage of low credit score consumers, as well as six variables capturing the share of census tract in a county with a median income that falls in one of the six deciles of income brackets below 60K. To control for demographic characteristics we include variables capturing the share of Black population, the share of Hispanic population, and the average immigration rate between 2000 and 2005. We also control for the housing supply elasticity given that it captures the propensity of house prices to experiences boom-bust cycles. We also control for the extent of the mortgage boom between 2003 and 2005, $\Delta_{03-05}Z_i$, captured by the growth in house prices and mortgage credit during that period. This is because a higher $Independent_{i,05}$ might be associated with a faster expansion in credit and house prices. We thus explore whether lending by independents had a significant effect on foreclosure beyond its association with certain county characteristics or with the extent of the housing boom in these counties. We finally also control for the share of originated loans in a county that were sold for private securitization. There is now substantial evidence that securitization has led to worse lending standards. Since independents securitized a higher share of loans we control for securitization to differentiate between the effect of securitization and that of the type of the originator.

4.2.2 Matching methods

A salient feature of our empirical exercise is that, in addition to standard regressions, we also address the problems that could arise from using a linear regression with a poor distributional overlap of control variables and the risk of placing undue weight on a linear model by using matching methods. A linear representation might be inappropriate if the underlying relations between variables are highly non-linear. Also, a regression alone does not fully address the possibility that county characteristics are unbalanced between counties with varying market share of independents. Therefore, we supplement the standard parametric approach with a matching exercise. The objective of this approach is to reduce our sample to a subsample of counties that are similar on a set of covariates that we find likely candidates to be correlated with both, the main explanatory variable and the outcome variable. This approach also allows us to address the concern that the market share of independents might be highly correlated with county characteristics, as it involves testing whether the selected subsamples of high and low market share of independents are indeed similar on a set of county characteristics. Matching alone is not a method of estimation. It requires a technique to compute estimates. The literature usually makes use of some matching estimator to test the differences in means between the treated and control samples. We use the Abadie-imbens bias corrected estimator for this specific purpose. However, an important aspect of our exercise is that in addition to such estimates we re-run the earlier linear regressions using the matched sample of treated and control counties. Therefore the matching exercise is serving in essence as a nonparametric pre-processing of the data. Pre-processing the data the way we do reduces the correlation between our key variable and the controls and therefore makes estimates based on the subsequent parametric analysis far less dependent on modeling choices and specifications.[12] Ho et al (2010), show that after preprocessing the data estimates are less sensitive to changes in the parametric modeling assumptions. Furthermore, the exercise serves as a stringent robustness test for our earlier results by restricting our sample to characteristically similar counties.

4.3 Parametric Results

We first run a set of regressions following the linear model in (1) where the dependent variable is the change in new foreclosure filing rates between the second quarter of 2005 and the second quarter of 2007. The results are shown in Table 5. In the first column we run the regression with all controls included except for our key variable, the market share of independents (some regressors not shown in table due to space limit) . We also include state dummies and cluster error at the state level. We find that securitization was associated with an increase in foreclosure filings. This result is not surprising as there is now evidence showing that the OTD model has led to deterioration in lending standards (see e.g. Keys et al. 2010; Purnanandam, 2010). The estimate of the coefficient on securitization implies that an increase in one standard deviation of the securitization rate leads to an increase of 0.04 in the foreclosure filing rate. That means that 4 properties in every 10000 properties per quarter or 1/5th in the increase in average filings between 2005Q2 and 2007Q2. The estimate of the coefficient on per-capita income is not significant, but that is likely due to the inclusion of the census income level variables. The results also imply that counties that experience faster economic growth during the boom experienced less rise in foreclosures and that counties with a higher share of low credit score consumers and a higher share of Black population also display a more important rise in foreclosures. In the second column of Table 5 we include the market share of independent as a regressor. The estimate of the coefficient on

[12]See e.g. Rosenbaum and Rubin (1984), Rubin and Thomas (2000) and Imai and van Dyk (2004).

this variable is positive and significant at the 1%. It implies that an increase in one standard deviation in the market share of independents is associated with an increase in 0.08 in the rate of foreclosure filings, which is of important magnitude as it stands around 40% of the increase in average filings between 2005Q2 and 2007Q2. Interestingly, we find that the estimate of the coefficient on securitization loses its significance and becomes significantly smaller. This suggests that the coefficient in column (1) was capturing the effect of independents via their higher securitization rate. But as we control for the market share of independents we find that the type of lender is a more significant explanatory variable than securitization per se.[13] In the third column we control, in addition, for the house price growth between 2003 and 2005, and the growth rate in mortgage credit over that same period. We find that these factors do not significantly affect the coefficient on independents, and the estimates of their coefficient are not significant. This is likely due to the fact that we are studying the early rise in foreclosures, at which time the boom, particularly in house prices, was still ongoing.

In the fifth column, we show the result from a second stage regression of the change in foreclosure on the benchmark regressors (see column 2) and the house price growth between 2005 and 2007 instrumented by the housing supply elasticity and the lagged house price growth.[14] We find that even when we control for house prices the relation between independents and the rise in foreclosure remains strong despite a slightly smaller coefficient. The estimate of the coefficient on the instrumented house price growth is negative, in line with expectations, but not significant. These results minimize the concern that the relation between the market share of independents and the change in foreclosure rate could be driven by unobserved factors that affected house prices during that period.

In columns (5), (6), and (7) we repeat the above steps but replace the endogenous variable

[13]This finding is very robust and we later show that it also holds when controlling for different measures of securitization. In a regression of the change in foreclosure rates on the market share of independents and the share of securitization, *alone*, the estimates of both coefficients are positive and significant at the 1%. However, as we control for geographical and county characteristics, the share of securitization loses its significant, but the estimate of the coefficient on independents always remains signficant.

[14]The first stage F-statistic=14.1 and gives a partial $R^2 = 0.05$. The Sargan and Bassmann overidentificaton test yield a p-value of 0.96 and 0.97, respectively.

with the change in foreclosures between 2006 and 2007. This fully places the endogenous variable in the downturn period and allows us to address concerns related to our choice of studying the early rise in foreclosures and the possibility that some of our results might be reflecting correlations that are present during the boom but not during the bust episode. When the endogenous variable is the increase in foreclosure filings over one period only, the estimated coefficient on independents decreases in magnitude but remains significant, as shown in column (5). When we also control for the house price growth and the growth in mortgage credit in column (6) we find a positive and significant coefficient on mortgage credit growth, which also captures some of the effect of independents. The interpretation of this finding is relatively straightforward. Between 2006 and 2007 more U.S. counties have entered the downturn phase, in which case it is expected that the contraction to be at least partly explained by the extent of boom, as in most boom-bust episodes. As for the impact this has on the estimated coefficient on independents, it is expected that due to the fast expansion of independents, their market share in 2005 will be correlated with the growth rate of credit at county level. In the last column we also instrument for the house price growth in 2007 and find a negative and significant coefficient.[15] This also has an effect of decreasing the magnitude of the coefficient on independents; as we will see shortly, the market share of independents also predict a contraction in house prices, and therefore this explains the impact on its coefficient in column (7).

Figure 6 shows that the contraction in mortgage credit started in 2005, albeit to a mild degree as mortgage credit was still higher than that of 2003 and 2004 levels. In 2007, credit contracted substantially further bringing total credit to a significantly lower level than in the boom years. One might be concerned about how these movements in credit supply could affect the documented relation between independents and foreclosures. Arguably, however, movements in credit are only likely to affect foreclosures through their effect on house prices, and we do control for this variable. Nevertheless we also run regressions where

[15] The F-statistic from first stage is equal to 31.2 and the partial $R^2 = 0.11$. The Sargan and Bassman overidentification tests yield a p-value of 0.52 and 0.55, respectively.

the dependent variable is the change in total mortgage credit between 2005 and 2006 and find similar results. We also study the relation between independents and foreclosures in subsamples of counties based on their mortgage growth in 2005 and 2006. We find that the relation is more important in magnitude in the subsample of counties that were still experiencing a mortgage boom in 2005 and in 2006. These results are shown in Table 6. In the first column we run a simple regression on the full sample, of the change in foreclosures between 2005Q2 and 2006Q2 on a constant. In Column 2, we re-run the regression selecting only the subsample of counties that recorded higher than median growth in 2005 and in 2006. We find the constants in both regressions comparable which suggests that counties with fast growing mortgage market as of 2005 and 2006 also experienced a similar early rise in foreclosure. In column (3) we include the benchmark regressors in Table 5 using the full sample and find a positive and significant coefficient on independents. In column (4) we restrict the regression to the same sample of fast growing counties, while in column (5) we restrict it to the subsample of slow growing counties (below median growth in credit in 2005 and 2006) and find that the estimated coefficient on independents in column (4) is larger in magnitude. In summary, the aggregate patterns, together with the IV regressions from Table 5 and the results in Table 6 severely minimize the concern that the relation between independents and foreclosure is driven by factors related to house price and credit movements at the start of the downturn.

Credit, house prices, and unemployment We next explore whether counties with a higher market share of independents also experienced a more severe housing downturn and whether their regional economies were more impacted by the downturn. The rise in foreclosures alone can have important consequences on the regional economy through its effect on house prices (see e.g. Rogers and Winter, 2009; Mian et al., 2011). Lenders might also shy away from these counties due to an increase in the perceived riskiness of borrowers in these counties. These several hard-to-dissociate factors amplify the impact of foreclosures and might lead to a when-it-rains-it-pours effect. Disentangling the amplification mechanism is beyond the scope of this paper, however, and our objective in this subsection is to examine whether the presence of independents was also associated with worse outcomes in terms of credit, house prices, and unemployment. We focus on the early credit and house price contraction between 2005 and 2007 in order to minimize, to the best extent possible, the impact of the liquidity crunch.[16] As for unemployment, which is one of the hallmarks of the Great Recession, it started its rise only in 2008. Therefore we also include 2008 in our analysis while keeping in mind that some of this relation could be affected by the event of the liquidity crunch. The results are shown in Table 7. The first column shows the results from a linear regression similar to the one in equation (1) except that the endogenous variable is now the change in total mortgage credit in the county between 2005 and 2007. We first find that the market share of independents as of 2005 has a strong and significant negative impact on mortgage credit growth during the downturn. An increase in one standard deviation is associated with a contraction of around 5% in mortgage credit between 2005 and 2007 (0.1*-0.498). This sharper decline of credit in areas with higher pre-crisis market share of independents could be due to a combination of both demand and supply effects, as discussed earlier, both of which are likely related to the more important rise in foreclosures in these areas. We also find that the higher market share of securitization is associated with a sharper contraction in credit. However, this effect loses its significance when we control in the second column for the expansion in credit and house prices during the boom. Column (2) also suggests that the increase in house prices during the boom was also significantly negatively associated with credit growth during the downturn. This is expected as the extent of the boom is likely to be an important factor in explaining the severity of the bust. Controlling for the mortgage boom, however, only slightly decreases the magnitude of the coefficient on the market share of independents, which remains significant at the 1%. In the third and fourth columns, the dependent variable is the change in house prices between 2005 and

[16]The impact of the liquidity crunch on lenders could widely vary based on lenders' size and liability structure, and its impact on credit supply could be in part unrelated to lending standards during the boom.

2007. We find that there is a negative relation between the market share of independents and house price growth, but that this relation is only significant when we control for credit and house price growth during the boom. Note that unlike credit growth between 2005 and 2007, a more substantial housing boom predicts an increase in house prices between 2005 and 2007. This finding is likely due to the fact that there is a significant persistence in house prices as they only started to decline substantially in late 2007 and during 2008. In the fifth column the dependent variable is the change in unemployment between 2005 and 2007. The coefficient on independents is positive but not significant. As mentioned earlier, however, unemployment only started to increase during 2008.[17] We thus regress, in column (6), the change in unemployment between 2005 and 2008 on the benchmark regressors. We find that the market share of independents is a significant predictor of the rise in unemployment, and that a one standard deviation increase in the market share is associated with an increase of 0.16 points in unemployment rate.

4.4 Matching results

We use the Abadie-Imbens matching estimator which allows us to match counties with respect to both categorical and continuous variables. Since continuous observations cannot be exactly matched, the procedure allows for bias-correction for that purpose. Our matching procedure and the post-matching balancing tests are carried in a way similar to a recent literature that uses these methods. The matching strategy consists first of isolating a subsample of counties that share similar characteristics based on our key explanatory variable, the percentage of independent loans in 2005. The procedure is often used when the explanatory variable is categorical so that there is a clear cutoff between what is treated and what is not. In our case, our explanatory variable is continuous and therefore we choose an ad-hoc cutoff of the independent variable and we vary this cutoff for robustness. Such practice is standard when the variable is continuous (see e.g. Almeida et al., 2010). Our benchmark

[17] U.S. unemployment rate in 2007 was in fact only slightly higher than that in 2005, 5% in comparison to 4.9% respectively. Source: BEA.

cutoff is the upper 15% of counties in terms of their market share of independents as of 2005.[18] The smaller our sample is, the better our matches are, but decreasing our sample too much might jeopardize our statistical tests. We denote this subsample as the sample of "treated" counties. We end up with a sample of 107 treated counties. The objective is to match this subsample to another subsample of counties that are similar in characteristics.

We choose our covariates with the main endogenous variable in mind, the change in the rate of foreclosures.[19] The covariates that we have to control for should be variables that are likely to be correlated with both the market share of independents and the rise in foreclosures. It is absolutely important, however, to avoid using a covariate for which we suspect a direct causality from the market share of independents, such as, for example, the change in house prices during the boom. Such variables will be included in the linear regression that we run on the sample of treated and control counties, but cannot be included in the matching process (see e.g. Ho et al, 2010). Our choice of covariates is self explanatory: we choose to match on the county's per capita income, average credit score, housing supply elasticity, and unemployment rate. These are variables for which a causality from the treatment variable is highly unlikely, yet they are likely to be correlated with both the market share of independents and the rise in foreclosures. In the benchmark exercise we match counties in the U.S. without geographical restrictions. We also show the results from an exercise where we impose the matching to be restricted within a state, i.e., intrastate matching. We do so to address concerns that state foreclosure laws could play an important role, although we do control for state dummies in the post-matching regression stage.

4.4.1 Balancing tests

Upon completion of the matching estimation we conduct balancing tests. The objective of these tests is to ensure that the distribution of the conditioning variables, the covariates, does

[18] This cutoff corresponds to a market share of independents of 0.3854; choosing a cutoff corresponding to the higher 10% or higher 20% gives similar results.

[19] The fact that the outcome variable reflects a change in a flow variable addresses issues with unobservable time-non varying county characteristics.

not significantly differ across the treatment and the control groups. We use the Kolmogorov-Smirnov (KS) test of *distributional* differences as well as t-test to compare the means. The first row of Table 8 shows the change in foreclosures between treated and the control groups of counties. A visual comparison of the means and medians across the two groups suggest that the treated group experienced distinctly worse outcomes during the downturn. The KS and t-tests suggest these differences are significant. The next four rows compare the distribution of covariates between the treated and control subsamples. We find a strong similarity and the KS test cannot reject that they are generated by the same distribution, while the p-values from t-tests show that we cannot reject the equality of the mean. Table 10 shows similar results from the exercise in which, in addition to matching counties on the four covariates, we also impose on the counties to be from the same state. This constraint, indeed, makes it harder to find counties that are characteristically similar, nevertheless we find that the KS and t-test suggest that the differences in the distribution of the covariates and their means, respectively, are not significantly different between the treated and control subsamples. Note that the p-value from the KS test on income is relatively small (0.12), however, we find that on average it is the treated counties, i.e., counties with a higher market share of independents, that have a slightly higher per-capita income; this is a lesser reason for concern. The first row in Table 10 shows that, just like in the benchmark interstate matching, foreclosure outcomes are significantly worse in the treated sample.

4.4.2 The Abadie-Imbens Estimator

We next show the results from the Abadie-Imbens matching estimator. We show results from three different estimators: the sample average treatment effect (SATE), the sample average treatment effect on the treated (SATT), and the population average treatment effect on the treated (PATT). The results for the benchmark matching exercise are shown in Table 9 which reports the differential change in foreclosure filings rate, mortgage credit growth, and the change in unemployment rate between the treated and control samples.

The results confirm that treated counties had experienced a significantly sharper increase in foreclosures, as can be seen from all estimators which yield results of a similar magnitude. The treatment effect, i.e., having a high market share of independents, is estimated to be associated with an increase in foreclosure filings rate by around 0.26, which is higher than the average increase in foreclosure filings rate over that period. The results on mortgage credit and unemployment also confirm earlier findings, although we note that the impact on unemployment varies substantially depending on the estimator used. Table 11 shows the results from the intrastate match. The SATE estimator yields substantially lower difference but results from all estimators are again significant for the three variables. Interestingly we find that the SATT and PATT yield very similar results on the main outcome variable, foreclosures, in the benchmark and the intrastate matching exercises.

4.4.3 OLS on the matched subsample

The third step of our matching exercise consists of running the benchmark linear regression on the subsample of matched counties. The results are shown in Table 12. Note that we control, but do not show, for all previously used economic and demographic controls as well as for state dummies (see Table 5), and we cluster errors at the state level. The first three columns are regressions on the full sample for the three endogenous variables, change in foreclosures, credit growth and unemployment. The next three are from the benchmark matched subsample, while the last three are from the intrastate matched subsample. Looking at the coefficients on foreclosure first, we find that the estimated coefficients on the matched subsample are significantly larger in magnitude. In fact, the estimated coefficient in column (4) is twice the size of that in column (1). The estimated coefficient from the intrastate match, as shown in column (7), is even higher. These results are very encouraging as they show that as we focus our study on characteristically similar counties our key finding becomes sharper. As for the coefficients on mortgage credit and unemployment we find that they are similar in magnitude in the interstate match, although the coefficient on mortgage credit growth becomes only significant at the 10%.

4.5 Discussion

In the earlier section we have shown that independents contributed disproportionally to the lending boom and that, during the boom, the expansion in their market share was more pronounced in areas with a higher percentage of low credit score consumers, and areas experiencing higher house price growth. These findings alone hint to more severe deterioration in the lending standards of independents when compared to banks, particularly in light of the findings from the earlier literature that shows that the mortgage boom was to a great extent caused by an outward shift in the supply of mortgage which was fueled by greater moral hazard due to securitization (Mian and Sufi, 2009a). In this section, we examine the outcome of this mortgage boom and focus particularly on foreclosures, a variable that is more directly related to lending standards. We show that, even after controlling for county characteristics, counties where a higher share of mortgage lending was channeled by independents experienced a sharper rise in foreclosures. Indeed, it is the heterogeneity in the market share of independents that allows us to carry this exercise. Despite the correlation between the presence of independents and some of the county characteristics, it is far from a perfect correlation. A large share of the market share of independents as of 2005 is explained by their market share prior to the mortgage boom, as these lenders were concentrated in several geographical pockets. Many counties which did experience high price growth during the boom, and that had relatively lower average income and credit score were prior to the boom, and also as of 2005, largely dominated by banks. We control for county characteristics not only with standard parametric methods, but also by matching counties. These matching methods allowed us to verify the claim that the type of lender is not perfectly correlated with county characteristics.

These findings strongly indicate that the expansion of independents came at the expense of a significant deterioration in lending standards, one which led them to either lend to a riskier category of lenders, expend less effort in collecting soft information from the average borrower, design riskier contracts (but possibly more attractive for the less risk-averse borrowers), or all of the above. Such differential between the lending standards of banks and independents alone can explain the above results. Exploring the risks associated with independents' lending is, however, beyond the scope of this paper, but would be an important avenue for future research, possibly using disaggregated data. Our findings from the county level data establish correlations that are quantitatively important at the aggregate level and thus shed light on the aggregate contribution of independent lenders.

5 Exploring the Role of Regulation and Alternative Hypotheses

Compared to banks, a conspicuous characteristic of independents is their weak regulation and supervision. This difference offers a very plausible explanation to the patterns documented in this paper.[20] Less tightly regulated and supervised lenders, by definition, face fewer constraints when it comes to their lending policy. They are thus able to, under favorable circumstances such as the housing boom and the availability of the OTD technology, gain market shares by originating increasingly risky loans. We nevertheless check the robustness of this argument by (a) testing alternative hypotheses and (b) exploiting variation in mortgage related regulation across states.

5.1 Alternative Hypotheses

A long standing finance literature that examines the relation between competition and lending standards offers ambiguous results (see e.g. Jarayatne and Strahan, 1996; Black and Strahan, 2002; Campbell, 2006; Gabaix and Laibson, 2006; Dick and Lehnert, 2010).

[20]Keys et al. (2009) use this distinction between independents and banks to test for the impact of regulation.

Nevertheless, it suggests that competition can have a substantial effect on lending policy. One might ask, therefore, whether the market share of independents, our key variable in the analysis, is correlated with the degree of competition on the local market. To control for the regional competition effect we control for a Herfindahl index constructed for the top, 15, 30 and 50 lenders in the county (see e.g. Barth et al., 2009). We sequentially add these indexes on the right hand side of our benchmark regression of foreclosures on county characteristics. The results are shown in Table 13. In the first column we show the outcome of the benchmark regression for comparison. We then in columns (2), (3), and (4) control for our measures of market competition and find that the estimated coefficient on each of the Herfindahl measures are far from significant. Note that when we control for the Herfindahl indexes constructed for the top 30 and 50 lenders, in columns(2) and (3) respectively, our sample of counties becomes smaller, as there are counties with fewer than 30 and 50 lenders. Nevertheless, we find that the coefficient on independents remains positive and significant in all three, and becomes larger in magnitude as the sample size shrinks in (3) and (4).

Another concern is related to the geographical diversification of lenders. Recently, Loutskina and Strahan (2011) showed evidence that geographically concentrated lenders act like informed investors and tend to collect more information on the applicants, while geographical diversification has the opposite effect. One might argue that our results could be driven by a difference in the degree of geographical diversification of lenders, which could have an impact on the outcome of their lending. This is unlikely to explain our results, however, as the bulk of bank lending was originated by geographically diversified lenders. Nevertheless we control for this factor by computing the same index of lender diversification as in Loutskina and Strahan (2011) from which we compute a weighted measure of diversification at the county level.[21] We control for this measure in column (5) of Table 13 and find that it has virtually no impact on the coefficient of independents and that the estimated coefficient on the index is small and not significant.

[21]See Data Appendix.

One might also argue that differences in lending standards between banks and independents could be due to differences in their liability structure. In particular, banks typically rely on core deposits (in varying degrees across banks) while independent lenders are essentially wholesale lenders. There are two opposing predictions of the impact of deposit-taking on lending standards. On one hand, the presence of subsidized deposit insurance might lead to imprudent lending from banks. On the other hand, retail-lenders are more involved in relationship lending (see e.g. Song and Thakor, 2007) and thus might be better placed to efficiently screen applicants on soft information (see e.g. Purnanandam, 2010).[22] We address the question of whether the relations that we see in the data are driven by differences in deposit-taking activity rather than by differences in the regulatory framework by exploiting the heterogeneity in the extent of deposit-taking within banks. The increasing reliance on wholesale funding by banks during recent decades (see e.g. Feldman and Schmidt, 2001) makes our sample of banks a very heterogeneous one in terms of the ratio of core deposits to assets. To exploit this heterogeneity we obtain data on the ratio of core deposits to assets from the Reports of Income and Condition and from Statistics on Depository Institutions.[23] The median core deposits to assets ratio in our sample banks, as of 2005, is 0.51. A significant share of banks rely on deposits as a secondary source of funding as several large banks have ratios lower than 0.2. We therefore compute the share of loans originated in each county by banks with an above the median core deposits ratio, and also by banks above the upper quartile cutoff. The non-bank lending is, by definition, done by independents which can be characterized by a core deposits ratio equal to zero. We compete these measures with our measure of the market share of independents in columns (6) and (7). The results strongly suggest that the relation that we document is unlikely to be driven by the differences in deposit taking. We also control for other cutoffs as well as a weighted average measure of

[22] Another argument that would lead to a similar prediction is one related to the fragility induced by demand deposits as in Calomiris and Kahn (1991). However, wholesale funding or market borrowing are also subject to a sudden stop and recent literature suggests that wholesale lenders could be more vulnerable to withdrawal in episodes of liquidity shocks (see e.g. Gatev and Strahan, 2006; and Huang and Ratnovski, 2008).

[23] See Data Appendix.

core deposits in a given county (by imposing a ratio of core deposits to assets equal to zero for independents) and find similar results.

Several studies have recently established a negative relation between securitization and lending standards (see e.g. Keys et al., 2010; Purnanandam, 2010). This finding can be explained, as earlier studies argued, by a moral hazard argument by which an originate-to-distribute model diminishes banks' screening and monitoring incentives (see e.g. Petersen and Rajan, 1994; and Parlour and Plantin, 2008). In light of this finding, one might ask whether the heterogeneity in the rate of securitization between banks and independents can explain the relation between independents and the rise in foreclosures. We address this question in our benchmark regressions by controlling for the share of securitized loans at the county level. To compute this share we follow closely Mian and Sufi (2009a)'s definition of private securitization. The results suggested that securitization explains at best a small fraction of the effect of independents. We further address this question using other proxies for securitization. Specifically, in addition to private securitization we control in columns (8) and (9) for measures of the share of loans sold to GSEs and the share of loans that were kept on the balance sheet of the originator, respectively. We see that in column (8) the estimated coefficient on *Percent sold to GSE* to be negative but not significant. It slightly reduces the estimated coefficient on independents which however remains very significant. The result suggests that securitization to GSEs, unlike private securitization, is negatively correlated with the rise in foreclosures. Indeed, GSEs required minimum standards on the loans their purchased which could explain this correlation. The decline in the estimated coefficient on independents could thus be explained by the fact that they sold a relatively smaller share of their loans to GSEs. Nevertheless, this relation is weak and has only a small impact on the benchmark regression. Finally, in column (9) we control for the share of all non-securitized loans and find that the estimated coefficient to be positive and not significant.

5.2 State regulation

We next explore whether the strong association between lending by independents and the rise in foreclosures varied with the extent of mortgage market regulations across states. If this association can be explained by the lack of sufficient regulation of independents, then one might expect to find that this association is less (more) pronounced in more (less) regulated states. The premise is the following: *if* state mortgage-related regulations are effective in limiting risky loans, they are likely to have a more important effect on the lending of the otherwise less regulated lenders, i.e., independent lenders. The challenge in identifying such relation is the difficulty in measuring *effective* state regulation and supervision. State laws that regulate the mortgage market vary widely across states, however, market observers have pointed to a lack of enforcement problem (see e.g. Belskey and Retsinas, 2008; Treasury Blueprint, 2008; Immergluck, 2009). With these caveats in mind, we explore two datasets on state regulation. One dataset is constructed by Bostic et al. (2008) and reflects the extent of state restrictions on predatory lending laws. The second dataset is on state regulation of mortgage brokers and comes from Pahl (2007).[24] Note that in most states, brokers and lenders were supervised by the same state agency (see e.g. Immergluck, 2009), making this index a good candidate for a proxy of mortgage regulation and supervision of both mortgage brokers and lenders. These datasets thus focus on distinctive aspects of the mortgage market. Arguably, however, more regulation and supervision of mortgage brokers and more restrictive predatory lending laws should both act as constraints on risky lending. We thus run regressions where we interact the share of independents as of 2005 with one of these indexes on new state mortgage-related regulations. We focus on new regulations for several reasons. First, Bostic et al.(2008) make the distinction between pre- and post-1999 state regulations on anti-predatory lending, as the modern laws were patterned differently, akin the Home Ownership and Equity Protect Act (HOPEA) that congress enacted in 1994. They find that these new laws with broader coverage had an effect above and beyond the old laws. Second,

[24]See Data Appendix

since our dependent variable measures the change in new foreclosure filings, one would expect that examining new laws would also be more appropriate in our context. Third, since many of the state regulations were not effectively implemented during the mortgage boom (see e.g. Immergluck, 2009) a concern about effectiveness leads us to place more weight on new regulations which are a better proxy of a state's regulatory reaction to the mortgage boom. For these reasons, and for comparability with the data from Bostic et al. (2008), we examine the new state regulation on broker regulations which are available from Pahl (2007) between 1996 and 2005.[25] For each regulation measure, the anti-predatory lending laws and the broker regulations, we rank states and assign a dummy for the upper quartile of most regulated states. We do so to minimize the effect of the judgmental nature in which these indexes were constructed by sometimes a linear sum of subcomponents. Finally, since we are examining laws at the state level one cannot control simultaneously for state dummies. Instead we also control, in addition to the county characteristics, for state characteristics that could affect foreclosures such as the state GDP, and three dummies capturing foreclosure related laws (see Pence, 2006).

The results from these regressions are shown in Table 14. In the first column we show the results from the benchmark regression of the rise in foreclosure on the county and state controls, to which we add the dummy for states with high broker regulation. We find that the estimated coefficient on the dummy is negative, meaning that these states experienced on average a smaller increase in foreclosures during the downturn. In the second column we interact the broker dummy with the market share of independents (third row) and find a negative and significant coefficient. This result supports the premise that more regulation lessened the impact of independents on foreclosures. Note that the coefficient on the regulation dummy turns positive. This is surprising but could be due to a host of factors that we cannot control for, such as state specific effects. In the third column we cluster errors at the

[25]The data are also available for 2006 but we exclude this year out of a concern for possible endogeneity with the outcome variable. Nevertheless we include it in a robustness exercise and find that it does not affect our results (not shown).

state level and find that the coefficient remains significant at the 5%. In columns (4), (5) and (6) we re-do the exercise in the first three columns this time replacing the dummy for broker regulation with the dummy on the anti-predatory lending laws. Column (4) shows that there is a negative correlation between the dummy and the increase in foreclosure, yet it is far from significant. Interestingly however, when we interact this dummy variable with the market share of independents we find that the results mirror our earlier finding from the broker dummy, with however a smaller magnitude on the interaction variable which is also only significant at the 10% in the last column where errors are clusters at the state level. While the regulation variables we use are far from ideal, as they are not direct measures of effective state regulation of mortgage lenders, the results do suggest that the effect of independents on foreclosure is weaker in states that implemented stricter mortgage related regulations during the boom. Taken together with our robustness analysis, the findings suggest that regulation could be key in explaining the lender effect on foreclosures.

6 Conclusion

The evidence in this paper suggests that the lightly regulated independent lenders contributed disproportionately to the recent boom-bust housing cycle. We show that, to a large extent, the mortgage boom was fueled by a fast expansion of credit from independent lenders. We then show that the market share of these independents as of 2005 is a strong predictor of the increase in foreclosure between 2005 and 2007. We carefully control for county characteristics using both parametric and semi-nonparametric methods and show that these patterns are unlikely to be driven by factors unrelated to the lending standards of independents. We show robustness tests that suggest that this strong association between independents and the rise in foreclosures is most likely be due to the weak regulatory structure. We illustrate the macroeconomic consequences of these relations by showing that the presence of independents also predicts the contraction in credit and house prices and the subsequent rise in unemployment between in 2007 and 2008. Overall our findings lend support to the view that more stringent regulation could have averted some the volatility in the housing market during the recent boom-bust episode. Our study sheds light on the aggregate contribution of the least regulated lenders. An interesting avenue for future research is to identify, using disaggregated data, the characteristics that made lending by independents riskier than that by banks.

References

Abadie, A., Imbens, G., 2002. Simple and Bias-Corrected Matching Estimators for Average Treatment Eects. NBER Technical Working Paper No. 0283.

Almeida, H., Campello, M., Laranjeira, B., and Weisbenner, S., 2010. Corporate Debt Maturity and the Real Effects of the 2007 Credit Crisis. University of Illinois at Urbana-Champaign.

Barth, J. R., Lin, C., Lin, P., Song, F. M., 2009. Corruption in bank lending to firms: cross-country micro evidence on the beneficial role of competition and information sharing. Journal of Financial Economics 91, pp. 361-388.

Belsky, E., Retsinas, N.P., 2008. Borrowing to live: consumer and mortgage credit revisited. Brookings Institution Press 2008.

Belsky, E., Richardson, N., 2010. Understanding the boom and bust in the nonprime mortgage market. Harvard University working paper.

Berlin, M, Mester, L.J., 1999. Deposits and relationship lending. Review of Financial Studies, Volume 12, Issue 3, pp. 579-607.

Bostic R. W., Engel, K. C., McCoy P. A., Pennington-Cross A., and Wachter, S. M., 2008. State and local anti-predatory lending laws: The effect of legal enforcement mechanisms. Journal of Economics and Business 60, pp. 4766.

Black, S. E., Strahan, P., 2002, Entrepreneurship and bank credit availability. Journal of Finance 67, pp. 2807 2833.

Calomiris, C. W., Kahn, C., 1991. The Role of Demandable Debt in Structuring Optimal Banking Arrangements. American Economic Review 81, No. 3, pp. 497-513.

Calomiris, C. W., Stanley, L. D., and William, M., 2008. The foreclosure-house price nexus: lessons from the 2007-2008 housing turmoil. NBER working paper No. 14294.

Campbell, John Y., 2006, Household finance, Journal of Finance 61, 1553- 1604.

Demyanyk, Y. S., and Van Hemert, O., 2009. Understanding the subprime mortgage crisis. Review of Financial Studies.

Demyanyk, Y. S., 2010. Ten Myths about the Subprime Crisis. In: Lessons from the Financial Crisis: Causes, Consequences, and Our Economic Future. (Robert W. Kolb Series).

Dick, A., Lehnert, A., 2010. Personal Bankruptcy and Credit Market Competition. Journal of Finance 65, Issue 2, pp. 655-686.

Dell'Ariccia, G., Igan, D., Laeven, L., 2008. Credit booms and lending standards: Evidence from the subprime mortgage market. IMF Working Paper.

Doms, M., Furlong, F., Krainer, J, 2007. Subprime Mortgage Delinquency Rates. Federal Reserve Bank of San Francisco Working Paper.

Favara, G., Imbs, J., 2010. Credit Supply and the Price of Housing. Working Paper . HEC University of Lausanne.

Fellowes, M., 2006. Credit Scores, Reports, and Getting Ahead in America. Brookings Institute.

Feldman R., Schmidt J., 2001. Increased Use of Uninsured Deposits: Implications for Market Discipline. Federal Reserve Bank of Minneapolis-Fed Gazette.

Gabaix, X., Laibson, D., 2006, Shrouded attributes, consumer myopia, and information suppression in competitive markets, Quarterly Journal of Economics 121, pp. 505- 540.

Gaetev, E., strahan, P. E., 2006. Banks's advantage in hedging liquidity risk: theory and evidence from the commercial paper market. Journal of Finance 61, Issue 2, pp. 867-892.

Gabriel, S., and Rosenthal, S., 2007. Secondary Markets, Risk, and Access to Credit: Evidence from the Mortgage Market. Unpublished working paper. Syracuse University Working Paper.

Gerardi, K., Shapiro, A., and Willen, P., 2007. Subprime Outcomes: Risky Mortgages, Homeownership Experiences, and Foreclosures. Federal Reserve Bank of Boston Working Paper.

Glaeser, E. L., Gyourko, J., and Saiz, A., 2008. Housing supply and housing bubbles. Journal of Urban Economics 64, volume 2, pp. 198-217.

Ho, D. E., Imai, K., King, G., Stuart, E.A., 2007. Matching as Nonparametric Preprocessing for Reducing Model Dependence. Political Analysis 15, pp. 199-236.

Huang, R., Ratnovski, L., 2010. The dark side of wholesale funding. Journal of Financial Intermediation 20, Issue 2, pp. 248-263.

Imai, K., van Dyk, D., 2004. Causal inference with general treatment regimes: Generalizing the propensity score. Journal of the American Statistical Association 99, pp. 85466.

Immergluck, D., 2009. Foreclosed: high-risk lending, deregulation and the undermining of America's mortgage market. Cornell University Press.

Ivashina, V., and D., Scharfstein, 2010. Bank lending during the financial crisis of 2008. Journal of Financial Economics 97, pp. 319-338.

Keys, B., Mukherjee, T., Seru, A., Vig, V., 2009. Financial regulation and securitization: Evidence from subprime loans. Journal of Monetary Economics 56, Issue 5.

Keys, B., Mukherjee, T., Seru, A., Vig, V., 2010. Did securitization lead to lax screening? Evidence from subprime loans. Quarterly Journal of Economics 125.

Jarayatne, J., Strahan, P., 1996. The finance-growth nexus: Evidence from bank branch deregulation. Quarterly Journal of Economics 111, pp. 636- 670.

Loutskina, E., Strahan, P., 2009. Securitization and the declining impact of bank finance on loan supply: Evidence from mortgage originations. Journal of Finance.

Loutskina, E., Strahan, P., 2011. Informed and uninformed investment in housing: the downside of diversification. Review of Financial Studies, forthcoming.

Mayer, C., and Karen, P., 2008. Subprime Mortgages: What, Where, and to Whom? Federal Reserve Board Working Paper.

Mayer, C., Karen P., Shane S. M., 2009. The rise in mortgage defaults. Journal of Economic Perspectives 23, Number 1, pp. 27 50.

Mian, A., Sufi, A., 2009a. The Consequences of Mortgage Credit Expansion: Evidence from the US Mortgage Default Crisis. Quarterly Journal of Economics.

Mian, A., Sufi, A., 2009b. House Prices, Home Equity-Based Borrowing, and the U.S. Household Leverage Crisis , American Economic Review, forthcoming

Mian, A., Sufi, A., 2010. Household Leverage and the Recession of 2007 to 2009, IMF Economic Review 58, Issue 1, pp. 74-117

Mian, A., Sufi, A., Trebbi, F., 2011. Foreclosures, House Prices, and the Real Economy. Unpublished working paper. University of California, Berkeley.

Pence, K., 2006. Foreclosing on Opportunity: State Laws and Mortgage Credit. Review of Economics and Statistics.

Pahl, C., 2007. A Compilation of State Mortgage Broker Laws and Regulations, 1996 to 2006. Federal Reserve Band of Minneapolis.

Parlour, C., Plantin, G., 2008. Loans Sales and Relationship Banking. Journal of Finance, 63, 1291-1314.

Peterson, M.,Rajan, R., 1994. The Benefits of Lending Relationships: Evidence from Small Business Data, Journal of Finance 49, pp. 3-37.

Purnanandam, A. K., 2010. Originate-to-distribute model and the subprime mortgage crisis. Review of Financial Studies.

Rogers, W. H., Winter, W., 2009. The impact of foreclosures on neighboring housing sales. Journal of Real Estate Research 31, Number 4., pp. 455-479.

Rosenbaum, Paul R., and Donald B. Rubin. 1984. Reducing bias in observational studies using subclassification on the propensity score. Journal of the American Statistical Association 79, pp. 516 524.

Rubin, D. B., Thomas, N., 2000. Combining propensity score matching with additional adjustments for prognostic covariates. Journal of the American Statistical Association 95, pp. 573 585.

Saiz, A., 2010. The Geographic Determinants of Housing Supply. Quarterly Journal of Economics.

Song, F., Thakor, A., 2007. Relationship Banking, Fragility, and the Asset-Liability Matching Problem. Review of Financial Studies 20, pp. 2129 -2177.

Treasury Blueprint, 2008. The Department of the treasury blue print for amodernized financial regulatory structure.

7 Data Appendix

HMDA Data

We use a comprehensive sample of mortgage applications and originations that have been collected by the Federal Reserve under the provision of the Home Mortgage Disclosure Act (HMDA). Under this provision, the vast majority of mortgage lenders are required to report data about their house-related lending activity.[26] HMDA data covered around 95% of all mortgage originations in 2005 (see e.g. Dell'Ariccia et al., 2008), and has a better coverage within MSAs due to stricter reporting requirements in these areas.

The HMDA data provide information on the year of the application (the data is available on an annual basis), the amount of the loan, the lender's decision, and the income of the applicant. The data also provide information on the gender and race of the applicant, as well as other information on the census tract of the property such as the median income and share of minority households.

The raw HMDA data in our sample covering the sample period 2003 to 2008 period contain around 190 million applications. Of these, we keep only loans that are either approved or denied (Action code 1,2, and 3). We further restrict our loans types to be conventional (we exclude Federal Housing Agency, Veterans Administration, Farm Service Agency or Rural Housing Service), the property types to be one to four-family, the loan purpose to be

[26]Lenders are required to report if they meet certain criteria related to size, geographical location, the extent of housing-related lending activity, and regulatory status. Regarding size, a depository institution is subject to HMDA reporting requirements if it has assets of $34 million or more, as of December 31, 2004. In 2010, the Board raised this threshold to $40 million. For a non depository institution, total assets must exceed $10 million, as of December 31 of the preceding year, taking into account the assets of any parent corporation. Regarding the geographical location, lenders must report if they have offices in a Metropolitan Statistical Area (MSA) or if they are non-depository institutions with lending activities on properties located in an MSA. Lenders must also report if they are depository institutions with at least one home purchase loan or if they are non-depository institutions and they originate 100 or more home-purchase and refinancing loans. As for the regulatory status, lenders must report if they are non-depository institutions or if they are depository institutions that are federally insured or regulated.

home purchase only (excluding home improvement, refinancing purposes), and the occupancy status to be owner-occupied as principal dwelling. This leaves us with 34 million applications.

We distinguish between the type of lenders based on information available from HMDA on their regulatory agencies. Depository institutions and their affiliates (which we refer to as banks) are listed under the following agencies: Federal Deposit Insurance Corporation, Federal Reserve System, Office of the Comptroller of the Currency, Office of Thrift and Supervision, and National Credit Union Administration. Non-bank mortgage originators (independents) are listed under the Department of Housing and Urban Development.

We restrict our study to mortgage originations in counties situated in an Metropolitan Statistical Area (MSA) for which HMDA has better coverage and data on house prices and on house supply elasticity are available. This leaves us with 773 counties. These counties cover around 80% of total mortgage originations in HMDA in 2005.

We aggregate our data on mortgage originations at the county level which gives us the volume of loans originated in a county during a year. We can also distinguish between the originators. We calculate, in a county, the percentage of loans originated by independent mortgage companies and by banks.

HMDA provides information on the securitization process. Lenders are asked to report whether the originated mortgage was sold to a third party during the same calendar year in which it was originated. HMDA defines 8 types of purchasers. In the benchmark exercise we follow the approach of Mian and Sufi (2009a) and define securitization as being "private securitization", i.e., loans sold to private securitization pools, or sold to life insurance companies, credit unions, mortgage banks, and finance companies. We also supplement this measure with several other measures of securitization such as the share of of GSE securitization, as well as the share of non-securitized loans.

With the originated loan volume information, HMDA data allows us to construct measures on credit growth, bank competition (Herfindahl index) and geographic diversification. More specifically, for Herfindahl index we sum for each county the square of the percentage share of originated loans of the top 15 , 30, and 50 mortgage originators to create three respective competition indicators. The Herfindahl index ranges from near 0 for a county that has much bank competition to 1 for a county that has only bank, i.e. no competition.

For lender geographic diversification, we follow closely the method used in Loutskina and Strahan (2011). The variable measures the extent to which a lender concentrates its lending within a Metropolitan Statistical Area (MSA). The measure equals the sum of squared shares of loans made by a lender in each of the MSAs in which it operates, where the shares are based on originated loans. The geographic diversification measure ranges from near 0 for lenders operating cross most U.S. MSAs to 1 for lenders operating in a single MSA. We construct our county level index by taking weighted average of the indexes of geographical diversification for each lender in the region, weighted by their share of originated loans.

Inter-University Consortium for Political and Social Research

Inter-University Consortium for Political and Social Research (ICPSR), an affiliated institute of the University of Michigan, maintains a database on demographic and economic characteristics of U.S. counties. The sources of the database include the Bureau of the Census, the Bureau of Economic Analysis, the Bureau of Labor Statistics, as well as other sources (website: http://www.icpsr.umich.edu/icpsrweb/ICPSR/). For our county level analysis, we include the following economic and demographic characteristics: per capita personal income in 2005 ($CA0N0030\_05$), Percent of Black resident population in 2005 (PctBlack05), percent of Hispanic resident population in 2005 (PctH05), and average net international migration from 2001 to 2005 (IntlMig01,02,03,04,05). We also compute the per capita income growth between 2003 and 2005 using annual growth measures from the U.S. Bureau of Economic Analysis (BEA).

RealtyTrac Foreclosure Market Trend Data

The RealtyTrac U.S. Foreclosure Market Trend Report provides comprehensive data on foreclosures at the county level. Data is taken from more than 2,200 counties in the U.S. that account for more than 90 percent of the population. RealtyTrac's report provides foreclosure rates at the county level based on five types of documents filed in all three phases of foreclosure. Two filings, the Notice of Default and the *lis pendens* correspond to the first stage of foreclosure, prior to a foreclosure auction. Two filings are associated with the foreclosure auction, which are the Notice of Trustee Sale and the Notice of Foreclosure Sale. When a foreclosure auction is unsuccessful, the lender will legally repossess the property which is then filed as a REO, or Real Estate Owned. Our measure of foreclosure filings reflects all three stages of foreclosure and is a sum of all filings on properties in the county divided by the number of households in the county which is also provided by RealtyTrac. To avoid double counting, RealtyTrac only reports the most recent filing on a property. The report also checks if the same type of document was filed against a property in a previous month or quarter. When this is the case, the report does not count the property if a previous filing occurred within the estimated foreclosure time frame for the state the property is in. The reports are available from April 2005. We took the second quarter of 2005, 2006, and 2007 and use them to compute year on year changes as a measure of the increase in foreclosure filing rates.

Federal Housing Finance Agency

House Price Index (HPI) is a quarterly data published by the U.S. Federal Housing Finance Agency, an entity created in 2008 from the merging of the U.S. Office of Federal Housing Enterprise Oversight and the U.S. Federal Housing Board. As a weighted, repeated sales index, the HPI measures average price changes in repeat sales or refinancing on single family properties with mortgages that have been purchased or securitized by Fannie Mae or Freddie Mac. The HPI includes indexes for all nine Census Divisions, the 50 states and the District of Columbia, and every Metropolitan Statistical Area (MSA) in the U.S., excluding Puerto Rico. Compared to S&P/Case-Shiller indexes, the HPI offers a more comprehensive coverage of housing price trends in the U.S. metropolitan areas. We use the HPI data at MSA level (most disaggregated level that is available for this variable) and compute the year on year changes as a measure of house price growth in a given MSA.

TransUnion Trend Data

TransUnion is a leading consumer credit information company is the U.S., which offers credit-related information to potential creditors. It compiles the Trend Data, an aggregated consumer credit database that offers quarterly snapshots of randomly selected consumers, which enables the evaluation of actual consumer credit data over time. Data aggregations are available at national, state, metropolitan statistical area (MSA) and county levels. We use two categorical measures on credit scores in a county: Average Consumer Credit Score (ACCS) in 2004 and the Proportion of Low Consumer Credit Scores (PLCCS) as in Fellowes (2006).

Housing Supply Elasticity

Saiz (2010) provides a measure of housing supply elasticity at the MSA level computed based on topological factors. These factors are exogenous to house market conditions and population growth and are computed using both water and land slope constraint information obtained using Geographic Information System (GIS), United State Geographic Service (USGS), and USGS Digital Elevation Model (DEM). The data covers 269 Metropolitan areas using the 1999 county-based MSA or NECMA definitions. The geographic data is calculated using the principal city in the MSA, i.e., the first one on the list of a MSA name.

Call Report data

All regulated depository institutions in the United States are required to file their financial information periodically with their respective regulators. Reports of Condition and Income data are a widely used source of timely and accurate financial data regarding banks' balance sheets and the results of their operations. Specifically, every national bank, state member bank and insured non-member Bank is required by the Federal Financial Institutions Examination Council (FFIEC) to file a Call Report as of the close of business on the last day of each calendar quarter. The specific reporting requirements depend upon the size of the bank and whether or not it has any foreign offices. The availability of agency specific bank IDs in HMDA (Federal Reserve RSSD-ID, FDIC Certificate Number, and OCC Charter Number) allows us to match HMDA lenders that are depository institutions with their financials from the Call report. For savings institutions, i.e. depository institutions regulated by the OTS, we use the balance sheet information from Statistics on Depository Institutions (SDI), available from the FDIC, and match them with HMDA using OTS docket number.[27] We use the financial information to compute a core deposit ratio as total deposit minus time deposit over $100,000 divided by total asset (see e.g. Berlin and Mester, 1999). Naturally, for non-depository institutions we assign a zero for this ratio. We then rank lenders based on their core deposit (CD) and pick two thresholds for CD, 0.51 and 0.61, which correspond to the lower quartile and median values. We then compute the percentage share of banks in a county that is above these thresholds.

State Broker Regulation

We use Pahl's (2007) compilation of mortgage broker regulation in fifty states and the District of Columbia. These regulations pertain to requirements on the financial entity's controlling individual and managing principal (such as age, state of residency, pre-licensing education, examination results as well as net worth), requirements on the entity to maintain a minimum net worth or a surety bond, as well as physical office requirements such as maintaining a physical office in the state, obtaining a license or certificate and paying various fees. Pahl assigns a value for the intensity of each of twenty-four regulatory components. We focus on new regulations that were put in place by the various states between 1996 and 2005.

State Anti-predatory Law Index

In 1994, Congress enacted the first modern, comprehensive anti-predatory lending statute, the Home Ownership and Equity Protection Act (HOEPA). Starting in 1999, many states began adopting anti-predatory lending laws akin to HOPEA; these were labeled mini-HOPEA laws. These mini-HOEPA laws display considerable variation across states. Bostic et al. (2008) constructed a legal dataset of these laws in 50 states and the District of Columbia.

[27] http://www2.fdic.gov/sdi/

They also computed a state level index which scores the degree of restrictiveness on anti-predatory lending. The subcomponent of this index are indexes that measure the extent of: 1) Restrictions (limits on prepayment penalties, restrictions on balloon payments, requirements for credit counseling, and limits on judicial relief), 2) Coverage (number of loan types, APR trigger for first lien/subordinate mortgages, points and fees trigger) and 3) Enforcement mechanisms (assignee liability, enforcement against originators). We use their additive state level index of new mini-HOPEA laws which is available in Table 2 in their paper.

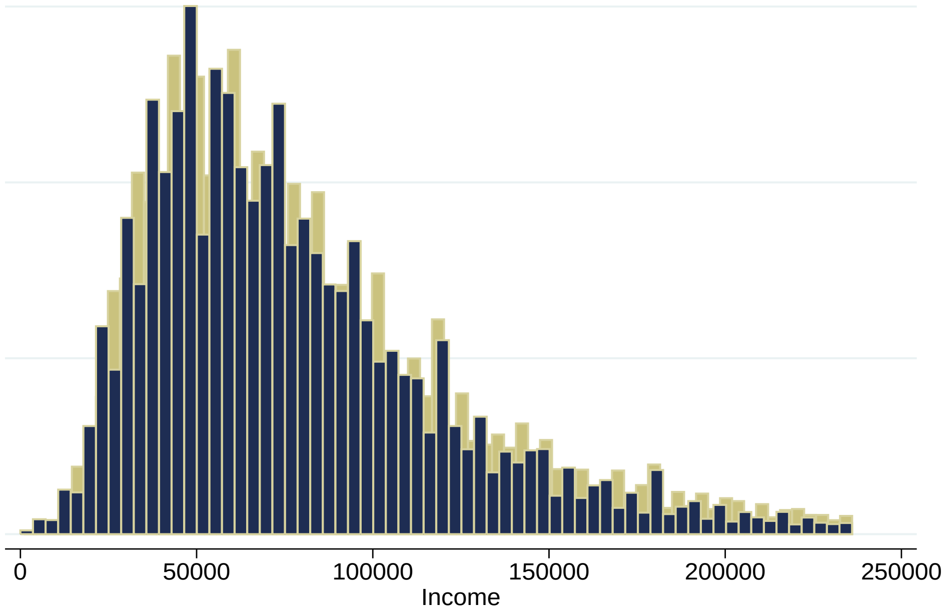


Figure 1: Income distribution.

Notes: This figure compares the income distribution of originated loans for each type of lender. The histogram of applicants' income for loans originated by independents is in black (dark blue in color).

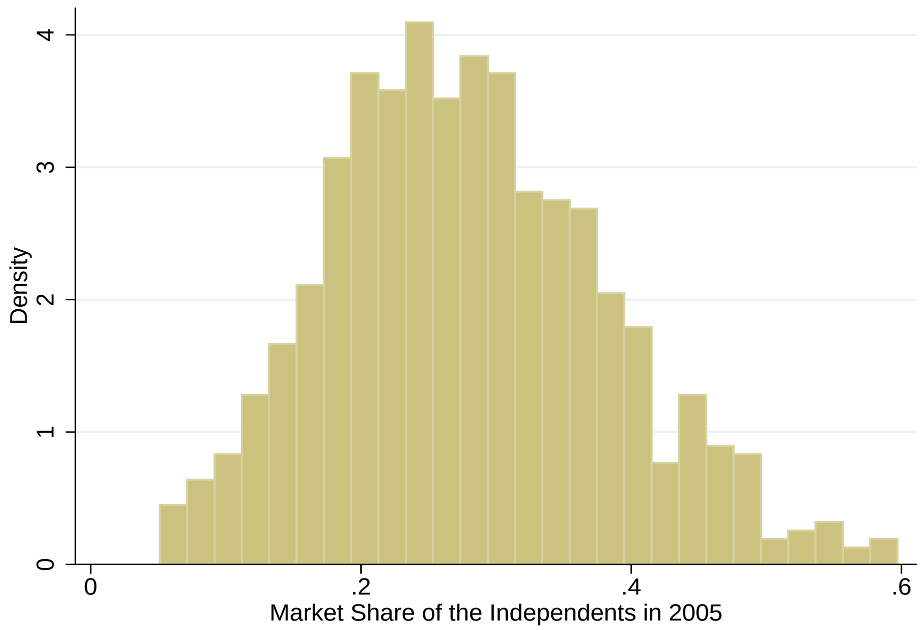


Figure 2: Market share of independents.

Notes: This figure shows a histogram of the market share of independents in our sample of 773 counties.

Figure 3: The expansion of independent lenders.

Notes: This figure shows the shift in Independents' share of the mortgage market between 2003 and 2005. For comparison we plot the 45 degree line to underline the upward shift.

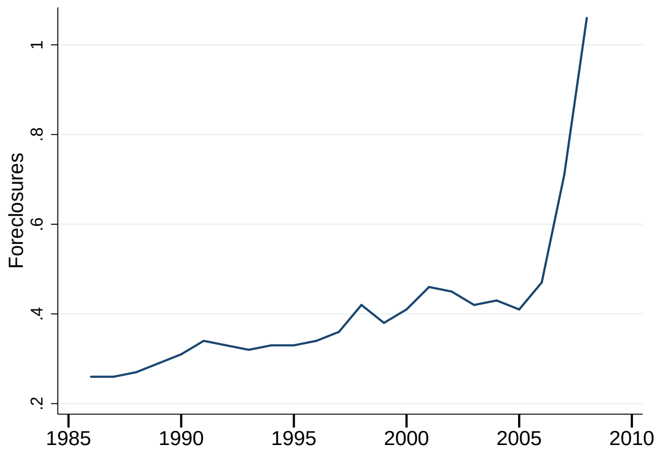


Figure 4: Foreclosures

Notes: This figure shows the evolution over time of foreclosure filings in percentage of originated mortgages. Source: HUD.

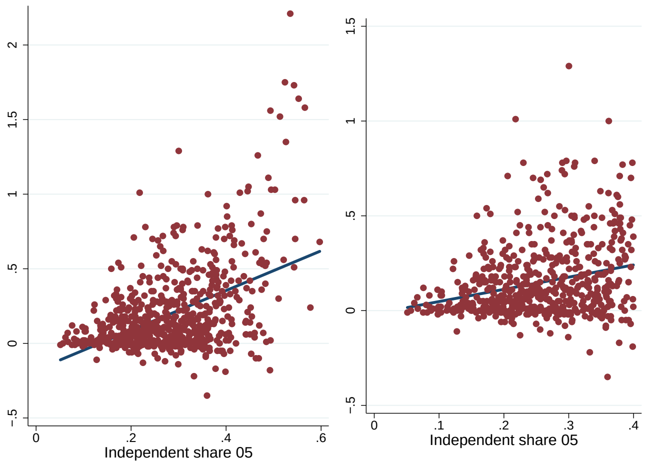


Figure 5: Foreclosures

Notes: This figure shows a scatter of the change in foreclosure filing rate (05Q2-07Q2) on the market share of independents as of 2005. In the left diagram we show the full sample. In the right doagram we show the close-up of the scatter eliminating counties with a market share of independents that is higher than 0.4. In both scatters we also fit a line from the regression of foreclosures on independents alone, in both samples we find a positive and significant coefficient.

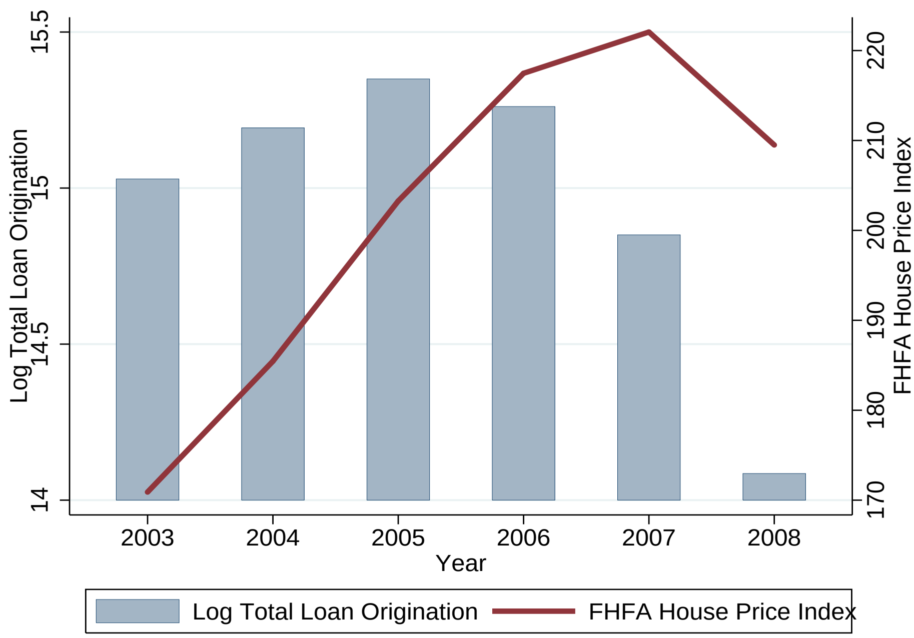


Figure 6: Mortgage credit boom and bust.

Notes: This figure plots the logarithm of total mortgage credit in our sample (bars) and an index of house prices in the U.S. (line, yearly average of quarterly data). Source: HMDA data (our sample, see Data Appendix) and FHFA.

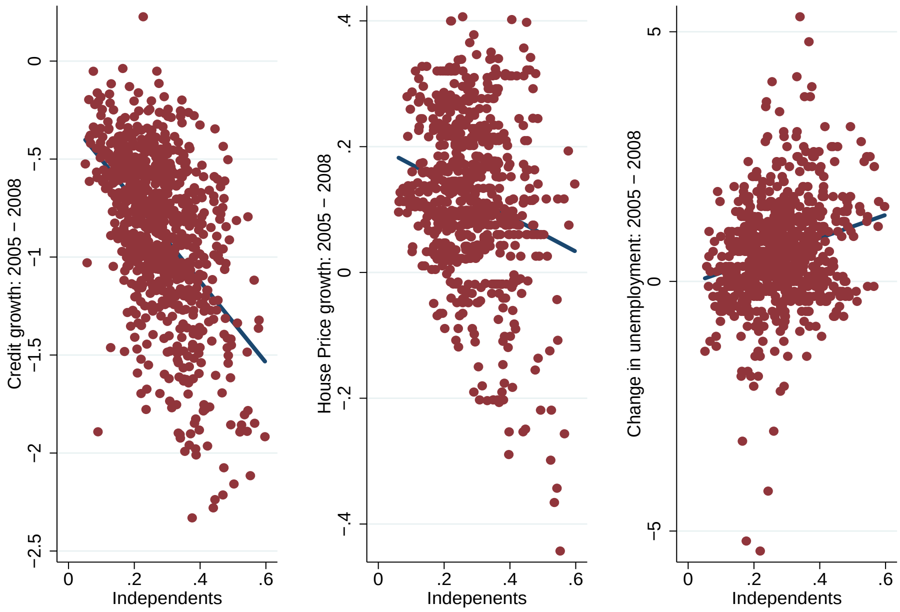


Figure 7: Credit, house prices, and unemployment.

Notes: This figure shows a scatter of mortgage credit growth between 2005 and 2008 (left) house price growth between 2005 and 2008 (middle) and change in unemployment rate between 2005 and 2008 (right), against the market share of independents as of 2005.

Table 1: Summary statistics: loan originations

This table presents summary statistics for the originated loans by both Independents and Banks for three years in our HMDA sample. Jumbo loan cutoffs are selected using information on loan limits from Fannie Mae and Freddie Mac for the corresponding year. We use the following limits for one family house mortgage loans: $322,700 for 2003, $359,650 for 2005, and $417,000 for 2007. See Data Appendix for detailed information on the selection of our sample.

Full Sample

		Banks			**Independents**			
	Year	N	Mean	Median	N	Mean	Median	p-value
Loan amount (000's)	2003	2,309,677	181,229	146,000	1,056,122	164,044	142,000	0.00
	2005	2,770,440	203818	154000	1,867,061	174,472	136,000	0.00
	2007	2,155,242	220008	165000	658,369	208,325	176,000	0.00
Applicant Income	2003	2,227,064	89,468	69,000	1,013,923	79,904	66,000	0.00
	2005	2,665,797	98,122	75,000	1,769,365	89,749	74,000	0.00
	2007	2,100,790	109,093	80,000	629,392	98,470	78,000	0.00
Loan to income	2003	2,227,064	2.32	2.26	1,013,911	2.58	2.28	0.00
	2005	2,665,797	2.31	2.29	1,769,365	2.12	2.11	0.00
	2007	2,100,790	2.35	2.32	629,392	2.56	2.47	0.00

Non-Jumbo Loans

		Banks			**Independents**			
	Year	N	Mean	Median	N	Mean	Median	p-value
Loan amount (000's)	2003	2,051,601	141,935	133,000	973,496	140,219	133,000	0.00
	2005	2,400,392	147,281	135,000	1,679,344	137,970	122,000	0.00
	2007	1,922,485	165,175	150,000	607,013	177,034	165,000	0.00
Applicant Income	2003	1,979,014	74671	64,000	935,856	72,310	63,000	0.00
	2005	2,309,230	79619	68,000	1,593,676	80,226	70,000	0.00
	2007	1,876,602	88270	73,000	581,551	88,171	74,000	0.39
Loan to income	2003	1,979,014	2.23	2.18	935,847	2.49	2.21	0.00
	2005	2,309,230	2.16	2.14	1,593,676	1.98	1.95	0.00
	2007	1,876,602	2.26	2.23	581,551	2.47	2.42	0.00

Table 2: Statistical summary of the county level variables

This table show summary statistics for main the county level variables in our dataset. See Data Appendix for detailed description of the sources and construction of these variables.

Source	Variable	N	Mean	Median	Min	Max	S.D.
HMDA data							
	Mortgage credit growth, 2003-2005	773	0.32	0.30	-0.30	1.61	0.22
	Mortgage credit growth, 2005-2008	773	-0.87	-0.81	-2.33	0.22	0.41
	Market share of independents, 2003	773	0.23	0.22	0.02	0.55	0.09
	Market share of independents, 2005	773	0.27	0.27	0.05	0.59	0.10
	Share of private securitization, 2005	773	0.13	0.13	0.02	0.59	0.05
	Herfindhal index 1, 2005	765	0.10	0.09	0.04	0.64	0.04
	Herfindhal index 2, 2005	743	0.09	0.09	0.04	0.35	0.03
	Herfindhal index 3, 2005	660	0.09	0.09	0.04	0.24	0.036
	Lender geographical diversification, 2005	773	0.19	0.16	0.03	0.67	0.12
Realty Trac							
	Foreclosure rate, 2005Q2	697	0.10	0.06	0.01	0.66	0.11
	Foreclosure rate, 2006Q2	684	0.19	0.11	0.01	1.94	0.24
	Foreclosure rate, 2007Q2	730	0.29	0.19	0.01	2.33	0.33
ICPSR							
	Per capita income, 2005	746	10.34	10.32	8.54	11.44	0.22
	Unemployment, 2005	766	5.03	4.9	2.3	15.9	1.38
	Share of Black population, 2005	773	11.02	6.1	0.06	78.57	13.04
	Share of Hispanic population, 2005	773	7.81	3.48	0.37	89.36	11.27
	International Immigration, 2000-05	773	0.010	0.006	-0.0007	0.087	0.013
BEA							
	Per capita income growth, 2003-2005	746	.13	.13	.03	.30	.06
FHFA							
	House price growth, 2003-2005	721	0.27	0.19	0.041	0.98	0.20
	House price growth, 2006	721	0.04	0.04	-0.05	0.19	0.04
	House price growth, 2006-2007	721	0.05	0.05	-0.21	0.30	0.07
Trans Union							
	Average consumer credit score	722	2.99	3	1	5	1.25
	Percentage of low consumer credit score	718	2.95	3	1	5	1.26
Saiz (2010)							
	Housing supply elasitcity	773	2.37	2.23	0.59	12.14	1.24

Table 3: Expansion of the Independents.

This table compares the mortgage expansion of banks with that of independents between 2003 and 2005. The first column regresses the change in total mortgage credit on a constant and state dummies. In columns (2) and (3) the same is repeated for banks and independents respectively. Column (4) simply regresses the difference between the growth rate of each type of lender. In columns (5) and (6) we show regressions in which the endogenous variable is the change in the market share of independents between 2003 and 2005. We cluster errors at the state levbels in the regressions corresponding to columns (5) and (6).

	(1) Mortgage Credit 03-05	(2) Bank Credit 03-05	(3) Independent Credit 03-05	(4) Δ Indep. share %03-05	(5) Δ Indep. share 03-05	(6) Δ Indep. share 03-05
Constant	0.329***	0.265***	0.500***	0.041***	0.059***	0.093***
	(40.4)	(33.2)	(41.51)	(21.06)	(8.17)	(8.35)
Independents					-0.076***	-0.116***
					(-2.31)	(-3.57)
Housing supply elasticity						-0.010***
						(-4.13)
N	773	773	773	773	773	773
adj. R^2	0	0	0	0	0.717	0.731

t statistics in parentheses

*$p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Expansion and house prices.

This table shows regressions of mortgage volume growth on house price growth. The first column shows the results from regressing mortgage growth for banks between 2003 and 2005 on house price growth in 2002, a set of economic and demographic variables, and state dummies. In the second column we show the same for independents. In the third column we show the result from the first stage of a IV regression where the dependent is the change in volume between 2003 and 2005, the instrumented endogenous variable is the change in house prices over the same period, and the instrument is the house price change over that same period. In the fourth and fifth column we show the results for the second stage where the dependent variables are bank credit growth and independents' credit growth, respectively. Erros are clustered at state level.

	(1)	(2)	(3)	(4)	(5)
	Bank	Indep.	House price growth	Bank	Indep.
	03-05	03-05	05	05	03-05
				IV	IV
House price growth, 2002	1.656***	3.317***			
	(4.00)	(3.69)			
Housing supply elasticity			-0.018***		
			(-2.75)		
House price growth, 2003-05				0.0242	1.112*
				(0.05)	(1.92)
Contstant	0.271***	0.396***	0.217***	0.170**	0.0229
	(20.44)	(13.81)	(4.23)	(2.35)	(0.20)
N	721	721	670	670	670
adj. R^2	0.382	0.246	0.813	0.447	0.330

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: An OLS of the rise in foreclosures on county characteristics.

This table shows results from the linear regression in equation (1). The dependent variable is the change in new foreclosures rates between 2005Q2 and 2007Q2 in columns 1 to 4, and between 2006Q2 and 2007Q2 in columns 5 to 7. See Table 2 and Data Appendix for details on the regressors. We also control for, but do not show, six income variables that capture the percentage of census tracts with a median income that falls into on of six income brackets, for the average immigration rate 2000-05, housing supply elasticity, and for state dummies. Errors are clustered at the state level.

	(1) 05-07	(2) 05-07	(3) 05-07	(4) 05-07 IV	(5) 06-07	(6) 06-07	(7) 06-07 IV
Market share of independent, 2005		0.831***	0.810***	0.648***	0.647**	0.544**	0.489***
		(2.94)	(2.85)	(2.80)	(2.41)	(2.05)	(3.14)
Private securitization, 2005	0.846**	0.315	0.149	-0.195	0.0138	-0.188	-0.385
	(2.64)	(0.94)	(0.41)	(-0.44)	(0.05)	(-0.57)	(-1.40)
Per-capita income, 2005	-0.00304	0.0325	0.0357	-0.0325	-0.0481	-0.00927	-0.0574
	(-0.03)	(0.30)	(0.34)	(-0.34)	(-0.44)	(-0.09)	(-0.82)
Income growth, 2003-05	-1.200***	-1.096***	-1.109***	-0.936***	-0.318	-0.395	-0.368
	(-3.78)	(-3.83)	(-3.50)	(-3.07)	(-1.44)	(-1.67)	(-1.49)
Unemployment,, 2005	0.00405	-0.00222	-0.00200	-0.0227	-0.00105	-0.000697	-0.0188
	(0.40)	(-0.22)	(-0.18)	(-1.13)	(-0.08)	(-0.05)	(-1.57)
Percentage of low credit score, 2004	0.0528*	0.0399	0.0423	0.0509***	0.0300	0.0296	0.0386***
	(1.89)	(1.65)	(1.67)	(2.91)	(1.54)	(1.48)	(2.76)
Foreclosure rate,, 2005Q2	0.0553	-0.0876	0.00687	0.0233	-0.485**	-0.409*	-0.388***
	(0.37)	(-0.49)	(0.04)	(0.20)	(-2.22)	(-1.80)	(-3.94)
Percent Black, 2005	0.00319**	0.00325**	0.00317**	0.00386**	0.0000652	0.0000815	0.000610
	(2.26)	(2.27)	(2.04)	(2.55)	(0.05)	(0.05)	(0.50)
Percent Hispanic 2005	0.00311	0.00214	0.00230	0.00392*	0.000795	0.00101	0.00168
	(0.88)	(0.73)	(0.78)	(1.96)	(0.29)	(0.36)	(1.15)
House price growth, 2003-05			0.122			0.165	
			(0.72)			(1.24)	
Mortgage credit growth, 2003-05			0.0338			0.134**	
			(0.50)			(2.69)	
House price growth, 2005-07				-2.531			
				(-1.23)			
House price growth, 2007							-2.599***
							(-2.58)
Constant	0.0814	-0.484	-0.533	0.551	0.405	-0.0235	0.667
	(0.07)	(-0.44)	(-0.51)	(0.46)	(0.36)	(-0.02)	(0.89)
N	624	624	583	583	594	557	557
adj. R^2	0.472	0.495	0.506	0.465	0.436	0.444	0.468

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Early rise in foreclosures in subsamples of counties selected based on mortgage growth.

This table shows the output of simple linear regressions where the endogenous variables is the change in new foreclosure rates between 2005Q2 and 2006Q2. In the first two columns we regress the dependent variable on a constant, first in the full sample (Full) and second in a susbsample of counties with mortgage growth above median both in 2005 and 2006 (High). In columns 3,4 and 5 we regress the dependent variable on our benchmark controls from Table 3 (second column) for the full sample, the subsample of counties with mortgage growth above median both in 2005 and 2006 and the subsample of counties with mortgage growth below median both in 2005 and 2006 (Low). The table only shows the coefficients on our key explanatory variable, the market share of independents. Errors are clustered at the state level.

	(1) Full	(2) High	(3) Full	(4) High	(5) Low
Market share of independents, 2005			0.344***	0.564**	0.337**
			(2.73)	(2.17)	(2.44)
Constant	0.0975***	0.105***	-0.611	0.369	-0.488
	(5.86)	(2.98)	(-0.85)	(0.29)	(-0.40)
N	632	176	593	161	188
adj. R^2	0.000	0.000	0.402	0.344	0.530

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Measures of the severity of the crisis and pre-crisis county characteristics.

This table shows results from linear regerssion of mortgage credit, house price and unemployment growth on the benchmark regressors. The dependent variable is the growth rate between 2005 and 2007, except for the last column which is the growth rate of unemployment between 2005 and 2008. See Table 5 and Data Appendix for details on the regressors. We also control for, but do not show, six income variables that capture the percentage of census tracts with a median income that falls into on of six income brackets, for the average immigration rate 2000-05, housing supply elasticity, percentage Black, percentage Hispanic, and for state dummies. Errors are clustered at state level.

	(1) Credit 05-07	(2) Credit 05-07	(3) House price 05-07	(4) House price 05-07	(5) Unemployment 05-07	(6) Unemployment 05-08
Market share of	-0.498***	-0.454***	-0.0939	-0.168***	0.687	1.639**
independents, 2005	(-3.21)	(-3.76)	(-1.47)	(-3.10)	(1.36)	(2.61)
Private Securitization,	-0.550**	-0.235	-0.00794	-0.0932	0.644	0.631
2005	(-2.25)	(-1.49)	(-0.08)	(-1.18)	(1.07)	(0.45)
Per capita	0.0368	0.0292	-0.0337	-0.0171	-0.265	-0.356
income, 2005	(0.52)	(0.51)	(-1.23)	(-0.66)	(-1.53)	(-1.42)
Income growth	0.702***	0.738***	0.178**	0.126*	-0.0990	0.251
2003-05	(3.38)	(3.89)	(2.11)	(1.72)	(-0.17)	(0.26)
Unemployment ,	-0.0107	-0.0135	-0.00965	-0.00803	-0.245***	-0.125
2005	(-0.85)	(-1.32)	(-1.32)	(-1.23)	(-3.07)	(-1.15)
Percentage of low	0.0110	0.0118	0.00501	0.00361	-0.0648*	-0.0256
credit score, 2005	(0.90)	(1.28)	(0.87)	(0.63)	(-1.70)	(-0.52)
Foreclosure rate,	0.150*	0.0619	-0.0225	0.0185	0.202	0.520*
2005Q2	(1.84)	(0.94)	(-0.73)	(0.77)	(1.09)	(1.80)
House price		-0.471***		0.227*		0.862
growth, 2003-05		(-4.01)		(2.00)		(1.61)
Mortgage credit		-0.0650		0.0696**		0.201
growth, 2003-05		(-1.08)		(2.06)		(0.99)
Constant	-0.423	-0.238	0.571*	0.365	2.793	4.350
	(-0.58)	(-0.39)	(1.91)	(1.31)	(1.46)	(1.55)
N	644	599	599	599	644	599
adj. R^2	0.719	0.755	0.696	0.743	0.713	0.711

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Comparison between treated and control, interstate matching.

This table compares our outcome variables and covariates (*) between the treated and control samples. The KS test compares the distribution across both samples while the t-test is test of the difference in means.

		Mean	25%	Median	75%	KS Test	T Test
Foreclosure 05-07	Treated	0.432	0.130	0.360	0.610	0.000	0.000
	Control	0.154	0.000	0.090	0.230		
Credit Risk*	Treated	3.174	2.0	3.0	4.0	1.000	0.965
	Control	3.168	2.0	3.0	4.0		
Elasticity*	Treated	1.651	1.067	1.529	2.241	0.207	0.286
	Control	1.735	1.196	1.629	2.302		
Unemployment*	Treated	5.2	4.3	5.0	5.6	1.000	0.716
	Control	5.2	4.2	4.9	5.8		
Income*	Treated	10.385	10.207	10.385	10.567	0.697	0.817
	Control	10.379	10.205	10.387	10.536		

Table 9: Matching estimators, interstate.

This table shows the Abadie-Imbens matching estimators from the benchmark interstate matching exercise. We compare the change in mortgage volume, foreclosure, and unemployment between the matched samples. The matching estimators shown in columns are the average treatment effect, the average treatment effect on the treated and the average treatment effect on the treated where standards error are adjusted for the population.

	SATE	SATT	PATT
Foreclosure 05-07	0.2556***	0.2606***	0.2606***
	(0.0398)	(0.0371)	(0.0418)
Volume 05-07	-0.1624***	-0.1652***	-0.1652***
	(0.0226)	(0.0225)	(0.0284)
Unemployment 08	0.2670***	0.3190***	0.5175**
	(0.0990)	(0.0896)	(0.1604)

Table 10: Comparison between treated and control, intrastate matching.

This table compares our outcome variable and covariates (*) between the treated and control samples. The KS test compares the distribution across both samples while the t-test is test of the difference in means.

		Mean	25%	Median	75%	KS Test	T Test
Foreclosure 05-07	Treated	0.416	0.135	0.385	0.605	0.000	0.000
	control	0.170	0.05	0.19	0.27		
Credit Risk*	Treated	3.2333	2	3.5	4	0.928	0.594
	Control	3.122	2	3	4		
Elasticity*	Treated	1.717	1.100	1.550	2.553	0.704	0.721
	Control	1.758	1.068	1.605	2.175		
Unemployment*	Treated	5.222	4.1	5.1	5.7	0.179	0.240
	Control	4.99	3.9	4.9	5.8		
Income*	Treated	10.375	10.171	10.353	10.582	0.126	0.712
	Control	10.362	10.168	10.253	10.547		

Table 11: Matching estimators, instrastate

This table shows the Abadie-Imbens matching estimators from the state matching exercise. We compare the change in mortgage volume, foreclosure, and unemployment between the matched samples. The matching estimators shown in columns are the average treatment effect, the average treatment effect on the treated and the average treatment effect on the treated where standards error are adjusted for the population.

	SATE	SATT	PATT
Foreclosure 05-07	0.1964**	0.2600***	0.2600***
	(0.0991)	(0.0392)	(0.0428)
Volume 05-07	-0.0873***	-0.1302***	-0.1302***
	(0.0352)	(0.0238)	(0.0242)
Unemployment 05-08	0.1969**	0.2310***	0.2310***
	(0.0767)	(0.0688)	(0.0747)

Table 12: Comparison of OLS on full, interstate matched, and intrastate matched samples.

The endogenous variables are the change in foreclosure (05-07), denoted by For, mortgage credit growth (05-07), denoted by Vol., and the change in unemployment rate (05-08). The first three columns show regressions of these dependents variables on the market share of independents controlling for county economic, demographic characteristics, and for the growth in mortgage and house prices during the boom. The next three columns (4)-(6), run the same regressions on the subsample of matched counties from the benchmark *interstate* matching exercise. The last three columns (7)-(9), run the same regressions on the subsample of matched counties from *intrastate* matching exercise. As in the benchmark regressions, we control for county characteristics and state dummies (but do not show) and cluster errors at the state level.

	All sample			Interstate matching			Intrastate matching		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	For.	Vol.	Un.	For.	Vol.	Un.	For.	Vol.	Un.
Market share of	0.840***	-0.499***	1.900***	1.685***	-0.423*	2.078**	1.982***	-0.536**	1.177*
independents, 2005	(3.37)	(-3.83)	(2.90)	(3.97)	(-1.76)	(2.45)	(5.54)	(-2.14)	(1.75)
House price	0.124	-0.488***	0.859	0.309*	-0.301	0.757	0.242	-0.511***	1.582**
growth, 2003-05	(0.80)	(-4.36)	(1.54)	(1.71)	(-1.37)	(1.14)	(1.09)	(-7.64)	(2.84)
Mortgage credit	0.0339	-0.0517	0.211	-0.0683	-0.158	0.531**	-0.0779	-0.145	0.807
growth, 2003-05	(0.47)	(-0.84)	(1.07)	(-0.58)	(-1.37)	(2.04)	(-0.40)	(-1.36)	(1.18)
Constant	-0.205	-0.596	5.125*	-0.385	-0.342	-5.549	-1.853	2.289	-15.64**
	(-0.21)	(-0.98)	(1.94)	(-0.25)	(-0.27)	(-1.04)	(-0.72)	(1.63)	(-2.13)
N	583	599	599	278	280	280	162	162	162
adj. R^2	0.504	0.752	0.714	0.572	0.790	0.796	0.414	0.796	0.849

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Robustness analysis.

This table shows variation on the benchmark regression in the second column of Table 5. The first column shows the benchmark regression. Columns (2), (3) and (4) add measures of local market lender competition to the regressors: a Herfindahl index for the top 15, 30 and 50 lenders, respectively. Note that only 376 counties have more than 50 lenders. Column (5) controls for a measure of the geographical diversification of lenders in the county (see Strahan and Louskina, 2011; and the Data Appendix). Columns (6) and (7) control for the share of loans originated by banks with core deposits ratio (CD) above 0.51 and 0.61, respectively (see text and Data Appendix). Columns (8) control for the share of loans that are sold to GSEs. Column (9) controls for the share of originated loans that are not sold. As in the benchmark regression, we control for county characteristics (see Table 5), state dummies, and we cluster errors at state level.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Market share of independents	0.831***	0.835***	1.003***	1.353***	0.833**	0.866***	0.892***	0.790***	0.922***
	(2.94)	(2.92)	(3.27)	(3.36)	(2.66)	(2.70)	(2.92)	(2.88)	(3.36)
Private securitization	0.315	0.312	0.247	0.202	0.316	0.325	0.336	0.235	
	(0.94)	(0.92)	(0.58)	(0.40)	(0.94)	(0.96)	(1.00)	(0.73)	
Herfindahl index top 15		0.0716							
		(0.27)							
Herfindahl index top 30			-0.0289						
			(-0.07)						
Herfindahl index top 50				0.258					
				(0.36)					
Geographic diversification of lenders					0.00526				
					(0.04)				
Percent originated by $CD > 0.5$ *banks*						0.0624			
						(0.42)			
Percent originated by $CD > 0.6$ *banks*							0.150		
							(0.88)		
Percent sold to GSEs								-0.199	
								(-1.37)	
Percent not sold									0.106
									(0.54)
Constant	-0.484	-0.477	-0.241	-0.331	-0.490	-0.530	-0.540	-0.279	-0.404
	(-0.44)	(-0.43)	(-0.20)	(-0.27)	(-0.45)	(-0.48)	(-0.48)	(-0.24)	(-0.36)
N	624	622	558	376	624	624	624	624	624
adj. R^2	0.495	0.494	0.500	0.582	0.494	0.494	0.495	0.495	0.494

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Foreclosures and state regulations

This table shows results from regressions of the change in foreclosures filings rates on the benchmark controls (see Table 5, column 2), state controls, and dummies for state mortgage-related regulations. We do not control for state dummies since the regulation variables are at the state level. The dummy variable "States with high broker regulation" indicates that the state is in the top quartile on the broker regulation index constructed based on Pahl's (2007) index of new mortgage broker regulations between 1996 and 2005 (see text and Data Appendix). The dummy variable "States with high anti-predatory laws" indicates that the state is in the top quartile on the anti-predatory lending laws index constructed by by Bolstic et al. (2008) based on various indicators of new state regulations between 1999 and 2005 (see column 5, Table 2, p. 55 in their paper). In the third and fifth rows we control for the interaction of these dummies with the market share of independents. We also control for property laws that affect foreclosures which are taken from Pence (2006).

	(1)	(2)	(3)	(4)	(5)	(6)
Market share of independents, 2005	0.871***	1.086***	1.086***	0.859***	1.054***	1.054***
	(6.40)	(7.61)	(3.14)	(6.27)	(7.04)	(2.99)
States with high broker regulation	-0.0596***	0.213***	0.213*			
	(-2.72)	(3.23)	(1.91)			
Independents#broker		-0.956***	-0.956**			
		(-4.37)	(-2.10)			
States with high anti-predatory lending laws				-0.00623	0.180***	0.180*
				(-0.29)	(2.83)	(1.92)
Independents#anti-predatory					-0.660***	-0.660*
					(-3.11)	(-1.77)
Judicial foreclosure	0.0601***	0.0481**	0.0481	0.0577***	0.0546***	0.0546
	(2.93)	(2.36)	(1.06)	(2.80)	(2.67)	(1.08)
Statutory right of redemption required	-0.0194	-0.0258	-0.0258	-0.0140	-0.0129	-0.0129
	(-0.71)	(-0.95)	(-0.69)	(-0.51)	(-0.47)	(-0.31)
Deficiency judgment prohibited	0.0704**	0.0601*	0.0601	0.0711**	0.0662**	0.0662
	(2.17)	(1.88)	(0.84)	(2.18)	(2.04)	(0.81)
State per capita GDP, 2002	0.0382***	0.0352***	0.0352	0.0404***	0.0382***	0.0382
	(2.81)	(2.63)	(1.17)	(2.96)	(2.81)	(1.05)
Benchmark controls	YES	YES	YES	YES	YES	YES
Cluster errors at state level	NO	NO	YES	NO	NO	YES
Constant	-1.455*	-1.537*	-1.537	-1.355	-1.078	-1.078
	(-1.72)	(-1.84)	(-1.19)	(-1.59)	(-1.27)	(-0.77)
N	594	594	594	594	594	594
adj. R^2	0.183	0.208	0.208	0.173	0.185	0.185

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Group Lending or Individual Lending? Evidence from a Randomized Field Experiment in Mongolia*

Orazio Attanasio,† Britta Augsburg,‡ Ralph De Haas,§ Emla Fitzsimons,¶ Heike Harmgart‖

December 29, 2011

Abstract

Although microfinance institutions across the world are moving from group lending towards individual lending, this strategic shift is not substantiated by sufficient empirical evidence on the comparative impact of both types of lending on borrowers. We present such evidence from a randomized field experiment in rural Mongolia. We find a positive impact of access to group loans on food consumption and entrepreneurship. Among households that were offered group loans the likelihood of owning an enterprise increases by ten per cent more than in control villages. Enterprise profits increase over time as well, particularly for the less-educated. For individual lending on the other hand, we detect no significant increase in consumption or enterprise ownership. These results are in line with theories that stress the disciplining effect of group lending: joint liability may deter borrowers from using loans for non-investment purposes. Our results on informal transfers are consistent with this hypothesis. Borrowers in group-lending villages are *less* likely to make informal transfers to families and friends while borrowers in individual-lending villages are *more* likely to do so. Importantly, we find no significant difference in repayment rates between the two lending programs, neither of which entailed weekly repayment meetings.

Keywords: Microcredit; group lending; poverty; access to finance; randomized field experiment

JEL Codes: 016, G21, D21, I32

*The authors thank Marco Alfano, Artyom Sidorenko, and Veronika Zavacka for excellent research assistance and Erik Berglöf, Marta Serra Garcia, Robert Lensink, David Roodman, Jeromin Zettelmeyer, and participants at the Women for Women/J-Pal/EBRD Conference on 'Banking on Women: Finance and Beyond', the 2nd European Research Conference on Microfinance (Groningen), and seminars at the EBRD, the Economic Research Institute of the National University of Mongolia, the Frankfurt School of Finance & Management, and XacBank for useful comments. This project would not have been possible without the tireless support from Ariunbileg Erdenebileg, Jargalmaa G., Maria Lotito, Bold Magvan, Norov Sanjaajamts, Otgochuluu Ch., and Benjamin Shell at XacBank; Oksana Pak at the EBRD; Tsetsen Dashtseren at MARBIS; Erin Burgess, Stephen Butler, Pamela Loose, and Jeffrey Telgarsky at NORC, and the Mongolian Women's Federation.

†University College London, Institute for Fiscal Studies, and J-PAL (o.attanasio@ucl.ac.uk).

‡Institute for Fiscal Studies (britta_a@ifs.org.uk).

§European Bank for Reconstruction and Development (dehaasr@ebrd.com, corresponding author).

¶Institute for Fiscal Studies (emla_f@ifs.org.uk).

‖European Bank for Reconstruction and Development (harmgart@ebrd.com).

Group Lending or Individual Lending? Evidence from a Randomized Field Experiment in Mongolia*

December 29, 2011

Abstract

Although microfinance institutions across the world are moving from group lending towards individual lending, this strategic shift is not substantiated by sufficient empirical evidence on the comparative impact of both types of lending on borrowers. We present such evidence from a randomized field experiment in rural Mongolia. We find a positive impact of access to group loans on food consumption and entrepreneurship. Among households that were offered group loans the likelihood of owning an enterprise increases by ten per cent more than in control villages. Enterprise profits increase over time as well, particularly for the less-educated. For individual lending on the other hand, we detect no significant increase in consumption or enterprise ownership. These results are in line with theories that stress the disciplining effect of group lending: joint liability may deter borrowers from using loans for non-investment purposes. Our results on informal transfers are consistent with this hypothesis. Borrowers in group-lending villages are *less* likely to make informal transfers to families and friends while borrowers in individual-lending villages are *more* likely to do so. Importantly, we find no significant difference in repayment rates between the two lending programs, neither of which entailed weekly repayment meetings.

Keywords: Microcredit; group lending; poverty; access to finance; randomized field experiment

JEL Codes: 016, G21, D21, I32

---

1 Introduction

The effectiveness of microcredit as a tool to combat poverty is much debated now that after years of rapid growth microfinance institutions (MFIs) in various countries - including India, Bosnia and Herzegovina, and Nicaragua - are struggling with client overindebtedness, repayment problems, and in some cases a political backlash against the microfinance sector as a whole. This heightened scepticism, perhaps most strongly voiced by Bateman (2010), also follows the publication of the findings - summarized below - of a number of randomized field experiments indicating that the impact of microcredit might be more modest than thought by its strongest advocates. These studies have tempered the expectations many had about the ability of microcredit to lift people out of poverty.

Much remains unclear about whether, and how, microcredit can help the poor to improve their lives. Answering these questions is even more important now that the microcredit industry is changing in various ways. In particular, increased scale and professionalization has led a number of leading MFIs to move from group or joint-liability lending, as pioneered by the Bangladeshi Grameen bank in the 1970s, to individual microlending.[1]

Under joint liability, small groups of borrowers are responsible for the repayment of each other's loans. All group members are treated as being in default when at least one of them does not repay and all members are denied subsequent loans. Because co-borrowers act as guarantors they screen and monitor each other and in so doing, reduce agency problems between the MFI and its borrowers. A potential downside to joint-liability lending is that it often involves time-consuming weekly repayment meetings and exerts strong social pressure, making it potentially onerous for borrowers. This is one of the main reasons why MFIs have started to move from joint to individual lending.

Somewhat surprisingly, there as yet exists very limited empirical evidence on the relative merits of individual and group lending, especially in terms of *impacts on borrowers*. Both the ample theoretical and the more limited empirical literature mainly center on the impact of joint liability on repayment rates. Armendáriz and Morduch (2005, p. 101-102) note that: "*In a perfect world, empirical researchers would be able to directly compare situations under group-lending contracts with comparable situations under traditional banking contracts. The best test would involve a single lender who employs a range of contracts (...). The best evidence would come from well-designed deliberate experiments in which loan contracts are varied but everything else is kept the same.*"

This paper provides such evidence from a randomized field experiment among 1,148 poor women in 40 villages across rural Mongolia. The aim of the experiment, in which villages were randomly assigned to obtain access to group loans, individual loans, or no loans, is to measure and compare the impact of both types of microcredit on various poverty measures. Importantly, neither the group nor the individual-lending programs include mandatory public repayment meetings and are thus relatively flexible forms of microcredit.

The loans provided by the programs we investigate are relatively small, targeted at female

[1] Liability individualization is for instance at the core of 'Grameen Bank II'. Large MFIs such as ASA in Bangladesh and BancoSol in Bolivia have also moved towards individual lending. Cull, Demirguç-Kunt, and Morduch (2009) show that joint-liability lenders tend to service poorer households than individual-liability lenders.

borrowers, and progressive in nature: successful loan repayment gives access to another loan cycle, with reduced interest rates, as is the case with many microcredit programs. Our evaluation is based on two data rounds of collections: a baseline survey collected before the start of the loans, and a follow-up survey collected 18 months (and potentially several loan cycles) after the baseline.

Though the loans provided under this experiment were originally intended to finance business creation, we find that in both the group- and in the individual-lending villages, about one half of all credit is used for household rather than business goals. Women who obtained access to microcredit often used the loans to purchase household assets, in particular large domestic appliances. Only among women that were offered group loans do we find an impact on business creation: the likelihood of owning an enterprise increases for these women by ten per cent more than in control villages. We also document an increase in enterprise profits but only for villages that had access to microcredit for longer periods of time. In terms of poverty impact, we find a substantial positive effect of access to group loans on food consumption, particularly of fruit, vegetables, dairy products, and non-alcoholic beverages.

In terms of individual lending, overall we document no increase in enterprise ownership, although there is some evidence that as time passes women in these villages are more likely to set up an enterprise jointly with their spouse. Amongst women in individual-lending villages we also detect no significant increase in (non-durable) consumption, though we find that women with low levels of education are significantly more likely to consume more.

The stronger impact on consumption and business creation in group-lending villages, after several loan cycles, may indicate that group loans are more effective at increasing the permanent income of households, though we detect no evidence of higher income in either individual- or group-lending villages, relative to controls. If one were to take at face value the evidence on the larger impact of group loans, one would want to ask *why* such loans are more effective at raising consumption (and probably long-term income). One possibility is that the joint-liability scheme better ensures discipline in terms of project selection and execution, so that larger long-run effects are achieved. We document results on informal transfers that support this hypothesis: women in group-lending villages decrease their transfers to families and friends, contrary to what we find for women in individual-lending villages.

The remainder of this paper is structured as follows. Section 2 summarizes the related literature, and this is followed by a description of our experiment in Section 3. Section 4 then explains our estimation methodology, and Section 5 provides the main results. Section 6 concludes.

2 Related literature

This paper provides a comparative analysis of individual versus joint-liability microcredit and as such is related to the theoretical literature on joint-liability lending that emerged over the last two decades.[2] Notwithstanding the richness of this literature, the impact of joint liability on

[2]See Ghatak and Guinnane (1999) for an early summary. Theory suggests that joint liability may reduce adverse selection (Ghatak, 1999/2000 and Gangopadhyay, Ghatak and Lensink, 2005); ex ante moral hazard by preventing excessively risky projects and shirking (Stiglitz, 1990; Banerjee, Besley and Guinnane, 1994 and Laffont and Rey, 2003); and ex post moral hazard by preventing non-repayment in case of successful projects (Besley and Coate, 1995 and Bhole and Ogden, 2010).

risk taking and investment behavior remains ambiguous. For instance, on the one hand, group lending may encourage moral hazard if clients shift to riskier projects when they expect to be bailed out by co-borrowers. On the other hand, joint liability may stimulate borrowers to reduce the risk undertaken by co-borrowers since they will get punished if a co-borrower defaults.

Giné, Jakiela, Karlan, and Morduch (2010) find, based on laboratory-style experiments in a Peruvian market, that contrary to much of the theoretical literature, joint liability stimulates risk taking - at least when borrowers know the investment strategies of co-borrowers. When borrowers could self-select into groups there was a strong negative effect on risk taking due to assortative matching. Fischer (2010) undertakes similar laboratory-style experiments and also finds that under limited information, group liability stimulates risk taking as borrowers free-ride on the insurance provided by co-borrowers (see also Wydick, 1999). However, when co-borrowers have to give upfront approval for each others' projects, ex ante moral hazard is mitigated. Giné and Karlan (2010) examine the impact of joint liability on repayment rates through two randomized experiments in the Philippines.[3] They find that removing group liability, or introducing individual liability from scratch, did not affect repayment rates over the ensuing three years. In a related study, Carpena, Vole, Shapiro and Zia (2010) exploit a quasi-experiment in which an Indian MFI switched from individual to joint-liability contracts, the reverse of the switch in Giné and Karlan (2010). They find that joint liability significantly improves loan repayment rates.

To the best of our knowledge, there as yet exists no comparative empirical evidence on the merits of both types of lending from the borrower's perspective. Earlier studies that focus on the development impact of microcredit study either individual or joint-liability microcredit, not both in the same framework. In an early contribution, Khandker and Pitt (1998) and Khandker (2005) use a quasi-experimental approach and find a positive impact of joint-liability microcredit on household consumption in Bangladesh, though one must acknowledge the possibility of omitted variable and selection bias. Morduch (1998) and Morduch and Roodman (2009) replicate the Bangladeshi studies and find no evidence of a causal impact of microcredit on consumption. Kaboski and Townsend (2005) also use non-experimental data and document a positive impact of joint-liability microcredit on consumption but not on investments in Thailand. Based on a structural approach the authors corroborate this finding in Kaboski and Townsend (2011). Bruhn and Love (2009) use non-random opening of bank branches in Mexico to analyze the impact of access to individual loans on entrepreneurship and income. They find that branch openings led to an increase in informal entrepreneurship amongst men but not women. Because women in 'treated' municipalities start to work more as wage-earners they eventually increased their income too.

More recently, randomized field experiments have been used to rigorously evaluate development policies, including microcredit (Duflo, Glennerster and Kremer, 2008). Banerjee, Duflo, Glennerster and Kinnan (2010) randomly phase in access to joint-liability microcredit in the Indian city of Hyderabad. The authors find a positive impact on business creation and investments by existing businesses, while the impact on consumption is heterogeneous. Those that start an enterprise reduce their non-durable consumption so they can pay for the fixed cost of the start-up

[3] Ahlin and Townsend (2007) empirically test various repayment determinants in a joint-liability context.

(which typically exceeds the available loan amount). In contrast, non-entrepreneurs increase their non-durable consumption. Crépon, Devoto, Duflo, and Parienté (2011) find that the introduction of joint-liability loans in rural Morocco led to a significant expansion of the scale of pre-existing entrepreneurial activities. Here as well there was a heterogeneous impact on consumption with those expanding their business decreasing their non-durable and total consumption.

Two other field experiments focus on individual-liability loans. Karlan and Zinman (2011) instructed loan officers in the Philippines to randomly reconsider applicants that had been labelled 'marginal' by a credit-scoring model. They find that access to loans *reduced* the number and size of businesses operated by those who received a loan. In a similar vein, Augsburg, De Haas, Harmgart and Meghir (2011) analyze the impact of microcredit on marginal borrowers of a Bosnian MFI. In contrast to Karlan and Zinman (2011), they find that microcredit increased entrepreneurship although the impact was heterogeneous - similar to Banerjee et al. (2010) and Crépon (2011). Because microloans only partially relaxed liquidity constraints, households had to find additional resources to finance investments. Households that already had a business and that were highly educated did so by drawing on savings. In contrast, business start-ups and less-educated households, with insufficient savings, had to cut back consumption. These households also reduced the school attendance of young adults aged 16-19.

Our paper is the first to use the same experimental context to compare the impact of individual versus joint-liability microcredit on borrowers.

3 The experiment

3.1 Background

Microfinance, as it is known today, originated in Bangladesh - one of the most densely populated parts of the world with 1,127 people per km^2 - but has also taken hold in less populated countries. One of these is Mongolia, which encompasses a land area half the size of India but with less than 1% of the number of inhabitants. This makes it the least densely populated country in the world with just 1.7 people per km^2.[4] This extremely low population density means that disbursing, monitoring, and collecting small loans to remote borrowers is very costly, particularly in rural areas. Mongolian MFIs are therefore constantly looking for cost-efficient ways to service such borrowers.

Mongolian microcredit has traditionally been provided in the form of individual loans, reflecting concerns that the nomadic lifestyle of indigenous Mongolians had impeded the build up of social capital outside of the family. Notwithstanding such concerns, informal collective self-help groups (*nukhurlul*) have developed and some of these have started to provide small loans to their members, in effect operating as informal savings and credit cooperatives. This indicates that group lending might be feasible in rural Mongolia too. Moreover, recent theoretical work suggests that when group contracts are sufficiently flexible, group loans can be superior to individual loans even in the absence of social capital (Bhole and Oden, 2010). This implies

[4]Source: United Nations World Population Prospects (2005). Mongolia has a semi-arid continental climate and an economy dominated by pastoral livestock husbandry, mining, and quarrying. Extreme weather conditions - droughts and harsh winters with temperatures falling below -35º C - frequently lead to large-scale livestock deaths.

that group lending may also work in countries were social connectedness and the threat of social sanctions is relatively limited.

This paper describes a randomized field experiment conducted in cooperation with XacBank, one of Mongolia's main banks and the second largest provider of microfinance in the country, to compare the impact of individual and group loans on borrowers' living standards.[5] While XacBank provides both men and women with microcredit, our experiment focused on extending credit to relatively less well-off women in rural areas. This target group was believed to have considerably less access to formal credit compared with richer, male, and urban Mongolians. According to the Mongolian National Statistics Office (2006, p. 54): "*Microcredit appears to be unavailable to most of the poor living in the aimag and soum centers. Their normal channels for credit are to borrow from a shop or kiosk where they often buy supplies or from a relative or friend*".[6]

3.2 Experimental design

The experiment took place in 40 soum centers (henceforth: villages) across five aimags (henceforth: provinces) in northern Mongolia. Figure A1 in the Annex maps the geographical location of all participating villages and provinces. The experiment started in January-February 2008 when XacBank loan officers and representatives of the Mongolian Women's Federation (MWF) organized information sessions in all 40 villages. [7]The goal and logistics of the experiment were explained and it was made clear to potential borrowers that there was a 2/3 probability that XacBank would start lending in their village during the experiment and that lending could take the form of either individual or group loans. Women who wished to participate could sign up and were asked to form *potential* groups of about 7 to 15 persons each. Because of our focus on relatively poor women, the eligibility criteria stated that participants should in principle own less than 1 million Mongolian tögrög (MNT) (USD 869) in assets and earn less than MNT 200,000 (USD 174) in monthly profits from a business.[8] Many of these women were on official 'poor lists' compiled by district governments.

Various indicators show that the households in our sample lie markedly below the Mongolian average in terms of income, expenditures, and social status. Data from the Mongolian statistical office indicate that the average rural household in 2007 had an annual income of MNT 3,005,000 (USD 2,610) whereas the average household in our sample earned MNT 1,100,000 (USD 955) (we define earnings as profits from household enterprises plus wages from formal employment

[5]According to XacBank's mission statement, it intends to foster Mongolia's socio-economic development by providing access to comprehensive financial services to citizens and firms, including those that are normally excluded such as low-income and remote rural clients. The bank aims to maximize the value of shareholders' investment while creating a profitable and sustainable institution.

[6]Mongolia is divided into 18 aimags or provinces which are subdivided into 342 soums or districts. Each soum contains a small village or soum center of on average 1 kilometer in diameter. The average soum in our experiment had 3,853 inhabitants of which on average 1,106 people (314 households) lived in the central village. The average distance from a village to the nearest province center - small towns where XacBank's branches and loan officers are based - is 116 kilometers. Because the distance between a village and the nearest paved road is on average 170 km, travel between villages, and between villages and province centers, is time consuming and costly.

[7]The MWF is a large NGO whose representatives worked together with XacBank and the research team to ensure a smooth implementation of the experiment. They signed up participants, facilitated group formation in the group-lending villages, provided information to loan applicants, and assisted the survey company.

[8]We use a MNT/USD exchange rate of 1,150 which corresponds to the average exchange rate during the first half of 2008.

by all household members). Similar patterns emerge when we compare expenditure levels, using data from the Mongolian statistical office or the EBRD 2006 Life in Transition Survey, or when we compare livestock ownership, a primary wealth indicator in Mongolia.

After about 30 women had signed up in each village, a detailed baseline survey was administered to all 1,148 participants during March-April 2008. Face-to-face interviews were conducted by a specialized survey firm hired by the research team and independent of XacBank. Interviews were held at a central location in each village where respondents and interviewers had sufficient time to go through the questions without interruptions. Use of a central location also minimized the risk that the female respondents would give biased answers due to the presence of older and/or male family members (as had happened during piloting). Interviews lasted approximately one hour. At the time of the baseline survey we also collected information on the main socioeconomic, demographic, and geographic characteristics of the 40 villages.

The baseline survey measured variables that reflect households' standards of living and that could be expected to change over the 1.5 year interval of the experiment. These include income, consumption, and savings; entrepreneurial activity and labor supply; asset ownership and debt; and informal transfers. In addition, information was elicited about household composition and education; exposure to economic shocks; and respondents' income expectations. The surveys also collected information on more context-specific poverty indicators such as livestock ownership and the quality and size of the dwelling, most often a *ger*.[9]

Randomization took place *after* completion of the baseline survey so that at the time of the interview, respondents did not know whether or not they would be offered a group loan, an individual loan or no loan at all. Randomization took place at the village level, with 15 villages receiving access to individual loans, 15 receiving access to group loans, while in 10 control villages XacBank did not provide loans to the participating women for the duration of the experiment. In all three types of villages XacBank continued to provide individual microloans to regular, more wealthy clients most of whom were male. Randomization across rather than within villages was chosen because it was administratively and politically easier to manage. Moreover, randomization across villages avoids the possibility that the program affects even individuals who do not receive it directly, though informal transfers and connections. We also stratified at the province level because a completely randomized design could have resulted in a situation whereby some provinces contained only treatment or control villages, which was unacceptable to XacBank. Also, to the extent that geographical or economical differences between provinces are large, we might not have been able to detect treatment differences in an unstratified design.

After randomization, group formation proceeded in the 15 group-lending villages, but not in the individual-lending and control villages. Group formation consisted of the development of internal procedures, the election of a group leader, and the signing of a group charter. Groups were formed by the women themselves, not by XacBank. A maximum of two women per group were allowed to be from the same family. Group members lived in the same village and already knew each other to varying degrees. In many cases actual group composition differed substantially

[9]A ger is a portable tent made from a wood frame and felt coverings. Its size is measured by the number of lattice wall sections (*khana*). A basic ger consists of four or five khana, with larger and less common sizes including six, eight, or ten khana. Bigger gers are a sign of wealth as they are more costly to heat. A sufficiently insulated ger has two layers of protective felt, whereas poorer households often only have one layer. Gers are sometimes surrounded by (costly) wooden fences (*hashaa*) that offer protection from the wind.

from the potential groups that were identified at the very beginning of the experiment when women had to indicate their interest (or not) to participate in the project. After a group had collected enough internal savings it could apply for its first XacBank loan. We provide detailed information on the type of loans offered in Section 3.4 below.

The 'treatment period' during which XacBank provided loans in the group and individual lending villages lasted 1.5 years - from April 2008 to September 2009 - with some variation across villages. During this period participating women in treatment villages could apply for (repeat) loans, while XacBank refrained from lending in the control villages. In October-November 2009 we conducted a follow-up survey to again measure the poverty status and economic activity of our sample of participating women. We also obtained information on how women had used their XacBank loan(s). In addition, we conducted a second village-level survey to collect information on village characteristics that may have changed, such as the prices of important consumer goods. Lastly, XacBank collected repayment information on all of its loans for the period April 2008-June 2011. In October 2011 we revisited one individual-lending and two group-lending villages for structured interviews and discussions with a number of borrowers about how they had experienced the lending programs.

3.3 Randomization

Table 1 presents a statistical comparison between the control villages and the two types of treatment villages. We compare the means of various characteristics of the villages themselves and of the respondents and their households. Treatment and control villages are very similar overall, and in particular in terms of size, number of inhabitants, distance to the nearest province center and the nearest paved road, and the prices of various consumption goods (Panel A). Panel B shows that the respondents living in the treatment and control villages are on average very similar too. We find no significant differences in household structure, informal transfers, self-employment, wage earnings, the value of the dwelling, or consumption patterns. Households are also very similar in terms of a large number of other consumption and asset-ownership measures (not shown but available upon request).

[INSERT TABLE 1 HERE]

Panel C also shows no significant differences between control and treatment villages in terms of the number and type of businesses operated by our respondents and their households. We do find, however, some differences in terms of access to finance at the household level. A majority of the households had at least one loan outstanding at the time of the baseline survey and this percentage is higher in the individual-lending villages (67 per cent) than in the control villages (56 per cent). However, conditional on having at least one loan, there are no significant differences between the treatment and control villages in the average number of loans per household, the total debt value (in absolute terms and in percent of household income), and the debt-service burden.

These figures also indicate that at the time of our baseline survey the penetration of micro-credit was already well advanced in rural Mongolia. For our purposes, however, an important question is whether households were already using their access to microcredit to finance entrepreneurial activities by our female respondents. Our baseline data show that this appears not to be the case. First, from Panel C we see that around 75 per cent of all outstanding loans were used for consumption, mainly to buy electric household appliances, instead of income generation. This picture is the same across all types of villages. Second, fewer than 20 per cent of households had invested part of their loan(s) in a business owned by the female targeted by the loan. Furthermore, while access to credit at the household level was somewhat higher in individual-lending villages, Panel C shows that the amount and percentage of funds used for female enterprises did not differ significantly between the three types of villages. In control villages households had invested on average 15 per cent of their outstanding debt in a female-run business, whereas these percentages were 11 and 10 per cent in individual and group-lending villages. These percentages, as well as the absolute amounts, do not differ significantly between control and treatment villages.

We conclude that the randomization process was successful: we find very few significant differences between treatment and control villages, despite considering a broad range of variables. The few differences that do exist are small and do not provide evidence of a systematic disparity between treatment and control villages along any particular dimension. We are therefore confident that randomization ensured absence of selection bias so that we can attribute any post-treatment differences in outcomes to the lending programs.

3.4 The loan products

The purpose of both group and individual loans was to allow women to finance small-scale entrepreneurial activities.[10]Given the focus on business creation and expansion, loans had a grace period of either two months (for loans exceeding six months) or one month (for shorter-term loans).[11] The interest rate varied between 1.5 and 2 per cent per month and was reduced by 0.1 per cent after each successful loan cycle. Other dynamic incentives included the possibility to increase the loan amount and/or maturity after each repaid loan (Table 2).

[INSERT TABLE 2 HERE]

Group-loan contracts stated that loans were based on joint liability and that XacBank would terminate lending to the whole group if that group did not fully repay a loan. Most group loans were composed of individually approved sub-loans with a maturity between 3 and 12 months depending on the loan cycle (within a group all sub-loans had the same maturity). Groups could also apply for a joint loan to finance a collective business, for instance to grow crops. The maximum size of the first loan to a group member was MNT 500,000 (USD 435). Group members had to agree among themselves who would get a loan and for what purpose. They

[10]Besides agriculture - both animal husbandry and crop growing - the main village industries are baking, wood-processing, retail activities, and felt making.

[11]Field, Pande, and Papp (2010) provide evidence from a randomized field experiment in India that indicates that a two-month grace period - instead of the regular two weeks - and the associated flexibility led to more business creation and investments but also to lower repayment rates.

then had to apply for the loan and XacBank screened each application independently.[12] If a borrower's project was deemed too risky XacBank could exclude it while the other members would still get a loan. If most projects were judged to be too risky then the total group loan was rejected. Contrary to individual loans the screening of group loans thus involved a two-stage process: first by co-borrowers and then by a XacBank loan officer.

Before applying for a loan, groups had to build up savings in a joint savings account equivalent to 20 per cent of the requested loan amount. Group members were in principle allowed to pledge assets instead of the compulsory savings although XacBank encouraged borrowers to use savings. The savings not only served as collateral but were also a means of ascertaining whether potential borrowers had sufficient financial discipline. Group leaders were responsible for monitoring and collecting loan repayments and handing them over to the loan officer on a monthly basis. There were no public repayment meetings or other mandatory meetings.[13] Groups decided themselves on the modalities of their cooperation, including the frequency of meetings (typically once per month).

Individual loans were similar to the sub-loans provided to group members, though larger on average. XacBank did not use predetermined collateral requirements but took collateral if available. As a result 91 per cent of the individual loans were collateralized, with the average collateral value close to 90 per cent of the loan amount. The maturity of individual loans ranged from 2 to 24 months, depending on the experience of the borrower and the type of business being invested in. Group loans had a somewhat shorter maturity (192 days on average) than individual loans (245 days) which reflects the smaller size of the former. Similar to group loans, individual loans did not involve any mandatory group activities such as repayment meetings.

3.5 Loan take-up

After the baseline survey XacBank started disbursing individual (group) loans in individual (group) treatment villages. All women who had signed up and expressed an initial interest in borrowing were visited by a loan officer and received a first loan after a successful screening. After 1.5 years, 54 per cent of all treatment respondents had borrowed from XacBank: 57 per cent in the group-lending villages and 50 per cent in the individual-lending villages. Although other MFIs were also lending in both the treatment and control villages during the experiment, our intervention led to a significant increase in borrowing. The probability of receiving microcredit during the experiment was 24 percentage points higher in treatment than in control villages (50 per cent of respondents in control villages versus 74 per cent in treatment villages).

We use information from the follow-up survey to better understand why a relatively large proportion of women in treatment villages did not borrow. First, the data show that of the 326 women who had initially signed up in the treatment villages but who did not get a loan during the experiment, 167 (51 per cent) never actually applied for a loan. At the time of signing up women did not know whether they would get access to an individual or a group loan (or end

[12]The loan officers were all female, between 21 and 27 years old, married with one or two children, and had completed at least a four-year university degree. They normally assess between 35 (Hentii province) and 50 (Hovsgol province) loan applications per month with an approval rate of about 90 per cent.

[13]Field and Pande (2008) randomly assign weekly and monthly repayment meetings and find that a more flexible schedule can significantly lower transaction costs without increasing defaults.

up in a control village). Some women may only have been interested in an individual (group) loan and may therefore not have applied when their village was assigned to group (individual) lending.

Second, of the non-borrowers who *had* applied for a loan, 47 per cent refused the offer made by XacBank. The main reasons stated for not taking up the loan were that the amount was too small, the interest rate too high, or the repayment schedule unsuitable. In total, about 75 per cent of the 'non-treatment' was therefore due to women who either did not apply for a loan or who applied for one but subsequently refused the offer. This leaves about a quarter of all 'untreated' women who were actually refused a loan by XacBank.

When we asked respondents during the follow-up survey why XacBank had refused them a loan, the main answers were 'too much outstanding debt' and 'insufficient collateral'. As discussed in Section 3.3, the baseline survey revealed that many households already had at least one microloan, mainly for consumption purposes. Interviews with loan officers indicated that existing debt at the household level made them hesitant to provide additional loans to female household members, even though these new loans were intended for entrepreneurial purposes rather than for consumption. At the time the Mongolian Central Bank had also become increasingly concerned about overindebtedness in rural areas. Loan officers may have been particularly conservative in lending to poorer-than-usual borrowers, despite having been explicitly instructed to do so by XacBank management.[14]

The experiment also partly coincided with the global financial crisis during which Mongolian financial institutions suffered from reduced access to foreign funding. Domestic funding constraints also tightened. The Mongolian Central Bank imposed higher reserve requirements in an attempt to stem inflation while deposit inflows were below average as herders suffered from low international cashmere prices. The confluence of these three factors made interbank liquidity dry up between March and late June 2008 and correspondingly XacBank reduced its credit supply. The year-on-year growth rate of business lending even turned negative in November 2008, not reverting to positive until July 2009.

Table 3 displays the results of reduced-form probit regressions to explain the probability of loan take-up in more detail. We find a higher probability of borrowing in group-lending villages (significant at the 10 per cent level). A closer inspection of the underlying data indicates that the higher lending probability in group-lending villages is not driven by XacBank covering some (group) villages earlier than others or by the follow-up survey being conducted earlier in individual-lending villages. Instead, demand for loans may have been lower in individual-lending villages either because the availability of microcredit was somewhat higher in the first place (see Panel C of Table 1) or because access to group loans (previously unavailable to anyone in these villages) was valued more than access to individual loans (previously available).

Interestingly, the number (or amount) of *outstanding* loans at the time of the baseline survey is not negatively associated with the probability of obtaining a loan during the experiment (for instance because households had already reached their borrowing capacity, either according to their own judgment or that of the loan officer). We do find a negative but imprecisely measured

[14]XacBank provided 375 out of 534 applicants with a loan, an approval rate of 70.2 per cent. This is below XacBank's regular approval rate, which is about 95 per cent according to its own management information system and about 90 per cent according to the answers of the loan officers during the loan officer baseline survey.

association with *previous* loans, i.e. loans that had been repaid at the time of the baseline survey. Prior use of loans could indicate borrower quality in which case one would expect a positive sign. A negative sign may indicate that previous borrowers no longer require loans, or that they were not satisfied with the loan product. Note that the prior loan variable is significantly negative in the group-village specification (when province fixed effects are included) indicating that borrowers with no or limited borrowing experience were particularly likely to participate in a group loan. This may indicate that even when individual loans are available some women may only be interested in applying for a group loan.

Lastly, we find that households who own a well, fence, or tools and machinery had a higher probability of getting a loan, either because they are more wealthy or could use these items as collateral.

[INSERT TABLE 3 HERE]

3.6 Attrition

The follow-up survey took place approximately 1.5 years after the baseline survey and 86 per cent of respondents were successfully re-interviewed. While an attrition rate of 14 per cent is relatively low, there is always the concern that non-response was not random across treatment and control villages, which could bias the estimated treatment effects. To investigate this, we estimate the probability of attrition as a function of treatment village dummies as well as a range of respondent, village, and household characteristics.

Table 4 shows that respondents in individual-lending villages are almost 7 percentage points more likely to attrit compared with those in control villages, and this is of borderline statistical significance at conventional levels (depending on the inclusion of control variables and/or province fixed effects). We detect no differential patterns in attrition between group and control villages. On further investigation, we find that the differential attrition is driven by two individual-lending villages where the wedding season was underway at the time of the follow-up survey, resulting in many respondents being away from home temporarily. We are thus reassured that the reason for higher attrition is unlikely to be related to the program, and so we retain these two villages in the analysis. While one might think that loan use might be distorted due to the wedding season, we note that we also estimate all models excluding these two villages and find that our results are robust.

[INSERT TABLE 4 HERE]

Lastly, we note that other variables have the expected association with attrition: respondents that own a fence or a well and families with more women and small children are less likely to attrit - as one would expect, given that these characteristics are generally associated with less mobility. Households that live further from the province center and/or own horses or camels are more likely to attrit, presumably because they are more likely to live a semi-nomadic lifestyle and are thus more difficult to locate for interviews. Households that experienced a recent death were less likely to participate in the follow-up survey too.

4 Methodology

In what follows, we report the results of an intention to treat (ITT) analysis where we compare *all* women who initially signed up in treatment villages, irrespective of whether they borrowed or not, with those who signed up in control villages.[15] The advantage of this conservative approach is that we can interpret the experimental intervention as a policy and learn about the impact on the population that XacBank initially targeted, and not just on those who actually borrowed. We also employed an instrumental variables (IV) methodology in which we instrument *actual* borrowing status of participants with a dummy indicating whether or not the village was randomized to be a treatment village. These IV results are very similar to the ITT findings described below and are available on request.

Results reported here use a difference-in-differences technique to compare respondents in treatment and control villages before and after the loan treatment.[16] Whilst in principle we could attribute post-treatment differences to the lending programs, we improve precision slightly when we take various baseline characteristics into account that are strong determinants of the outcome variables. All findings remain very similar if we use post-treatment data only. Our basic regression framework is:

$$Y_{ivt} = \alpha_0 + I_v \cdot (\alpha_1 + \alpha_2 \cdot F_t) + G_v \cdot (\alpha_3 + \alpha_4 \cdot F_t) + \alpha_5 \cdot F_t + \alpha_6 \cdot X_{iv0} + \epsilon_{ivt} \quad (1)$$

where:

- Y_{ivt} is the outcome variable of interest for individual i in village v at time t ($t = 0$ (1) at baseline (follow-up) survey);
- I_v is a binary variable equal to 1 for individual-lending villages (0 otherwise);
- G_v is a binary variable equal to 1 for group-lending villages (0 otherwise);
- F_t is a follow-up binary variable (0 for baseline observations);
- X_{iv0} is a set of baseline characteristics of respondents, their households, and their villages;
- ϵ_{ivt} is an i.i.d. error term clustered at the village level.

In this specification α_2 and α_4 measure the impact of the individual and group lending treatment, respectively. In addition, we also run more flexible specifications where we allow for heterogeneous impacts. We first allow for variation by education level of the respondent, which we consider to be an indicator of long-term poverty of the household:

$$Y_{ivt} = \alpha_0 + I_v \cdot (\alpha_1 + \alpha_2 \cdot F_t) + G_v \cdot (\alpha_3 + \alpha_4 \cdot F_t) + \alpha_5 \cdot F_t + H_i \cdot Z + \alpha_{12} \cdot X_{iv0} + \epsilon_{ivt} \quad (2)$$

[15] One can calculate the impact of access to microcredit on those women who actually borrowed - i.e. the average effect of the treatment on the treated (ATT) - by dividing the ITT effect by the probability of receiving treatment (57 per cent in the group-lending villages and 50 per cent in the individual-lending villages). A caveat is that this may not generalize, as those who receive the treatment may be systematically different from those who do not. As the (heroic) assumption underlying consistent estimation of ATT is that unobservable characteristics do not affect the decision to participate, we only show ITT parameters.

[16] We estimate using OLS for continuous dependent variables, a probit model for binary dependent variables, and a tobit model for dependent variables that are censored at zero.

where

$$Z = \alpha_6 + I_v \cdot (\alpha_7 + \alpha_8 \cdot F_t) + G_v \cdot (\alpha_9 + \alpha_{10} \cdot F_t) + \alpha_{11} \cdot F_t$$

and H_i is one for individuals with a high education level (grade 8 or higher, or vocational training) and zero for individuals with a low education level (less than grade 8). All other variables are as previously defined.

Second, because respondents in some villages received more loans than in others and for longer periods of time, we also analyze the impact of treatment intensity over and above the basic impact of access to credit. We allow impact to vary by treatment intensity Int_v at the village level, either measured as the average number of loans ($Number_v$) or as the average number of months between the date when the first respondents in a village received a loan and the follow-up survey ($Months_v$):

$$Y_{ivt} = \alpha_0 + I_v \cdot (\alpha_1 + \alpha_2 \cdot F_t + \alpha_3 \cdot Int_v) + G_v \cdot (\alpha_4 + \alpha_5 \cdot F_t + \alpha_6 \cdot Int_v) + \alpha_7 \cdot F_t + \alpha_8 \cdot X_{iv0} + \epsilon_{ivt} \quad (3)$$

where α_3 and α_6 give the additional effect of treatment intensity in individual-lending and group-lending villages, respectively.

We measure treatment intensity at the village level to avoid endogeneity problems: more motivated and entrepreneurial individuals may make sure to get exposed to the lending program early on, which would lead us to erroneously attribute the effect of these borrower characteristics to early treatment. We should stress that the intensity of the program was not purposely varied in a random fashion among the treatment sample. One should therefore interpret with caution the results obtained estimating equation (3), as the intensity of the program might vary with unobserved village and/or individual characteristics and induce biases in the estimation of the coefficients of this equation. Having said that, numerous conversations with XacBank officials make us believe that the variation in intensity of the program across villages was by and large induced by administrative quirks and is unlikely to be endogenous.

The mean number of months between the date when the first respondents in a village received a loan and the date of the follow-up survey is 5.2 months (6.3 months in group-lending villages, 4.2 months in individual-lending villages) with a standard deviation of 2.7 months. The mean number of loans received is 0.78 (0.99 in group-lending villages, 0.57 in individual-lending villages) with a standard deviation of 0.48. This indicates that not only is the probability of borrowing higher in group villages, but so also is the intensity of the treatment.

5 Results

5.1 Loan use

We first provide a picture of what borrowers reported having used their loans for. Table 5 shows that women used the individual and group loans in very similar ways. Assuming that the purchase of livestock, tools, and machinery are business expenses, we find that 67 (66) per cent of group (individual) borrowers used their first loan mainly to invest in a new or existing enterprise, putting between 70 and 80 per cent of the loan to this purpose, with the remainder being used for household expenses. In the case of second loans, fewer women - 43 (51) per cent of the group (individual) borrowers - used the loan primarily for business purposes.

[INSERT TABLE 5 HERE]

We can also compare what women reported as the purpose of the loan at baseline and at follow-up. When we do this, we find that 86 (93) per cent of group (individual) borrowers who at follow-up stated that they had used their loan(s) mainly for business purposes, had consistently indicated at the start of the experiment that they would use the loan for entrepreneurial activities. However, 82 per cent of women in both types of treatment villages who used the loan mainly for consumption had reported at baseline that they would use it to invest in a business. We cannot say whether they intentionally misreported at baseline (as the loans were marketed as business loans) or whether they later on changed their minds.

5.2 Impact of the microcredit programs

A key objective of the microcredit programs was to encourage women to expand or invest in small-scale enterprises, with the ultimate aim of reducing poverty and improving well-being. To evaluate the extent to which the programme achieved these two objectives, we first look at the effect on enterprise creation and growth, and on whether enterprise profits increased. We then go on to estimate its effect on detailed household consumption, as a measure of well-being. To pre-empt, we find evidence of households in group villages increasing investment in enterprises, and corresponding increases in consumption. We detect no systematic effects in individual villages.

5.2.1 Did the programs affect business creation and growth?

As discussed, one of the main intermediate objectives of the programs was to encourage women to invest in new or existing small-scale enterprises. We have seen some suggestive evidence that this was the case, with a large majority of women reporting having used a substantial part of their loan(s) to invest in working capital and fixed assets. In this section we estimate the effect on business creation and growth. Table 6 shows estimates from equation (1) through (3). The odd (even) columns show the impacts for group (individual) loans.

We first estimate the basic impact using equation (1), and then estimate heterogeneous impacts by education level (equation (2)) and treatment intensity (equation (3)). Treatment intensity is measured as the number of borrowing months or as the number of loans, and is in both cases the average at the village level. In line with equation (3) the intensity effects measure the impact of longer *actual* exposure to loans over and above the basic ITT effect. We use the same estimation approach for the other outcome variables. All regressions include a standard set of baseline respondent and village-level covariates (listed in Table A1 in the Annex) and our results remain robust to the exclusion of these covariates.

[INSERT TABLE 6 HERE]

Columns (1) and (2) show the impact of access to microcredit on the probability that the household operates a small-scale business, whether the respondent's own one, her partner's, or their joint one (65 per cent of respondents are married or cohabitating). Columns (3) and (4) show similar regressions but specifically for the respondent's own enterprise. We see that access to group loans has a significant positive impact on female entrepreneurship and this effect is largely driven by less-educated women (see row II). At the end of the experiment, these women had a 29 per cent higher chance of operating a business compared with women in the control villages. This difference is 10 per cent for highly educated women.[17] Rows III and IV show that a large part of these effects is driven by women who had been exposed to (repeat) loans for a longer period of time.

The results for access to individual loans are less strong. Columns (2) and (4) indicate no impact on female entrepreneurship, although there is a positive impact on total entrepreneurship over time (row III). This latter effect is driven by joint enterprises which become more prevalent in individual-lending compared with control villages. In individual-lending villages where respondents borrowed on average for six months, the probability that a household operates any type of business is 12 percentage points higher than in the control villages. Interestingly, the nature of the businesses operated by women themselves and those operated jointly with their spouses differ. The former are mostly sewing businesses and small-scale retail activities whereas the latter comprise mainly animal husbandry and crop production.

Figure 1 depicts how the *actual* loan exposure at the village level influences entrepreneurship (for a typical respondent with average covariate values). The left-hand (right-hand) panels show individual- (group)-lending villages. The upper panels focus on the likelihood that women run their own business, whereas the lower panels indicate the probability that households operate any kind of business. The starting point of each graph indicates the probability of business ownership for the average respondent in treatment villages where in practice virtually no XacBank lending took place. Due to the randomization these values do not differ significantly between both types of treatment villages nor do they differ from the values in the control villages (where XacBank did not lend *by design*). The graphs then show similar point estimates, surrounded by a 95 per cent confidence interval, for the probability of business ownership in treatment villages where the actual average exposure was 2, 4, 6, 8, 10, or 12 months.

While in all four graphs the probability of business ownership increases with loan exposure, the confidence intervals are narrowest for female enterprises in group-lending villages and for all enterprises in individual-lending villages. For example, a typical respondent in a group-lending village where respondents were only exposed to credit for a few days, had a 36 per cent probability of operating her own enterprise (the same as in a control village). A similar respondent in a group-lending village where respondents had been borrowing for a full 12 months had a 53 per cent probability of running a business. This 53 per cent is outside the 95 per cent interval surrounding the point estimate of 36 per cent for respondents in relatively less treated villages. These results mirror those in Table 6: female enterprises became more prevalent in group-lending villages (compared with the control villages) whereas in individual-lending villages there was a gradual and significant increase in the number of businesses operated jointly by borrowers and their spouses.

[17]This also translates into a higher likelihood of operating any type of enterprise (column (1)). Unreported regressions show that there is no strong impact of access to group loans on enterprise ownership by, or jointly with, the borrower's partner. The effect in column (1) is thus driven by an increase in female entrepreneurship.

[INSERT FIGURE 1 HERE]

Columns (5) to (8) in Table 6 analyze whether access to credit resulted in more profitable enterprises. Even though enterprise profitability decreased in both treatment and control villages between the baseline and follow-up surveys, mainly due to the economic crisis, access to credit seems to have partly shielded borrowers from this impact. Columns (5) and (7) show that over time and after repeat borrowing, enterprises in group-lending villages were significantly more profitable than those in control villages. After half a year of exposure to credit, the difference in yearly profitability amounts to over 200,000 tögrög, or almost one third of the average annual enterprise profits at baseline. We find a similar positive impact on business profits in individual-lending villages, although here again the impact is mainly due to enterprises that are operated jointly with the borrower's partner.

Lastly, we look at whether households increased labour supply in line with this increased business creation. About a quarter of respondents were employed in wage activities at the time of the baseline interview and they received an average wage of MNT 130,000 (USD 113) per month. During the experiment the share of wage employment remained unchanged and there was a marked drop in salary levels, most likely due to the global crisis. We find no clear impact of the programs on total labor supply or income at the household level, nor do we find an impact when we split labor supply into wage labor and hours worked in own enterprises (Table 7). There is weak evidence (at the 10 per cent significance level) that over time group borrowers work less for a wage, which would be in line with the increase in female self-employment. We do not find a significant impact on enterprise labor for these group borrowers though. In contrast, there is some evidence that households in individual-lending villages start to work more in enterprises over time, in line with the evidence on gradual (joint) enterprise creation. Despite these impacts we do not find any significant effect on overall household income (or on wage income and income from benefits separately).

[INSERT TABLE 7 HERE]

5.2.2 Did household well-being increase? The impact on consumption and asset ownership

In order to assess whether borrowers' increased engagement in entrepreneurial activities fed through to improving household well-being - a key objective of the program - we next estimate the effects of the program on household consumption. We use detailed information on consumption patterns elicited in the surveys, in which food consumption is measured over the past week (at a disaggregate level as well as overall), and non-durable and durable consumption over the past month and year, respectively.

Interestingly, we find robust evidence that access to group loans led to more and healthier food consumption, in particular of fresh items such as fruits, vegetables and dairy products (Table 8). With the exception of dairy these effects are not only due to increased home production: we also see treated clients purchasing more. The probability that a household consumed dairy products, fruits and vegetables, and non-alcoholic drinks in the last week was 5, 10 and 13 percentage points higher in group-lending than control villages. Total food consumption was 17 percentage points higher. To put this into context, the average loan per borrower in group-lending villages is 300 USD and the average monthly pre-treatment food consumption in group-lending (and control) villages was 108 USD per household. So the estimated effect implies that over time food consumption increased by 19 USD more per household in group villages, i.e. 6.3 per cent of the loan amount. Over time we also see an increase in the use of combustibles and additional felt for ger isolation as well as other non-durable and total consumption. In line with Banerjee et al. (2010) we find a negative impact on the probability of smoking and the amount spent on cigarettes, a typical temptation good.

In contrast to households in group-lending villages, households in individual-lending villages do not experience much change in their consumption as a result of access to credit. We do not find any effects on aggregate consumption and expenditure variables - not even with increased exposure to treatment.

[INSERT TABLE 8 HERE]

Our evidence on consumption and business creation somewhat contrasts with recent evidence from other microcredit field experiments, such as Banerjee et al. (2010) in India and Augsburg et al. (2011) in Bosnia, who find that clients who start new businesses reduce consumption, at least in the short run and probably to be able to finance the new business. Our results could be explained by the fact that our follow-up survey is conducted 18 months after the start of the program and after several loan cycles. This would imply that the women who did start a new business might be already reaping the returns and the higher (permanent) income of such an activity.

We also consider whether asset ownership increased, and find evidence that overall asset wealth does increase over time in group-lending villages, but not in individual-lending villages - see Table 9. In particular, we detect a significant increase in the ownership of VCRs, radios, and large household appliances for both treatment types. At the end of the experiment the probability of owning a VCR or radio was 17 and 14 per cent higher in the group and individual-lending villages, respectively. For large household appliances the corresponding figures are 9 and 7 per cent.

[INSERT TABLE 9 HERE]

In unreported regressions we do not find a robust impact of access to either type of loan on the likelihood of owning the main dwelling or on the value of this house or ger. There is thus no evidence that loans encouraged borrowers to buy new property or invest in their existing main property. However, in columns (5) and (7) we do find some evidence that less-educated women in group-lending villages disinvest in second gers, land, and vehicles. This may indicate that less-educated women sold some of these assets in order to combine the proceeds with the loan amount and invest in small-scale businesses (see Section 5.2.1). In line with this interpretation, the results in column (17) show that these women are 30 per cent more likely to own tools at the end of the experiment, which closely matches the 29 per cent higher chance of operating a business (Table 6). Over time we document an increase in unsold stock and raw materials, cattle, and riding equipment in group-lending villages, again in line with an expansion of business activity.[18]

We find fairly similar results for individual-lending villages: over time a reduction in second houses and an increase in the ownership of land and second gers. We also find an increase, relative to control villages, in the ownership of VCRs/radios, large household appliances and also of televisions (over time). Lastly, there was a gradual increase in the ownership of tools in the individual-lending villages, in line with the increase in the (general) business activity that we document for these villages in Table 6 and Figure 1.

5.2.3 Do the programs crowd out transfers?

The results just shown paint a different picture of the impact of the program in group and individual villages, with evidence that the group loans were relatively more effective at achieving their objectives. One interesting question is the extent to which interpersonal transfers are affected by the programs, and whether they are affected differently in group and individual villages: as in many developing countries, access to informal credit/transfers from friends and family is important in Mongolia, in particular for women (National Statistics Office, 2006). Kinship and social networks are confined to relatively small groups of people as they derive from the traditional *khot ail* support system in which a small number of nomadic households travelled, camped, and herded together for one or more seasons (Enkhamgalen, 1995). Within *khot ail* and similar social networks rural Mongolians often share income from entrepreneurial activities as well as pensions and other allowances.

Access to formal credit may have changed informal lending and transfer behaviour in two different ways. On the one hand, the increased availability of formal credit in treatment villages may have strengthened informal support networks as additional funds could be shared. On the other hand, informal networks may have weakened as borrowers substitute formal for informal credit, thereby crowding out insurance systems based on implicit reciprocal agreements.

The survey asked households about their informal - monetary and in-kind - transactions with friends and family during the past year and the most recent month. Although we do not find an overall ITT effect of either lending program on informal transfers, we document that over time group borrowers received less transfers both from friends and family members (Table 10). They were also less likely to make transfers to friends. Those that had been exposed to group loans for at least six months were 6 percentage points less likely to receive transfers from friends, 14 percentage points less likely to receive transfers from family, and 8 percentage points less likely to make transfers to friends.

[INSERT TABLE 10 HERE]

Interestingly, we find opposite effects in individual-lending villages. For individuals exposed to more loans and over a longer period of time, we detect an increase in the probability of making transfers to and receiving transfers from friends during the past year. We also find an increase

[18] We do not find a significant increase in the total number of animals as measured by the number of standardized Mongolian livestock units or bod (one horse, yak, or cattle equals one bod; one camel equals 1.4 bod; one sheep equals 1/6 bod; and one goat equals 1/7 bod).

in such transfers to and from family members over the past month. The relationship between the intensity of exposure to credit and the probability of receiving or giving transfers is shown in Figures 2a and 2b.

[INSERT FIGURES 2A AND 2B HERE]

These results may indicate that group borrowers partly substitute their informal networks with the formal network of the borrowing group. The associated discipline may make them less amenable to use part of their loans to help friends and family smooth consumption. In contrast, individual borrowers increase their informal financial transactions with friends and family, perhaps using part of their new loan to help others out.

Such an interpretation would be in line with recent evidence for Sri Lanka and Ghana by De Mel, McKenzie and Woodruff (2009) and Fafchamps, McKenzie, Quinn, and Woodruff (2011), respectively. The latter paper finds that women who received cash transfers did not increase their business profits as large portions of the cash grants ended up in household consumption and, to a lesser extent, transfers to others. Self-control problems, i.e. borrowers' inability to commit themselves to invest large parts of the cash grants into their enterprises and to resist the temptation to spend money on competing demands, including from friends and family, were a core explanation for the ineffectiveness of cash grants. Our results are also in line with Karlan and Zinman (2011) who find that individual-liability loans may increase access to informal credit from friends and family in the case of emergencies. Lastly, our finding that cigarette consumption increased far less in group-lending villages than in control villages, may reflect similar mechanisms. Just like group discipline can reduce the temptation to pass on part of the new loan to friends and family, it may also reduce spending on temptation goods (see also Banerjee and Mullainathan, 2010).

5.3 Repayment

In the preceding sections we documented a positive impact of access to group loans on consumption and business activities as well as some weaker effects of access to individual loans on business activity. In this section, we analyze the repayment behavior of both types of borrowers. Giné and Karlan (2010) also compare repayment rates between group and individual lending programs - both with mandatory weekly repayment meetings - and find no significant differences. In contrast, Carpena, Vole, Shapiro and Zia (2010) find that joint liability is associated with better loan repayment.

To construct our repayment data we use monthly reporting files that XacBank compiled on the basis of its administrative software. These files contain for each borrower the loan amount, interest rate, disbursement and due dates, loan purpose, collateral, overdue principal and interest, paid penalties as well as whether the client defaulted on the loan (defined as customers that were at least 90 days late in repaying one or more loan installments).

Table 11 presents probit regressions to explain the probability of loan default. The dependent variable is a dummy that indicates whether a borrower defaulted ('1') or not ('0'). The first two columns are based on a sample of first-time XacBank loans whereas the third and fourth columns are based on the full sample that includes repeat loans.

[INSERT TABLE 11 HERE]

We find, regardless of whether we control for borrower and loan characteristics, no difference between the probability of default in group-lending and individual-lending villages. This confirms the findings of Giné and Karlan (2010) although in our case *neither* loan program included mandatory repayment meetings whereas in their experiment *both* programs included such meetings.

The covariates in columns (2) and (4) give additional information on the borrower and loan characteristics that influence default probability. While the size of the loan does not influence the likelihood of repayment, there is a negative impact (at the 10 per cent significance level) of the amount of outstanding debt at the time of the baseline survey. Respondents with outstanding debt at baseline where thus *more* likely to (be able to) repay the subsequent XacBank loan. Borrowers that had already successfully passed the screening of another bank, where less risky compared with first-time borrowers.[19] In addition, column (4) indicates that repeat borrowers were significantly less risky, possibly because they had already successfully passed XacBank's own screening procedures and subsequently paid on time. For both first-time and repeat loans we also find that as loans mature (increasing number of months since disbursement) the risk of default increases, all else equal (see also Carpena et al., 2010).

Interestingly, a number of covariates are only of importance for first-time loans. Those that owned land or operated an enterprise at baseline were less risky borrowers as were the relatively highly educated. Ownership of a TV at baseline increased the risk of default, perhaps because this identifies women who use(d) debt for consumptive purposes. None of these variables is statistically significant at the 5 per cent level in the regression based on the whole loan sample (column 4). For repeat borrowers these variables are less important compared to the information that is contained in the variable that measures the number of successful previous loans with XacBank during the experiment.

Lastly, in unreported regressions we look at interaction effects between the liability structure and the number of previous loans of the borrower. We find no evidence for such a differentiated impact of repeat borrowing under the two programs. We also try other interaction terms but none of these is statistically significant, implying that there is no apparent heterogeneity between group and individual borrowers in terms of their repayment behavior.

6 Conclusions

We present results from a randomized field experiment in rural Mongolia where group-lending and individual-lending programs were randomly introduced across villages. The aim of the study was to measure and compare the effectiveness of these two types of microcredit in reducing poverty – a topic that still lacks unequivocal evidence, in particular for rural settings. While

[19] To the extent that multiple borrowing and overindebtedness were a problem in rural Mongolia this is not picked up by our default analysis. The fact that we do not find differences in repayments rates does not imply, however, that borrowers with initial debt did not experience any difficulties; it just shows that in the end they managed to repay as well as first-time borrowers. High repayment rates can point to successful projects with high returns but may also mask underlying problems where borrowers need to borrow from other sources or sell assets in order to be able to repay.

earlier papers have separately assessed the poverty impact of group lending (Banerjee et al., 2010) and individual lending (Karlan and Zinman, 2011) this is the first field experiment to compare both in the same (rural) setting.

Our findings on the poverty impact of different modes of microcredit are mixed. In line with previous studies, we document that participants in both programs used part of their loans to acquire assets – VCRs, radios, and large household appliances. A second finding that holds for both treatment programs is that women with lower education seem to benefit more from the intervention than women with higher education. We interpret the level of education as a proxy poverty measure, more reliable than a wealth indicator given that it is not affected by the program and is more stable over time. The results therefore suggest that it is the poorer part of the targeted population that benefits more from the microcredit intervention, independent of how it is being delivered.

For group loans we also find a positive impact on food consumption and entrepreneurship though not on current income. Enterprise profits increase over time as well. Among households that were offered group loans the likelihood of owning an enterprise increases by ten percentage points more than in control villages (and even close to 30 percentage points for less-educated women).

Our findings for individual lending are weaker. We find no significant increase in consumption or income although over time there is an increase in the probability that women operate a business jointly with their spouse. Over time these joint enterprises, which engage in different types of activities compared with the female-operated enterprises in group-lending villages, also become more profitable. More generally, we find that effects observed for group borrowers are also experienced by women in individual-lending villages if they are exposed to credit for longer periods of time. For example, their likelihood of starting a business is higher the longer they have access to loans. Nevertheless, it is not clear whether these longer-term effects will translate in the same way as they do for group clients. For instance, we find no evidence that food consumption goes up with exposure in individual-lending villages.

Importantly, we find no difference in repayment rates between the two lending programs, both of which did not include weekly repayment meetings. This casts doubt on the hypothesis that microcredit repayment rates are high mainly due to the effect of weekly group meetings. Our results indicate that, at least in our context, even without such regular meetings group and individual microcredit can have similar and high repayment rates (also note that both our loan products required some form of collateral).

There is at this stage no evidence on changes in income as a result of either of the programs, though it may be too early for such effects to be observed. The more sustained and more generalized increase in consumption (of both non-durable consumption and the service of durable items) in group-lending villages seems to indicate that these loans are more effective at increasing the permanent income of households. It would be interesting to test this hypothesis further by considering long-run income levels.[20]

If one were to take at face value the evidence on the stronger impact of group loans, one would

[20] There might also be a measurement issue. In developing countries income is notoriously harder to measure than consumption and might be more affected by measurement error, therefore making the detection of relatively small impacts harder.

want to ask *why* such loans are more effective at raising consumption (and probably long-term income) than individual loans. One possibility is that the joint-liability scheme better ensures discipline so that larger long-run effects can be achieved.[21] Group discipline may not only prevent the selection of overly risky investment projects, it may also ensure that a substantial part of the loans is actually invested in the first place (instead of used for consumption or transfers to others). Our results on informal transfers can be interpreted to support this hypothesis: we find that women in group-lending villages decrease their transfer activities with families and friends, opposite to what we find in individual-lending villages. This could reflect that groups replace some of their informal financial networks but further analysis is needed to explore this. Such an analysis would also be important to assess the welfare impact of access to group loans for the borrowers as well as their friends and families. Increased within-group financial discipline may come at the cost of disrupting informal credit and insurance systems based on kinship and other social ties.

Lastly, to some extent our weaker results for individual loans may also reflect that borrowing at baseline was somewhat higher in individual-lending villages compared with group-lending villages. Moreover, since group-lending was an innovative way of lending in the Mongolian context, the unmet demand for such a product - and consequently its marginal impact - may have been higher. Loan take-up was indeed higher in group-lending villages. This could indicate that some women, in particular the less-educated, had not been comfortable with borrowing on an individual basis but *were* willing to borrow within the framework of a group. This would imply that group and individual lending are complementary financial services for which the demand may differ across borrower types. The continuing process of liability individualization by MFIs may therefore run the risk that certain borrowers, those that are not able or willing to borrow and invest on their own, may gradually lose access to formal financial services.

References

Armendáriz, B. and J. Morduch (2005), *The Economics of Microfinance*, MIT Press, Cambridge, MA.

Augsburg, B., R. De Haas, H. Harmgart and C. Meghir (2011), Microfinance at the margin: Experimental evidence from Bosnia, mimeo.

Banerjee, A., T. Besley, and T. Guinnane (1994), Thy neighbour's keeper: The design of a credit cooperative with theory and a test, *Quarterly Journal of Economics*, 109(2), 491-515.

Banerjee, A., E. Duflo, R. Glennerster, and C. Kinnan (2010), The miracle of microfinance? Evidence from a randomized evaluation, BREAD Working Paper No. 278.

Banerjee, A. and S. Mullainathan (2010), The shape of temptation: Implications for the economic lives of the poor, MIT Working Paper.

Bateman, M. (2010), *Why Doesn't Microfinance Work? The Destructive Rise of Local Neoliberalism*, Zed Books, London/New York.

Bhole, B. and S. Ogden (2010), Group lending and individual lending with strategic default, *Journal of Development Economics*, 91(2), 348-363.

Bruhn, M. and I. Love (2009), The economic impact of banking the unbanked, World Bank Policy Research Paper No. 4981, World Bank, Washington D.C.

Besley, T.J. and S. Coate (1995), Group lending, repayment incentives and social collateral, *Journal of*

[21] The savings requirement of the group product may also have helped to select disciplined borrowers.

Development Economics, 46(1), 1-18.
Carpena, F., S. Cole, J. Shapiro and B. Zia (2010), Liability structure in small-scale finance. Evidence from a natural experiment, World Bank Policy Research Working Paper No. 5427, World Bank, Washington D.C.
Crépon, B., F. Devoto, E. Duflo, and W. Parienté (2011), Impact of microcredit in rural areas of Morocco: Evidence from a randomized evaluation, mimeo.
Cull, R., A. Demirguç-Kunt, and J. Morduch (2009), Microfinance meets the market, *Journal of Economic Perspectives*, 23(1), 167-92.
De Mel, S., D. McKenzie and C. Woodruff (2009), Are women more credit constrained? Experimental evidence on gender and microenterprise returns, *American Economic Journal: Applied Economics*, 1-32.
Duflo, E., R. Glennerster and M. Kremer (2008), Using randomization in development economics research: A toolkit, In: *Handbook of Development Economics*, T. Paul Schultz and J. Strauss (ed.), Chapter 61, Elsevier, 3895-3962.
Fafchamps, M., D. McKenzie, S. Quinn, and C. Woodruff (2011), When is capital enough to get female microenterprises growing? Evidence from a randomized experiment in Ghana, World Bank Policy Research Working Paper No. 5706, World Bank, Washington D.C.
Field, E. and R. Pande (2008), Repayment frequency and default in micro-finance: Evidence from India, *Journal of the European Economic Association*, 6(2-3), 501-9.
Field, E., R. Pande, and J. Papp (2010), Does microfinance repayment flexibility affect entrepreneurial behavior and loan default?, mimeo.
Fischer, G. (2010), Contract structure, risk sharing, and investment choice, London School of Economics, mimeo.
Gangopadhyay, S., M. Ghatak and R. Lensink (2005), Joint liability lending and the peer selection effect, *Economic Journal*, 115(506), 1005-15.
Ghatak, M. (1999), Group lending, local information and peer selection, *Journal of Development Economics*, 60(1), 27-50.
Ghatak, M. (2000), Screening by the company you keep: Joint liability lending and the peer selection effect, *Economic Journal*, 110, 601-631.
Ghatak, M. and T. Guinnane (1999), The economics of lending with joint liability: A review of theory and practice, *Journal of Development Economics*, 60, 195-228.
Giné, X., P. Jakiela, D. Karlan, and J. Morduch (2010), Microfinance games, *American Economic Journal: Applied Economics*, 2, 60-95.
Giné, X. and D. Karlan (2010), Group versus individual liability: Long-term evidence from Philippine microcredit lending groups, mimeo.
Enkhamgalan, A. (1995), *The essential problems of Mongolian pastoral reform*, Institute of Agrocultural Economics, Ulanbaataar.
Kaboski, J.P. and R.M. Townsend (2005), Policies and impact: An analysis of village-level microfinance institutions, *Journal of the European Economic Association* 3(1), 1-50.
Kaboski, J.P. and R.M. Townsend (2011), A structural evaluation of a large-scale quasi-experimental microfinance initiative, *Econometrica*, 79(5), 1357-1406.
Karlan, D. and J. Zinman (2011), Microcredit in theory and practice: Using randomized credit scoring for impact evaluation, *Science*, 332(June), 1278-1284.
Khandker, S. (2005), Microfinance and poverty: Evidence using panel data from Bangladesh, *World Bank Economic Review*, 19(2), 263–86.
Laffont, J.-J. and P. Rey (2003), Moral hazard, collusion, and group lending, Working Paper No. 122, Institut d'Économie Industrielle (IDEI), Toulouse.
Mongolian National Statistics Office, Asian Development Bank, and World Bank (2006), *Participatory Poverty Assessment Mongolia*, Ulaanbataar.

Morduch, J. (1998), Does microfinance really help the poor? New evidence on flagship programs in Bangladesh, Princeton University Working Paper.

Morduch, J. and D. Roodman (2009), The impact of microcredit on the poor in Bangladesh: Revisiting the evidence, Center for Global Development Working Paper No. 174.

Pitt, M. and S. Khandker (1998), The impact of group-based credit programs on poor households in Bangladesh: Does the gender of participants matter?, *Journal of Political Economy*, 106(5), 958-98.

Stiglitz, J. (1990), Peer monitoring and credit markets, *World Bank Economic Review*, 4(3), 351-366.

Wydick, B. (1999), Can social cohesion be harnessed to repair market failures? Evidence from group lending in Guatemala, *Economic Journal*, 109, 463-475.

Table 1. Randomization and treatment-control balance

This table provides t-test results for means comparisons of household and village characteristics in individual-lending versus control villages and in group-lending versus control villages. P-values are reported between brackets. ***, **, * denote significance at the 0.01, 0.05 and 0.10-level. In case of household characteristics, the standard errors are clustered at the village level. Table A1 provides the definitions and sources of all variables. *N* indicates the number of villages (Panel A) or respondents (Panel B and C) for whom information about a given variable is available. *Conditional N* indicates the number of respondents for whom the value of the respective variable is strictly positive in the case of conditional variables. E.g. 1,148 women answered the survey question about wage earnings and 266 of them reported positive wage earnings.

Panel A. Village and district characteristics

	People in village	People in district	District area	Livestock in district	Banks in district	SCCs in district	Distance to paved road	Time to paved road	Distance to province center	Time to province center	Price		
											Milk	Mutton	Bread
Control	1,017	3,530	2,823	128,747	1.7	0.6	185	220	113	218	628	2,967	1,035
Treatment	1,136	3,961	3,415	167,728	2.2	0.7	165	272	117	200	797	2,833	790
P-value	(0.35)	(0.63)	(0.24)	(0.08)*	(0.13)	(0.55)	(0.73)	(0.64)	(0.82)	(0.7)	(0.19)	(0.53)	(0.25)
N	40	40	40	40	40	40	29	24	39	36	39	33	39

Panel B. Household characteristics: general, consumption, assets

	Children <16	Age respondent	Education respondent	HH death	Received transfers	Given transfers	Self-employed	Wage earnings	Value of dwelling	Consumption			
										Milk	Red meat	Vegetables	Fuel
Control	1.5	40.4	9.3	6.0	155	241	32.4	29.4	1.43	3.4	5.4	2.2	22.8
Individual	1.6	38.9	9.4	6.4	174	153	33.4	31.8	1.52	4.0	5.2	2.0	18.9
P-value	(0.65)	(0.16)	(0.66)	(0.84)	(0.73)	(0.17)	(0.78)	(0.39)	(0.71)	(0.32)	(0.78)	(0.57)	(0.42)
Group	1.6	39.7	9.6	5.1	196	158	33.5	30.1	1.57	3.2	5	2.0	23.3
P-value	(0.82)	(0.48)	(0.38)	(0.58)	(0.73)	(0.21)	(0.76)	(0.79)	(0.55)	(0.86)	(0.54)	(0.45)	(0.93)
N	1,148	1,147	1,143	1,147	1,147	1,147	1,148	1,148	1,147	1,146	1,139	1,143	1,055
Conditional N					103	174		266					

Panel C. Household characteristics: entrepreneurship and borrowing

	Operates business	Female business	Hours hired	At least one loan	Outstanding loans	Debt value	Debt/HH income	Debt service	Interest rate	Secured loans	Percentage private use	Percentage female business	Amount female business
Control	58.9	64.8	40.9	56	2.6	1.7	0.9	31.7	2.2%	73%	72%	15%	158
Individual	59.8	62.6	54.1	67	2.7	2.0	0.9	45.1	2.1%	77%	74%	11%	140
P-value	(0.88)	(0.71)	(0.40)	(0.00)***	(0.48)	(0.44)	(0.24)	(0.07)*	(0.43)	(0.44)	(0.73)	(0.13)	(0.71)
Group	60.3	59.3	35.1	62	3.0	1.9	1.1	40.8	2.3%	73%	79%	10%	140
P-value	(0.80)	(0.31)	(0.74)	(0.13)	(0.25)*	(0.53)	(0.27)	(0.29)	(0.53)	(0.95)	(0.13)	(0.07)*	(0.71)
N	1,148	1,148	1,148	1,148	1,148	1,148	1,148	1,148	1,148	1,148	1,148	1,148	1,148
Conditional N		686			591	584	553	518	553	615	614	714	714

Table 2. The loan products

This table describes the main characteristics of the individual and the group loan products. Average loan size is calculated conditional on having a loan. Average loan size of group loans refers to loans per borrower not per group. Loans were disbursed in tögrög not US$. Source of data on maturities and loan size: XacBank.

	Individual loans	Group loans
Progressive?	Yes: larger loans, lower interest rate, and longer maturity after each successfully repaid loan	
Monthly interest rate	1.5% to 2%	
Grace period	One or two months depending on loan maturity	
Repayment frequency	Monthly, no public repayment meetings. In case of group loans, the group leader collects and hands over repayments to the loan officer	
Liability structure	Individual	Joint
Collateral	Yes but flexible approach	Joint savings (20% of loan) sometimes supplemented by assets
Available maturity	2 to 24 months	3 to 12 months
Average maturity 1st loan	224 days	199 days
Average maturity 2nd loan	234 days	243 days
Average size 1st loan	US$ 411	US$ 279
Average size 2nd loan	US$ 472	US$ 386

Table 3. Loan take-up

This table presents probit regressions to explain the probability of loan take-up in the individual and group lending villages. Standard errors are reported in brackets. ***, **, * denote significance at the 0.01, 0.05 and 0.10-level. Table A1 provides the definitions and sources of all variables.

	All villages		**Group villages**		**Individual villages**	
	(1)	(2)	(3)	(4)	(5)	(6)
Group village	0.120*	0.120*				
	(0.0692)	(0.0638)				
Outstanding loans	-0.00414	-0.00207	-0.0525	-0.00377	0.0457	0.0349
	(0.0296)	(0.0285)	(0.0377)	(0.0393)	(0.0386)	(0.0407)
Prior loans	-0.00566	-0.00899	-0.00760	-0.0130**	-0.00335	-0.00488
	(0.00738)	(0.00777)	(0.00650)	(0.00569)	(0.0155)	(0.0164)
Highly educated	0.0435	0.0309	-0.0526	-0.0774	0.111*	0.110*
	(0.0577)	(0.0559)	(0.0982)	(0.0948)	(0.0608)	(0.0637)
Owns dwelling	0.0778	0.0887	0.0961	0.131	0.0431	0.0565
	(0.0730)	(0.0743)	(0.137)	(0.149)	(0.0792)	(0.0854)
Owns fence	0.0946**	0.0690	0.195***	0.0968*	0.00530	0.0249
	(0.0458)	(0.0424)	(0.0649)	(0.0543)	(0.0521)	(0.0504)
Owns well	0.142***	0.109**	0.109	0.145**	0.163***	0.0711
	(0.0547)	(0.0535)	(0.0829)	(0.0712)	(0.0505)	(0.0627)
Owns vehicle	-0.00679	-0.0234	0.00294	-0.00606	-0.00793	-0.0371
	(0.0419)	(0.0401)	(0.0602)	(0.0530)	(0.0576)	(0.0574)
Owns tools/machinery	0.0793*	0.128***	0.0268	0.117**	0.124**	0.148***
	(0.0405)	(0.0344)	(0.0522)	(0.0464)	(0.0528)	(0.0455)
Owns animals	0.00364	-0.0193	-0.0250	-0.0746*	0.0273	0.0366
	(0.0415)	(0.0408)	(0.0354)	(0.0393)	(0.0741)	(0.0707)
HH death	-0.0223	-0.0307	-0.153	-0.141	0.0716	0.0625
	(0.0789)	(0.0816)	(0.110)	(0.115)	(0.105)	(0.110)
Province fixed effects?	No	Yes	No	Yes	No	Yes
Observations	830	830	397	397	433	433
Pseudo R-squared	0.03	0.06	0.04	0.13	0.03	0.06

Table 4. Attrition

This table presents probit regressions to explain the probability of non-participation in the follow-up survey. P-values are reported in brackets. ***, **, * denote significance at the 0.01, 0.05 and 0.10-level. Table A1 provides the definitions and sources of all variables.

	(1)	(2)	(3)	(4)
Individual village	0.0696	0.0663*	0.0688**	0.0640*
	(0.106)	(0.0969)	(0.0392)	(0.0570)
Group village	0.0155	0.0145	0.0325	0.0322
	(0.726)	(0.708)	(0.388)	(0.356)
Highly educated			0.0253	0.0223
			(0.467)	(0.517)
Male adults in HH			0.0190	0.0203
			(0.142)	(0.117)
Female adults in HH			-0.0255**	-0.0250**
			(0.0158)	(0.0181)
Children < 16			-0.0193*	-0.0173
			(0.0628)	(0.104)
Age respondent			-0.00333**	-0.00337**
			(0.0174)	(0.0138)
Distance to province center			0.000390*	0.0004**
			(0.0647)	(0.0411)
Owns dwelling			0.0263	0.0254
			(0.145)	(0.161)
Owns fence			-0.0813***	-0.0761***
			(0.000)	(0.000)
Owns other property			-0.0339	-0.0342
			(0.189)	(0.173)
Ownes well			-0.0801**	-0.0823**
			(0.0235)	(0.0283)
Owns cattle			-0.0210	-0.0151
			(0.444)	(0.607)
Owns horses or camels			0.0634***	0.0649***
			-0.003	(0.003)
Owns other animals			-0.0184	-0.0220
			(0.399)	(0.323)
HH death			0.110**	0.111**
			(0.0401)	(0.0384)
Province fixed effects?	No	Yes	No	Yes
Observations	1,115	1,115	1,115	1,115
Pseudo R-squared	0.01	0.01	0.07	0.07

Table 5. Loan use

This table presents an overview of how borrowers used their loans. Borrowers could state more than one type of loan use. Source: Follow-up survey.

	Percentage of borrowers that used part of the loan for this purpose		**Percentage of loan amount when used for this purpose**	
	1 st group loan	***2 nd group loan***	***1 st group loan***	***2 nd group loan***
Other business expenses	0.57	0.37	0.89	0.78
Other household expenses	0.28	0.22	0.73	0.56
Mixed expenses	0.14	0.17	0.60	0.60
Education	0.06	0.06	0.74	0.54
Purchase tools/machinery	0.06	0.01	0.87	100
Purchase livestock	0.04	0.05	0.60	0.69
	***1* st individual loan**	***2* nd individual loan**	***1* st individual loan**	***2* nd individual loan**
Other business expenses	0.51	0.47	0.82	0.83
Other household expenses	0.28	0.19	0.70	0.68
Mixed expenses	0.12	0.08	0.71	0.75
Education	0.08	0.07	0.65	0.53
Purchase tools/machinery	0.06	0.03	0.73	100
Purchase livestock	0.09	0.02	0.73	0.45

Table 6. Impact on business creation and growth

This table presents the results of difference-in-differences ITT regressions to measure the impact of group (G) and individual (I) loans on business creation and growth. *Base effect* refers to the basic difference between the treatment and the control villages. *High education* refers to an interaction term between a dummy for highly educated women and the base effect. *Intensity: Months* refers to an interaction term between intensity measure *Months* and the base effect. *Intensity: Number* refers to an interaction term between intensity measure *Number* and the base effect. Regressions also include a standard set of unreported pre-treatment covariates (see Table A1). The standard errors are clustered by village and reported in brackets. ***, **, * denote significance at the 0.01, 0.05 and 0.10-level. Table A1 provides the definitions and sources of all variables.

		Probability of any type of business		Probability of female business		Profit of any businesses combined		Profit of female business	
		G	I	G	I	G	I	G	I
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
I.	Base effect	0.080	-0.028	0.105*	-0.018	-2,125	-8,169	-2,125	-24,569
		(0.055)	(0.061)	(0.063)	(0.060)	(118,787)	(89,233)	(118,787)	(40,061)
II.	Base effect	0.284***	-0.001	0.289**	-0.105	-277,351*	-110,834	-88,405	-21,485
		(0.090)	(0.123)	(0.141)	(0.137)	(161,751)	(98,292)	(80,372)	(61,399)
	High education	-0.277**	-0.031	-0.186*	0.106	316,773	122,015	80,882	-2,933
		(0.124)	(0.126)	(0.110)	(0.143)	(221,398)	(129,769)	(113,427)	(89,685)
III.	Base effect	0.079	-0.029	0.103	-0.019	-7,658	-10,137	-20,514	-25,505
		(0.055)	(0.061)	(0.063)	(0.059)	(118,932)	(89,197)	(55,142)	(40,222)
	Intensity: Months	0.007	0.021**	0.014**	0.017	41,503**	26,255***	25,894***	10,428***
		(0.007)	(0.010)	(0.006)	(0.012)	(15,874)	(9,629)	(7,740)	(3,539)
IV.	Base effect	0.008	-0.028	0.103	-0.019	-6,018	-10,028	-19,855	-25,325
		(0.056)	(0.061)	(0.063)	(0.059)	(118,719)	(89,031)	(55,095)	(40,130)
	Intensity: Number	0.005	0.102	0.058*	0.010	201,679**	136,893*	135,560***	24,564
		(0.047)	(0.103)	(0.033)	(0.126)	(81,670)	(75,678)	(38,970)	(46,477)
	Observations	2,055	2,055	2,055	2,055	2,052	2,052	2,054	2,054

Figure 1. Treatment intensity and business creation

This figure shows the probability of enterprise ownership by an average respondent in the individual lending villages (left-hand side) and group-lending villages (right-hand side) as a function of the number of months respondents in a village borrowed on average from XacBank. The top two graphs show the probability of female-owned businesses whereas the two graphs at the bottom show the probability that the average household operates any type of business (operated by the respondent, her spouse, or jointly). The blue lines indicate the expected probability while the white lines indicate a 95 per cent confidence interval.

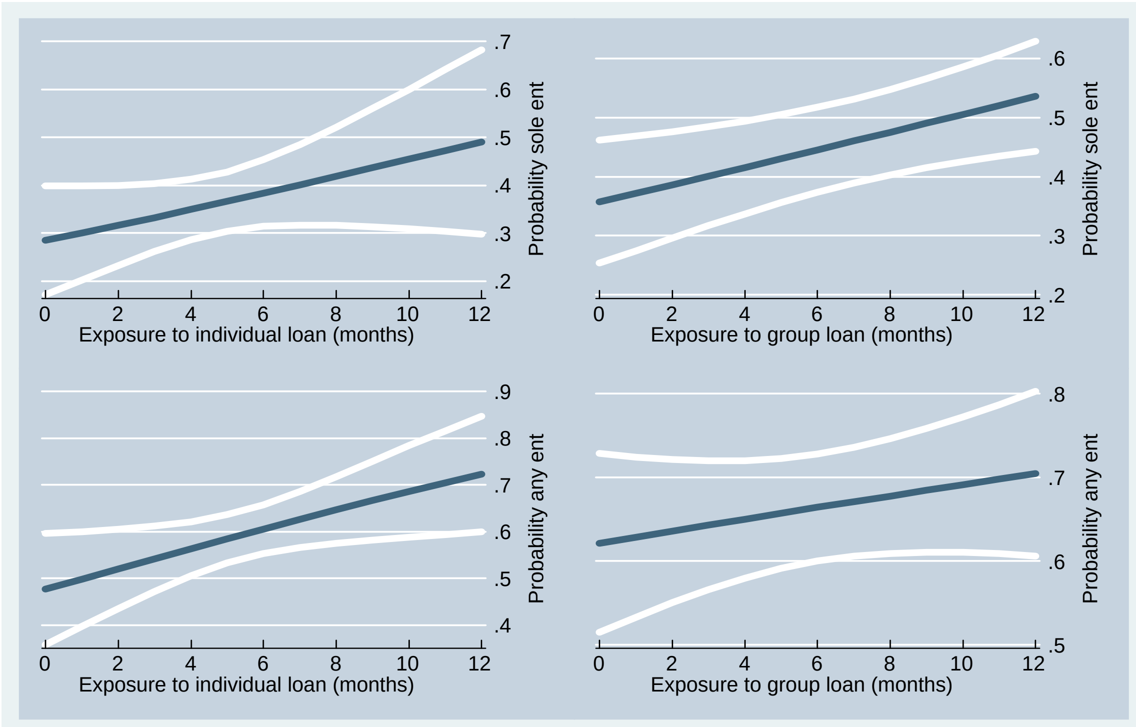


Table 7. Impact on labour supply and income

This table presents the results of difference-in-differences ITT regressions to measure the impact of group (G) and individual (I) loans on labour supply and income. *Base effect* refers to the basic difference between the treatment and the control villages. *High education* refers to an interaction term between a dummy for highly educated women and the base effect. *Intensity: Months* refers to an interaction term between intensity measure *Months* and the base effect. *Intensity: Number* refers to an interaction term between intensity measure *Number* and the base effect. Regressions also include a standard set of unreported pre-treatment covariates (see Table A1). The standard errors are clustered by village and reported in brackets. ***, **, * denote significance at the 0.01, 0.05 and 0.10-level. Table A1 provides the definitions and sources of all variables.

		Hours of wage labour by HH in average week		**Hours of enterprise labour by HH in average week**		**Total household income**	
		G	**I**	**G**	**I**	**G**	**I**
		(1)	(2)	(3)	(4)	(5)	(6)
I.	Base effect	-4.914	8.409	6.135	-8.472	-110,788	-131,659
		(9.775)	(10.03)	(12.98)	(13.99)	(204,082)	(209,531)
II.	Base effect	-45.090	0.037	21.23	-24.68	-224,480	91,786
		(28.950)	(25.24)	(37.24)	(33.18)	(224,003)	(229,403)
	High education	44.180	9.591	-16.80	18.83	146,491	-252,523
		(27.360)	(26.25)	(37.55)	(32.99)	(288,917)	(307,018)
III.	Base effect	-4.402	8.416	5.949	-8.495	-115,802	-133,925
		(9.717)	(10.04)	(12.99)	(13.94)	(203,265)	(210,005)
	Intensity: Months	-2.166*	-0.019	1.207	5.708***	45,995	24,518
		(1.217)	(3.278)	(1.626)	(1.580)	(33,618)	(33,512)
IV.	Base effect	-4.637	8.406	6.266	-8.463	-111,418	-134,153
		(9.706)	(10.01)	(13.05)	(13.96)	(203,382)	(209,871)
	Intensity: Number	-7.353	8.605	-2.213	38.18**	187,612	186,060
		(6.864)	(29.83)	(12.17)	(16.40)	(197,646)	(265,296)
	Observations	2,055	2,055	2,055	2,055	2,007	2,007

Table 8. Impact on consumption

	Total (log)		Food (log)		Non-durable (log)		Durable (log)		Dairy				Fruit and vegetables			
									Probit	*Tobit*	*Probit*	*Tobit*	*Probit*	*Tobit*	*Probit*	*Tobit*
	G	**I**	**G**	**I**	**G**	**I**	**G**	**I**	**G**	**G**	**I**	**I**	**G**	**G**	**I**	**I**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
I. Base effect	0.116	0.0347	0.173**	0.0183	0.0113	-0.00283	-0.0974	-0.0570	0.047**	22,031	0.0474***	-1,235	0.0960*	1,112*	0.0764	803.0
	(0.0805)	(0.0759)	(0.0712)	(0.0668)	(0.157)	(0.144)	(0.118)	(0.121)	(0.0189)	(18,544)	(0.0170)	(2,532)	(0.0570)	(634.8)	(0.0545)	(497.5)
II. Base effect	0.276	0.230	0.444*	0.367*	0.119	-0.137	-0.550*	-0.385	0.0603	28,877	0.0829***	10,932	0.142	1,192	0.132	1,276
	(0.238)	(0.204)	(0.220)	(0.204)	(0.393)	(0.396)	(0.326)	(0.246)	(0.0414)	(20,562)	(0.0288)	(9,215)	(0.101)	(1,156)	(0.0952)	(875.3)
High education	-0.185	-0.227	-0.317	-0.407*	-0.116	0.156	0.530	0.389	-0.0336	-7,922	-0.101	-14,020	-0.0838	-84.42	-0.0873	-541.0
	(0.272)	(0.246)	(0.239)	(0.229)	(0.425)	(0.418)	(0.332)	(0.235)	(0.0973)	(13,378)	(0.109)	(10,913)	(0.160)	(1,084)	(0.149)	(908.2)
III. Base effect	0.110	0.0339	0.166**	0.0163	0.00297	-0.00253	-0.102	-0.0571	0.0462**	21,295	0.0473***	-1,361	0.0975*	1,100*	0.0779	801.8
	(0.0800)	(0.0759)	(0.0703)	(0.0667)	(0.158)	(0.144)	(0.119)	(0.121)	(0.0184)	(18,263)	(0.0158)	(2,508)	(0.0565)	(632.6)	(0.0542)	(497.4)
Intensity: Months	0.049***	-0.00146	0.055***	0.0193	0.037**	-0.0184	0.035	-0.0114	0.0145***	7,110	-0.0160	-74.49	-0.0108	62.43	0.0227**	108.6
	(0.0128)	(0.0180)	(0.0160)	(0.0173)	(0.0174)	(0.0255)	(0.0225)	(0.0335)	(0.00475)	(4,535)	(0.0146)	(1,518)	(0.00881)	(53.30)	(0.0113)	(105.8)
IV. Base effect	0.111	0.0335	0.166**	0.0163	0.0075	-0.00287	-0.0992	-0.0569	0.0472**	21,137	0.0471***	-1,528	0.0966*	1,102*	0.0784	801.5
	(0.0802)	(0.0762)	(0.0707)	(0.0671)	(0.158)	(0.144)	(0.119)	(0.121)	(0.0183)	(18,353)	(0.0155)	(2,562)	(0.0568)	(633.5)	(0.0541)	(496.5)
Intensity: Number	0.272***	0.00143	0.359***	0.0581	0.123	-0.0816	0.0910	-0.0649	0.0790***	56,965*	-0.147	1,420	-0.0362	330.8	0.176*	1,061
	(0.0689)	(0.160)	(0.0907)	(0.194)	(0.102)	(0.186)	(0.141)	(0.233)	(0.0206)	(31,544)	(0.115)	(15,570)	(0.0419)	(311.9)	(0.0944)	(726.6)
Observations	2,055	2,055	2,050	2,050	1,993	1,993	2,048	2,048	2,034	2,034	2,034	2,034	2,034	2,034	2,034	2,034

	Non-alcoholic drinks				Combustibles				Cigarettes				Felt for ger	
	Probit	*Tobit*	*Probit*	*Tobit*	*Probit*	*Tobit*	*Probit*	*Tobit*	*Probit*	*Tobit*	*Probit*	*Tobit*	*Probit*	*Probit*
	G	**G**	**I**	**I**	**G**	**G**	**I**	**I**	**G**	**G**	**I**	**I**	**G**	**I**
	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(31)
I. Base effect	0.125**	1,426**	0.0700	786.6	0.0221*	-264.1	0.00442	6,015	-0.0681*	-2,644**	-0.0630	-943.5	-0.00483	-0.00452
	(0.0583)	(557.3)	(0.0604)	(555.0)	(0.0115)	(6,867)	(0.0224)	(7,474)	(0.0348)	(1,043)	(0.0440)	(957.0)	(0.0100)	(0.00903)
II. Base effect	-0.0196	-272.2	0.0844	995.2	-0.554**	11,140	0.000597	3,182	-0.0635	-3,685*	-0.0474	-658.3	0.972***	0.966***
	(0.178)	(1,885)	(0.173)	(1,562)	(0.256)	(26,035)	(0.00362)	(26,611)	(0.0963)	(1,927)	(0.110)	(2,496)	(0.0472)	(0.0271)
High education	0.163	1,867	-0.0146	-246.9	0.0222	-13,059	-0.00582	3,420	-0.00969	1,164	-0.0204	-270.9	-0.0327***	-0.0331***
	(0.230)	(2,094)	(0.188)	(1,849)	(0.0912)	(26,517)	(0.0281)	(26,692)	(0.105)	(2,150)	(0.106)	(2,580)	(0.0101)	(0.00656)
III. Base effect	0.122**	1,393**	0.0704	788.6	0.0159*	-848.3	0.00346	5,961	-0.0678*	-2,629**	-0.0621	-902.2	-0.00364	-0.00400
	(0.0580)	(560.1)	(0.0604)	(554.7)	(0.00913)	(6,839)	(0.0179)	(7,450)	(0.0350)	(1,059)	(0.0436)	(960.0)	(0.00934)	(0.00750)
Intensity: Months	0.00839	129.2*	-0.0114	-79.40	0.00728**	2,735***	-0.00120	-944.3	-0.00528	-270.0	-0.0125	-337.7	0.00337***	0.00155
	(0.00752)	(70.16)	(0.00890)	(117.5)	(0.00321)	(1,003)	(0.00170)	(1,074)	(0.00688)	(290.2)	(0.0197)	(492.2)	(0.000812)	(0.00141)
IV. Base effect	0.123**	1,397**	0.0708	787.9	0.0191*	-574.0	0.00385	5,966	-0.0677*	-2,636**	-0.0625	-917.6	-0.00361	-0.00412
	(0.0581)	(560.0)	(0.0604)	(555.0)	(0.0105)	(6,839)	(0.0200)	(7,456)	(0.0348)	(1,051)	(0.0436)	(964.8)	(0.00966)	(0.00783)
Intensity: Number	0.0363	588.8	-0.0991	-643.2	0.0282**	10,244**	-0.00990	-3,635	-0.0265	-1,163	-0.0412	-1,523	0.0166***	-0.00238
	(0.0399)	(389.9)	(0.0628)	(882.8)	(0.0143)	(5,029)	(0.0162)	(8,240)	(0.0339)	(1,425)	(0.174)	(4,238)	(0.00437)	(0.0128)
Observations	2,034	2,034	2,034	2,034	2,034	2,034	2,034	2,034	2,034	2,034	2,034	2,034	2,034	2,034

Table 9. Impact on asset ownership

This table presents the results of difference-in-differences ITT regressions to measure the impact of group (G) and individual (I) loans on asset ownership. *Base effect* refers to the basic difference between the treatment and the control villages. *High education* refers to an interaction term between a dummy for highly educated women and the base effect. *Intensity: Months* refers to an interaction term between intensity measure *Months* and the base effect. *Intensity: Number* refers to an interaction term between intensity measure *Number* and the base effect. Bod are standardized Mongolian livestock units. One horse, yak, or cattle equals one bod; one camel equals 1.4 bod; one sheep equals 1/6 bod; and one goat equals 1/7 bod. Regressions also include a standard set of unreported pre-treatment covariates (see Table A1). The standard errors are clustered by village and reported in brackets. ***, **, * denote significance at the 0.01, 0.05 and 0.10-level. Table A1 provides the definitions and sources of all variables.

	Value of all assets (incl. main dwelling)		Probability 2nd house		Probability 2nd ger		Probability land/well		Probability vehicle		Probability VCR or radio		Probability television	
	G	I	G	I	G	I	G	I	G	I	G	I	G	I
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
I. Base effect	-129,482	-325,163	0.001	0.071	0.009	0.064	-0.017	-0.105	0.062	0.018	0.172***	0.137**	-0.022	-0.001
	(527,000)	(542,918)	(0.072)	(0.071)	(0.05)	(0.042)	(0.125)	(0.113)	(0.050)	(0.046)	(0.054)	(0.054)	(0.022)	(0.014)
II. Base effect	-1,148,000	-905,922	-0.057	0.148	-2.08***	0.072	-0.335***	-0.124	-0.297***	-0.237***	0.169	0.192*	-0.005	-0.010
	(1,188,000)	(831,094)	(0.134)	(0.115)	(0.0611)	(0.122)	(0.113)	(0.151)	(0.065)	(0.083)	(0.143)	(0.107)	(0.036)	(0.037)
High education	922,123	357,832	0.069	-0.080	0.406**	-0.006	0.307**	0.023	0.516***	0.360**	0.004	-0.062	-0.012	0.012
	(1,367,000)	(1,019,000)	(0.142)	(0.105)	(0.178)	(0.106)	(0.131)	(0.157)	(0.118)	(0.146)	(0.161)	(0.140)	(0.054)	(0.041)
III. Base effect	-164,484	-331,615	0.005	0.072	0.008	0.062	-0.120	-0.110	0.0613	0.017	0.171***	0.136**	-0.020	-0.001
	(520,573)	(539,958)	(0.074)	(0.072)	(0.035)	(0.042)	(0.124)	(0.113)	(0.05)	(0.046)	(0.054)	(0.055)	(0.021)	(0.014)
Intensity: Months	264,751**	31,276	-0.03***	-0.03***	0.02***	0.022**	0.02***	0.045***	0.008	0.006	0.004	0.024*	-0.003	0.011***
	(103,886)	(202,940)	(0.007)	(0.010)	(0.005)	(0.010)	(0.007)	(0.0170)	(0.012)	(0.023)	(0.015)	(0.014)	(0.003)	(0.003)
IV. Base effect	-147,759	-335,491	0.004	0.072	0.07	0.063	-0.118	-0.111	0.062	0.018	0.172***	0.135***	-0.021	-0.001
	(522,313)	(540,709)	(0.073)	(0.072)	(0.035)	(0.042)	(0.124)	(0.113)	(0.05)	(0.046)	(0.054)	(0.055)	(0.021)	(0.014)
Intensity: Number	987,927*	880,953	-0.15***	-0.185**	0.081**	0.047	0.087**	0.399***	-0.03	0.043	0.010	0.173*	-0.070	0.098***
	(574,456)	(1,440,000)	(0.036)	(0.088)	(0.032)	(0.101)	(0.042)	(0.112)	(0.06)	(0.171)	(0.088)	(0.094)	(0.014)	(0.026)
Observations	2,055	2,055	2,055	2,055	2,055	2,055	2,055	2,055	2,055	2,055	2,055	2,055	2,055	2,055

	Probability large household appliances		Probability tools		Probability unsold stock and raw materials		Probability riding equipment		Number of cattle		Number of animals (in bod)	
	G	I	G	I	G	I	G	I	G	I	G	I
	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)
I. Base effect	0.085**	0.070*	0.060	0.161	0.011	-0.090	0.039	0.017	-0.601	-1.884*	-1.330	0.956
	(0.036)	(0.041)	(0.113)	(0.109)	(0.043)	(0.046)	(0.044)	(0.044)	(1.255)	(1.083)	(1.126)	(1.362)
II. Base effect	-0.048	-0.180	0.306**	0.366***	0.037	-0.037	-0.106	-0.131*	-3.356	-3.827	0.420	1.234
	(0.139)	(0.126)	(0.132)	(0.135)	(0.178)	(0.153)	(0.116)	(0.074)	(2.467)	(2.509)	(2.085)	(2.170)
High education	0.147	0.258**	-0.313*	-0.290*	-0.027	0.021	0.166	0.174**	3.135	2.237	-2.542	-0.410
	(0.131)	(0.105)	(0.160)	(0.158)	(0.159)	(0.170)	(0.127)	(0.083)	(2.621)	(2.644)	(2.400)	(2.264)
III. Base effect	0.084**	0.070*	0.059	0.161	0.010	-0.020	0.034	0.016	-0.822	-1.876*	-1.330	0.956
	(0.037)	(0.041)	(0.112)	(0.109)	(0.043)	(0.046)	(0.044)	(0.043)	(1.264)	(1.067)	(1.127)	(1.362)
Intensity: Months	0.013	0.020	-0.01	0.027**	0.012**	0.014	0.036***	0.014	1.268***	0.127	0.139	0.651
	(0.014)	(0.019)	(0.01)	(0.012)	(0.005)	(0.012)	(0.005)	(0.010)	(0.262)	(0.067)	(0.685)	(0.711)
IV. Base effect	0.084**	0.069*	0.058	0.161	0.010	-0.019	0.036	0.015	-0.777	-1.871*	-1.330	0.956
	(0.036)	(0.041)	(0.113)	(0.109)	(0.04)	(0.046)	(0044)	(0.043)	(1.256)	(1.064)	(1.127)	(1.362)
Intensity: Number	0.027	0.210	-0.078	0.207*	0.064**	0.120	0.143***	0.151*	6.047***	0.233	4.952	2.393
	(0.073)	(0.146)	(0.050)	(0.111)	(0.029)	(0.103)	(0.041)	(0.089)	(1.746)	(4.787)	(6.422)	(3.529)
Observations	2,055	2,055	2,053	2,053	2,055	2,055	2,055	2,055	2,051	2,051	1,874	1,874

Table 10. Impact on informal transfers

This table presents the results of difference-in-differences ITT regressions to measure the impact of group (G) and individual (I) loans on informal transfers to and from family and friends. *Base effect* refers to the basic difference between the treatment and the control villages. *High education* refers to an interaction term between a dummy for highly educated women and the base effect. *Intensity: Months* refers to an interaction term between intensity measure *Months* and the base effect. *Intensity: Number* refers to an interaction term between intensity measure *Number* and the base effect. Regressions also include a standard set of unreported pre-treatment covariates (see Table A1). Standard errors are clustered by village and reported in brackets. ***, **, * denote significance at the 0.01, 0.05 and 0.10-level. Table A1 provides the definitions and sources of all variables.

		Probability of *receiving* transfers from *friends* during the last *year*		**Probability of *making* transfers to *friends* during the last *year***		**Probability of *receiving* transfers from *family* during the last *year***		**Probability of *making* transfers to *family* during the last *year***		**Probability of *receiving* transfers from *family* during the last *month***		**Probability of *making* transfers to *family* during the last *month***	
		G	**I**	**G**	**I**	**G**	**I**	**G**	**I**	**G**	**I**	**G**	**I**
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
I.	Base effect	0.0454	-0.00322	-0.0216	0.0210	-0.0201	0.0389	-0.00984	0.115	-0.0364	-0.000815	-0.0367	0.0244
		(0.0499)	(0.0369)	(0.0526)	(0.0566)	(0.0644)	(0.0594)	(0.0599)	(0.0704)	(0.0447)	(0.0451)	(0.0444)	(0.0624)
II.	Base effect	0.0537	-0.0369	-0.0683	-0.0698	0.144	-0.0973	0.190	0.184	0.138	-0.0746	0.0269	0.0487
		(0.0944)	(0.0553)	(0.0784)	(0.0549)	(0.130)	(0.0930)	(0.130)	(0.170)	(0.154)	(0.0664)	(0.218)	(0.195)
	High education	-0.00794	0.0462	0.0710	0.134	-0.155	0.174*	-0.213	-0.0832	-0.125**	0.109	-0.0593	-0.0266
		(0.0715)	(0.0895)	(0.0996)	(0.106)	(0.101)	(0.0943)	(0.132)	(0.160)	(0.0605)	(0.111)	(0.198)	(0.166)
III.	Base effect	0.0491	-0.00222	-0.0194	0.0213	-0.0133	0.0389	-0.00680	0.115	-0.0329	-0.000895	-0.0348	0.0241
		-0.0509	-0.0366	-0.0527	-0.0561	(0.0647)	(0.0592)	(0.0594)	(0.0706)	(0.0447)	(0.0445)	(0.0443)	(0.0623)
	Intensity: Months	-0.0102***	0.00706	-0.0155***	0.0146**	-0.0256***	0.00866	-0.0156	0.0140	-0.0126***	0.0141***	-0.00854*	0.0264***
		-0.00253	-0.00433	-0.0059	-0.00736	(0.00683)	(0.00762)	(0.0119)	(0.0112)	(0.00399)	(0.00486)	(0.00499)	(0.00937)
IV.	Base effect	0.0491	-0.00186	-0.0194	0.0217	-0.0137	0.0390	-0.00788	0.115	-0.0336	-0.00103	-0.0358	0.0239
		(0.0511)	(0.0367)	(0.0526)	(0.0558)	(0.0646)	(0.0591)	(0.0597)	(0.0706)	(0.0444)	(0.0447)	(0.0443)	(0.0624)
	Intensity: Number	-0.0585***	0.0973**	-0.101***	0.166***	-0.136***	0.0805	-0.0718	0.126*	-0.0582**	0.0828*	-0.0271	0.179***
		(0.0128)	(0.0413)	(0.0337)	(0.0606)	(0.0355)	(0.0646)	(0.0642)	(0.0762)	(0.0230)	(0.0494)	(0.0293)	(0.0657)
	Observations	2,054	2,054	2,055	2,055	2,054	2,054	2,055	2,055	2,054	2,054	2,055	2,055

Figure 2a. Treatment intensity and informal transfers in group-lending villages

This figure shows the probability of receiving or giving informal transfers for an average respondent in the group-lending villages as a function of the number of months respondents borrowed on average from XacBank. The top two graphs show the probability of giving (left) and receiving (right) transfers to and from friends, while the bottom two graphs show the same for transfers to and from family members. The blue lines indicate the expected probability while the white lines indicate a 95 per cent confidence interval.

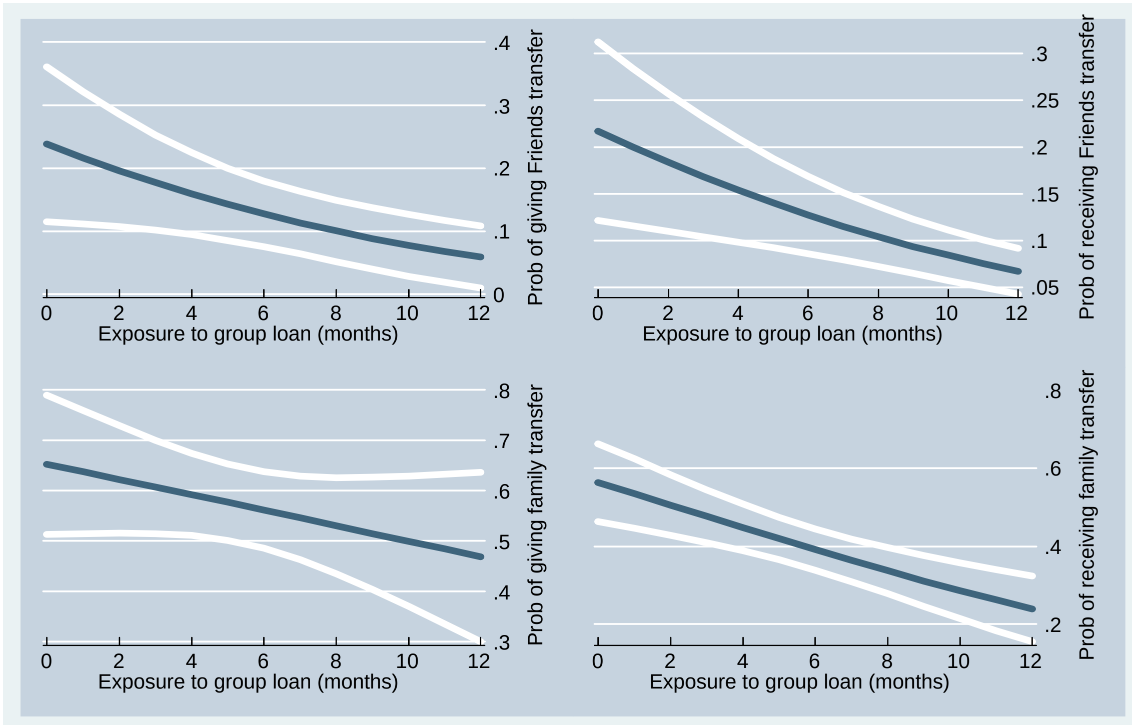


Figure 2b. Treatment intensity and informal transfers in individual-lending villages

This figure shows the probability of receiving or giving informal transfers for an average respondent in the individual-lending villages as a function of the number of months respondents borrowed on average from XacBank. The top two graphs show the probability of giving (left) and receiving (right) transfers to and from friends, while the bottom two graphs show the same for transfers to and from family members. The blue lines indicate the expected probability while the white lines indicate a 95 per cent confidence interval.

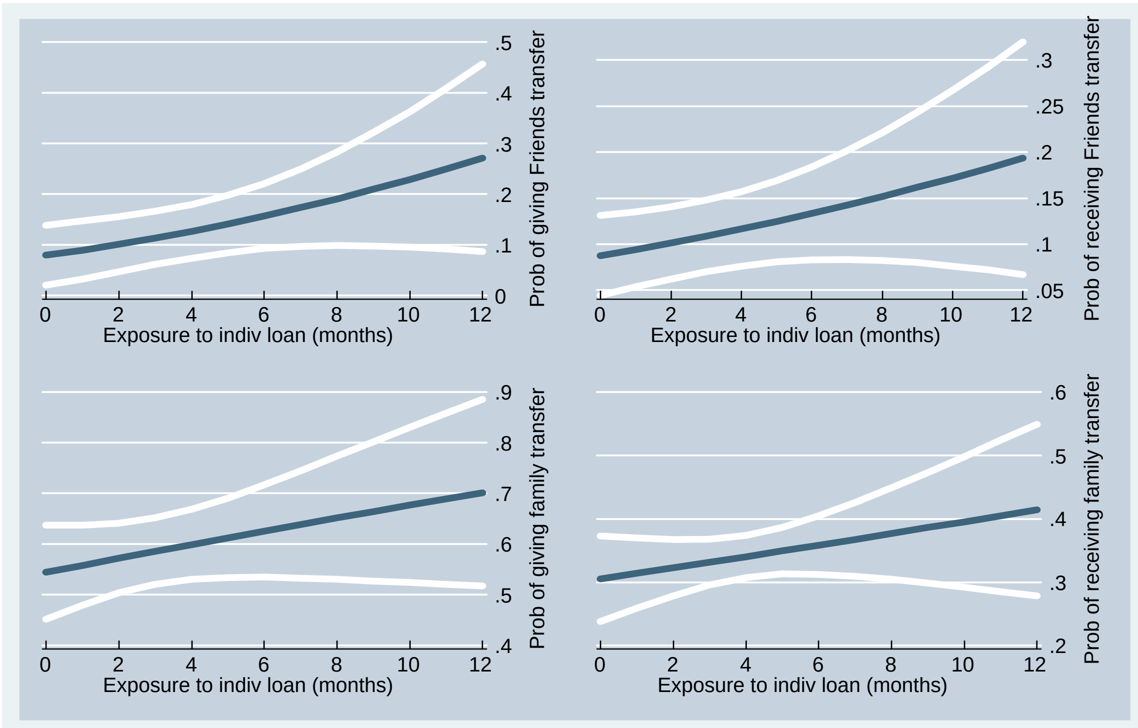


Table 11. Determinants of loan default

This table presents probit regressions to explain loan default. The dependent variable is a dummy that indicates whether a borrower defaulted (1) or not (0). *Loan amount* and *Debt at baseline* are measured in millions of tögrög. The following additional covariates were included but are now shown (all insignificant): *Household size*, *Collateral value*, *Male HH members >16*, *Female HH members >16*, *Children <16*, *Owns fence*, *House or flat*, *Owns vehicle*, *Saver*, *HH crop disaster*, *HH death*. Standard errors are clustered by village and reported in brackets. ***, **, * denote significance at the 0.01, 0.05 and 0.10-level. Table A1 provides the definitions and sources of all variables. Source of repayment data: XacBank.

	First loan		**All loans**	
	(1)	(2)	(3)	(4)
Group loan	0.029	-0.144	0.289	0.387
	(0.398)	(0.144)	(0.339)	(0.360)
Loan amount		-0.790		0.444
		(0.636)		(0.584)
Debt at baseline		-0.200*		-0.200*
		(0.140)		(0.117)
No. prior loans with XacBank				-0.161***
				(0.040)
Months since disbursement		0.096***		0.109***
		(0.024)		(0.021)
Owns land		-0.590***		-0.263
		(0.222)		(0.208)
Owns TV		1.262**		0.152
		(0.643)		(0.318)
Owns enterprise		-0.403*		-0.093
		(0.221)		(0.153)
Grade VIII education		-0.868***		-0.370*
		(0.297)		(0.218)
Vocational education		-0.809***		-0.359
		(0.325)		(0.225)
Age		-0.088		-0.023
		(0.090)		(0.066)
Age squared		0.001		0.000
		(0.001)		(0.001)
Buddhist		0.465		0.178
		(0.390)		(0.262)
Hahl		-0.763**		-0.707**
		(0.377)		(0.329)
Married		0.192		0.034
		(0.266)		(0.188)
Natural disaster		0.752*		0.300
		(0.404)		(0.277)
Observations	327	302	638	612
Pseudo R-squared	0.009	0.321	0.009	0.29

Figure A1. Overview of participating villages and provinces

This figure shows the geographical location of the 10 control soum centers (villages) as black dots, the 15 individual-lending villages (grey dots), and the 15 group-lending villages (white dots) across the five Mongolian provinces that participated in the experiment.

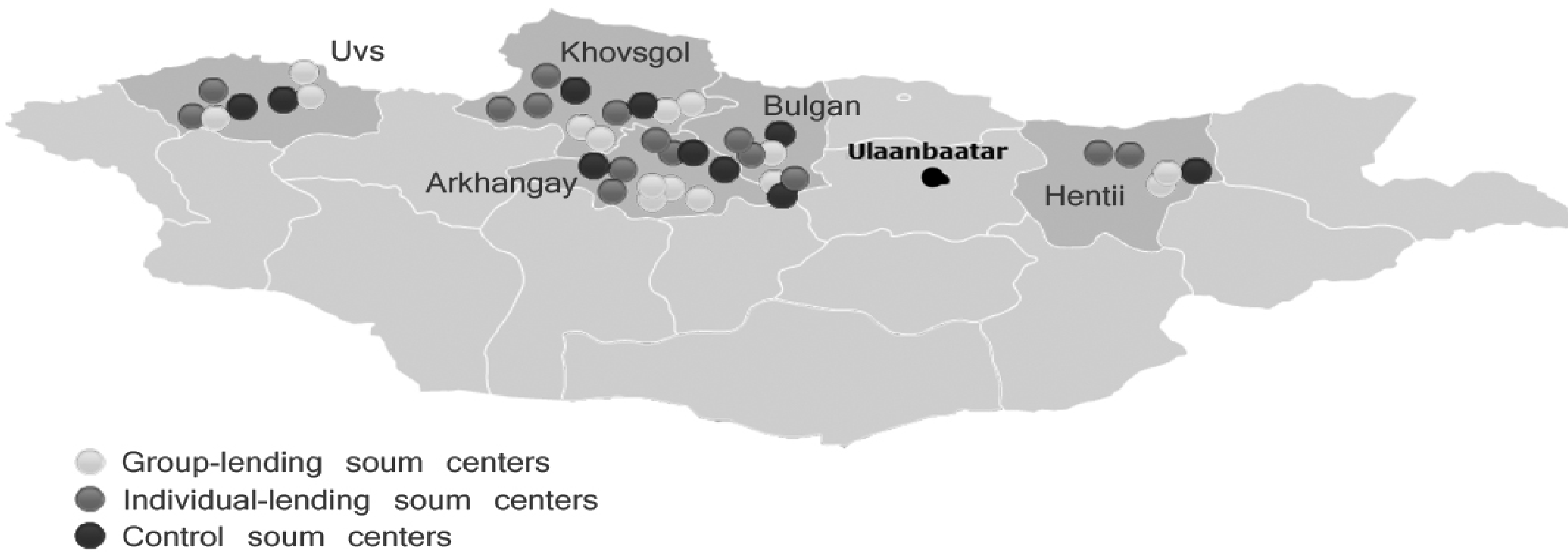


Table A1. Variable definitions

This table provides the names, definitions, and data sources of the variables used in the empirical analysis in alphabetical order. MNT = Mongolian t ögrög.

Variable name	Description	Standard control variable in impact analysis?
	Respondent and household (HH) level data (# respondents = 1,148). Source: Baseline survey	
Age	Age in years of respondent	X
Age squared	Age in years of respondent squared	X
Amount female business	Loan amount (in 000's MNT) that is used for a female-owned business	
At least one loan	Dummy variable that is '1' if the HH had at least one loan outstanding	
Buddhist	Respondent is of the Buddhist religion	X
Children <16	Number of children in the HH younger than 16 years	X
Collateral value	Estimated market value of the collateral (in 000's MNT)	
Consumption fuel	Quantity of fuel burned by the HH in the past week (in liters)	
Consumption milk	Quantity of milk consumed by the HH in the past week (in liters)	
Consumption red meat	Quantity of red meat consumed by the HH in the past week (in kilograms)	
Consumption vegetables	Quantity of vegetables consumed by the HH in the past week (in kilograms)	
Debt at baseline	Amount of loans outstanding at time of baseline survey (in million MNT)	
Debt service	Loan+interest (re)payment at HH level over past month (in 000's MNT) conditional on at least one loan outstanding	
Debt value	Amount of debt (in million MNT) at HH level that is still outstanding conditional on at least one loan outstanding	
Debt/HH income	Outstanding debt amount as proportion of annual HH income conditional on at least one loan outstanding	
Education respondent	Number of years of education of the respondent	
Education high	Dummy variable that is '1' if the respondent completed grade VIII or higher or vocational	
Education >VIII	Dummy variable that is '1' if the respondent completed grade VIII or higher	X
Education vocational	Dummy variable that is '1' if the respondent completed vocational training	X
Female business	Dummy variable that is '1' if the respondent operates her own business conditional on at least one HH business	
Female HH members >16	Number of female HH members aged 16 or older	X
Given transfers	Value of monetary and in-kind transfers given in last 12 months from non-relatives (in 000's MNT) conditional on giving	
Hahl	Respondent ethnicity is Hahl	X
HH crop disaster	Dummy variable that is '1' if the HH experienced severe crop losses during the previous year	
HH death	Dummy variable that is '1' if the HH experienced death of a HH member in the previous year	
HH robbery	Dummy variable that is '1' if the HH experienced a robbery in the previous year	
Highly educated	Dummy variable that is '1' if the respondent has completed vocational training or grade VIII or above	
Hours hired	Average number of hours worked per week in peak season by non-HH members in the respondent's enterprise	
Household size	Number of children and adults in the household	
House or flat	HH lives in a house, flat or apartment rather than a *ger*	
Interest rate	Monthly interest rate on a loan	
Joint enterprise	Dummy variable that is '1' if the respondent operates an enterprise together with her spouse	
Male HH members >16	Number of male HH members aged 16 or older	X
Married	Dummy variable that is '1' if the respondent is married or living together with partner	X

Loans at baseline	Dummy variable that is '1' if the HH had at least one loan outstanding at the time of the baseline interview	X
Operates business	Dummy variable that is '1' if the HH operates at least one business	
Outstanding loans	Number of loans taken by the HH that are still outstanding, conditional on at least one loan outstanding	
Owns animals	Dummy variable that is '1' if the HH owns animals for business purposes	
Owns dwelling	Dummy variable that is '1' if the HH owns at least one dwelling (ger, house, and/or apartment)	
Owns fence	Dummy variable that is '1' if the HH owns a fence around the dwelling	
Owns HH appliances	Dummy variable that is '1' if the HH owns large household appliances (refrigerator, cooler, washing machine)	
Owns tools/machinery	Dummy variable that is '1' if the HH owns tools and/or machinery for business use	
Owns vehicle	Dummy variable that is '1' if the HH owns a vehicle (car, lorry, tractor and/or motorbike)	
Owns well	Dummy variable that is '1' if the HH owns a well near the dwelling	
Partner enterprise	Dummy variable that is '1' if the respondent's spouse operates an enterprise but not jointly with the respondent	
Percentage female business	Percentage of total outstanding loan amount of the HH that is used for a female-owned business	
Percentage private use	Percentage of total outstanding loan amount of the HH that is used for private purposes	
Prior loans	Number of loans taken by the HH over the last five years that had been fully repaid at the time of the baseline survey	
Received transfers	Value of monetary and in-kind transfers received in last 12 months from non-relatives (in 000's MNT) conditional on receipt	
Saver	Respondent indicated that she saves	
Secured loans	Percentage of loans that is collateralized	
Self-employed	Dummy variable that is '1' if the respondent is self-employed	
Sewing or shop	Dummy variable that is '1' of the respondent operates a sewing business or shop conditional on having a business	
Sole enterprise	Dummy variable that is '1' if the respondent operates an enterprise independent from her spouse	
Value of dwelling	Value of the dwelling the HH lives in (in million MNT)	
Wage earnings	Average weekly wage earnings for wage earners (in 000's MNT)	
Years in existence	Number of years since the establishment of the respondent's business	
	Village-level data (# villages = 40). Source: Village survey in Spring 2008	
Banks in district	Number of bank branches in the district	
Distance to paved road	Distance (in km) from the village to the nearest paved road	
Distance to province center	Distance (in km) from the village to the province center	X
District area	Total surface are of the district in km2	
Doctors in district	Number of doctors in the district	X
Livestock in district	Number of livestock (cattle, camels, horses, sheep, goats) in the district	
Months	Average number of months between the date when respondents in a village received the first loan and the follow-up survey	
Number	Average number of loans received by the respondents in a village	
People in district	Number of people living in the district surrounding a village as well as that village itself	
People in village	Number of people living in a village	
Price bread	Price of a loaf of bread (in MNT)	
Price milk	Price of a liter of milk (in MNT)	
Price mutton	Price of a kilo of mutton meat (in MNT)	
Primary schools district	Number of primary schools in district	X
SCCs in district	Number of Savings and Credit Cooperatives in the district	
SS teachers	Number of secondary school teachers in the district	X
Time to paved road	Time (in minutes) to travel from the village to the nearest paved road by car or motorcycle	
Time to province center	Time (in minutes) to travel from the village to the province center by car or motorcycle	

Understanding the Incentives of Commissions Motivated Agents: Theory and Evidence from the Indian Life Insurance Market

Santosh Anagol
Wharton

Shawn Cole
Harvard Business School

Shayak Sarkar*
Harvard University

January 2, 2012

Abstract

We conduct a series of field experiments to evaluate two competing views of the role of financial service intermediaries in providing product recommendations to potentially uninformed consumers. One view argues intermediaries provide valuable product education, and guide consumers towards suitable products. Consumers understand how commissions affect agents' incentives, and make optimal product choices. The second view argues that intermediaries recommend and sell products that maximize the agents' well-being, with little or no regard for the customer. Audit studies in the Indian life insurance market find evidence supporting the second view: in 60-80% of visits, agents recommend unsuitable (strictly dominated) products that provide high commissions to the agents. Customers who specifically express interest in a suitable product are more likely to receive an appropriate recommendation, though most still receive bad advice. Agents cater to the beliefs of uninformed consumers, even when those beliefs are wrong.

We then test how regulation and market structure affect advice. A natural experiment that required agents to describe commissions for a specific product caused agents to shift recommendations to an alternative product, which had even higher commissions but no disclosure requirement. We do find some scope for market discipline to generate debiasing: when auditors express inconsistent beliefs about the product suitable from them, and mention they have received advice from another seller of insurance, they are more likely to receive suitable advice. Agents provide better advice to more sophisticated consumers.

Finally, we describe a model in which dominated products survive in equilibrium, even with competition.

*anagol@wharton.upenn.edu, scole@hbs.edu, and ssarkar@fas.harvard.edu.

iTrust provided valuable context on the Indian insurance market for this project. We also thank Daniel Bergstresser, Sendhil Mullainathan, Petia Topalova, Peter Tufano, Shing-Yi Wang, Justin Wolfers, and workshop participants at Harvard Business School, Helsinki Finance Summit, Hunter College, the ISB CAF Conference, the NBER Household Finance Working Group, the NBER Insurance Working Group, Princeton, the RAND Behavioral Finance Forum, and the Utah Winter Finance Conference for comments and suggestions. We thank the Harvard Lab for Economic and Policy Applications, Wharton Global Initiatives, Wharton Dean's Research Fund, Wharton Risk Management and Decision Processes Center, and the Penn Lauder CIBER Fund for financial support. Manoj Garg, Shahid Vaziralli and Anand Kothari provided excellent research assistance.

1 Introduction

The recent financial crisis has spurred many countries to pursue new consumer financial regulations that could drastically change the way household financial products are distributed. Both Australia and the U.K. Financial Services Authority have announced bans, to take effect in 2012, on the payment of commissions to independent financial advisors.[1] And as of August 2009, the Indian mutual funds regulator banned mutual funds from collecting entry loads, which had previously primarily been used to pay commissions to mutual fund brokers.[2] Opponents of these bans argue that commissions are important to motivate agents to provide financial advice and customer education, that competition and reputation concerns will discipline agents, and that consumers have demonstrated little willingness to pay for independent financial advice.

There is very little evidence to inform these important policy questions. In this paper, we use a set of field experiments conducted in the Indian life insurance market to provide quantitative evidence on the quality of advice provided by commissions motivated agents. In addition, we test recent theories on how commissions motivated agents will respond to disclosure requirements, greater competition, or more sophisticated consumers.

We focus on the market for life insurance in India for the following reasons. First, given the complexity of life insurance, consumers likely require help in making purchasing decisions. Second, popular press accounts suggest the market may not function well: life insurance agents in India engage in unethical business practices, promising unrealistic returns or suggesting only high commission products.[3] Third, the industry is large, with approximately 44 billion dollars of premiums collected in the 2007-2008 financial year, 2.7 million insurance sales agents who collected approximately 3.73 billion dollars in commissions in 2007-2008, and a total of 105 million insurance customers. Approximately 20 percent of household savings in India is invested in whole life insurance plans (IRDA, 2010). Fourth, agent behavior is extremely important in this market, as approximately 90 percent of insurance purchasers buy through agents.

[1] Independent Financial Advisors received commissions to sell mutual funds and life insurance products. See Reuters (2009), Vincent (2009) and Dunkley (2009) for more information on the U.K. ban on commissions. See "Australia Proposes Ban on Commission" in the Financial Times, September 4, 2011.

[2] For newspaper accounts of the importance of entry loads as the primary source of commissions see (1) "MFs Look For Life Beyond Entry Load Ban," Times of India, July 19, 2010 (2) "Mutual Fund Industry Struggling to Woo Retail Investors," Business Today, February 2011 Edition.

[3] See for example, "LIC agents promise 200% return on '0-investment' plan," Economic Times, 22 February 2008.

Lastly, commissions motivated sales agents are of particular importance in emerging economies where a large fraction of the population has little or no experience with formal financial markets. Commissions may motivate agents to identify potential consumers, educate them about the range of available products, and identify the most suitable products. Opponents, however, argue that the commissions motivated agents will encourage consumers to purchase expensive, complicated products that are not necessarily welfare maximizing for households. Systematic empirical evidence is needed to inform the policy debate about whether commissions motivated agents are necessary for encouraging the adoption of complicated household financial products.

This project consists of three closely related field experiments. All of these experiments use an audit study methodology, in which we hired and trained individuals to visit life insurance agents, express interest in life insurance policies, and seek product recommendations. The goal of the first set of audits was to test whether, and under what circumstances, agents recommend products suitable for consumers. In particular, we focused on two common life insurance products: whole life and term life. We chose these two products because, in the Indian context, consumers are generally much better off purchasing a term life insurance product than whole life. In section II, we detail how large this violation of the law of one price can be. The combination of a savings account and a term insurance policy can provide over six times as much value as a whole life insurance policy.

An important source of friction in financial product markets is that consumers may not know which products are best for them. A range of evidence suggests that individuals with low levels of financial literacy make poor investment decisions (Lusardi and Mitchell, 2007). An important role of agents may be to identify suitable products. In our first experiment, we randomly vary both the stated belief of the customer as to which product is most suitable, as well information the client provides about his or her actual needs. Thus, we have some treatments where the customer has an initial preference for term insurance but where whole insurance is actually the more suitable product, and vice versa (whole insurance could be a suitable product for an individual who has difficulty committing to saving). If an agent's role is to match clients to suitable products, only the latter information should affect agent recommendations. In fact, we find agents are just as responsive to consumers self-reported (and incorrect) beliefs as they are to consumers needs.

Interestingly, this is true even when the commission on the more suitable product is higher, and hence the agent has a strong incentive to de-bias the customer. We view this result as important because it suggests that agents have a strong incentive to cater to the initial preferences of customers in order to close the sale; contradicting the initial preference of customers, even when they are wrong, may not be a good sales strategy. Thus, salesmen are unlikely to de-bias customers if they have strong initial preferences to products that may be unsuitable for them.

Our second, third, and fourth experiments test predictions on how disclosure, competition, and increased sophistication of consumers affect the quality of advice provided by agents.

In our second experiment, we study whether competition amongst agents can lead to higher quality advice. We find that that agents who face greater competition, which we induce by having our auditor state that they have already talked to another agent, leads to better advice. This evidence is consistent with standard economic models which suggest that, at least under perfect competition, agents will have an incentive to provide good advice.

In our third experiment we test how disclosure regulation affects the quality of advice provided by life insurance agents. Mandating that agents disclose commissions has been a popular policy response to perceived mis-selling. In theory, once consumers understand the incentives faced by agents, they will be able to filter the advice and recommendations, improving the chance they choose the product best suited for them, rather than the product that maximizes the agents commissions. We take advantage of a natural experiment: as of July 1, 2010, the Indian insurance regulator mandated that insurance agents disclose the commissions they earned on equity linked life insurance products. We have data on 149 audits conducted before July 1, and 108 audits conducted after July 1. We find that following the implementation of the regulation, life insurance agents are much less likely to propose the unit-linked insurance policy to clients, and instead recommend whole life policies which have higher, but opaque, commissions.

In our last experiment, we test whether the quality of advice received varies by the level of sophistication the clients demonstrate. We find that less sophisticated agents are more likely to receive a recommendation for the wrong product, suggesting that agents discriminate in the types of advice they provide. This result suggests that the selling of unsuitable products is likely to have the largest welfare impacts on those who are least knowledgeable about financial products in the first place.

This paper speaks directly to the small, but growing, literature on the role of brokers and financial advisors in selling financial products. This literature is based on the premise that, in contrast to the market for consumption goods such as pizza, buyers of financial products need advice and guidance both to determine which product or products are suitable for them, and to select the best-valued product from the set of products that are suitable.

The theoretical literature can be divided into two strands: one posits that consumers are perfectly rational, understand that incentives such as commissions may motivate agents to recommend particular products, and therefore discount such advice. A second literature argues that consumers are subject to behavioral biases, and may not be able to process all available information and make informed conclusions.

Bolton et al. (2007) develops a model in which two intermediaries compete, each offering two products, one suitable for one type of clients, the other for the other type of clients. While intermediaries have an incentive to mis-sell, competition may eliminate misbehavior. Inderst and Ottaviani (2010) show that even in a fully rational world, producers of financial products will pay financial advisors commissions as a way to incentivize them to learn what products are actually suitable for their heterogenous customers. Del Guerico and Reuter (2010) take a different tack, arguing that sellers of mutual fund products in the US that charge high fees may provide intangible financial services which investors value.

A second, more pessimistic, view, argues that consumers are irrational, and market equilibria in which consumers make poorly informed decisions may persist, even in the face of competition. Gabaix and Laibson (2005) develop a market equilibrium model in which myopic consumers systematically make bad decisions, and firms do not have an incentive to debias consumers. Carlin (2009) explores how markets for financial products work in which being informed is an endogenous decision. Firms have an incentive to increase the complexity of products, as it reduces the number of informed consumers, increasing rents earned by firms. Inderst and Ottaviani (2011) present a model with naive consumers, where naivete is defined as ignoring the negative incentive effects of commissions, and find that naive consumers receive less suitable product recommendations.

The theoretical work is complemented by a small, but growing, empirical literature on the role of competition and commissions in the market for consumer financial products. In a paper that precedes this one, Mullainathan, Noth, and Schoar (2010) conduct an audit study in the United States, examining the quality of financial advice provided by advisors. Woodward (2008) demonstrates mortgage buyers in the U.S. make poor decisions while searching for mortgages. A series of papers (e.g. Choi et al 2009, 2010) demonstrate that consumers fail to make mean-variance efficient investment decisions, paying substantially more in fees for mutual funds, for example, than they would if they consistently bought funds from the low-cost provider. In work perhaps most closely related to this paper, Bergstresser et al. (2009) look at the role of mutual fund brokers in the United States. They find that funds sold through brokers underperform those sold through other distribution channels, even before accounting for substantially higher fees (both management fees and entry/exit fees). Buyers who use brokers are slightly less educated, but by and large similar to those who do not. They do not find that brokers reduce returns-chasing behavior.

In the next section we describe the basic economics of the life insurance industry in India, discuss why whole insurance policies are dominated by term policies, and economic theories of why individuals might still purchase whole policies. Section III discusses the theoretical framework that guides our empirical tests. Section IV presents the experimental design, while Section V and VI present our results. In section VII, we describe an equilibrium model of insurance markets in which dominated products survive, even with competition. Section VIII concludes.

2 Term and Whole Life Insurance in India

Life insurance products may be complicated. In this section, we lay out key differences between term and whole life insurance products, and demonstrate that the insurance offerings from the largest insurance company in India violate the law of one price, as long as an individual has access to a bank savings accounts. Rajagopalan (2010) conducts a similar calculation and also concludes that purchasing term insurance and saving strictly dominates purchasing whole or endowment insurance plans.

We start by comparing two product offerings from the Life Insurance Corporation of India (LIC), the largest insurance seller in India. For many years, LIC was the government-run monopoly provider of life insurance. We consider the LIC Whole Life Plan (Policy #2), and LIC Term Plan (Policy #190), for a 25-year old male seeking at least Rs. 2,500,000 in coverage (approximately USD $50,000), commencing coverage in 2010.

For a whole life policy, such a customer would make 55 annual payments (until the age of 80 is reached) of Rs. 55,116 (ca. $1,110 at 2010 exchange rates). The policy has a face value of Rs. 2,500,000 if the client dies before age 80. In case the client survives until age 80, which would be the year 2065, the product pays a maturation benefit equal to the coverage amount. The coverage amount is not necessarily constant: it may be increased via LIC's "bonus" policy, which the insurance company may declare if it earns profits. For the past several years, bonuses have ranged from 6.6% to 7% of the original coverage amount of the insurance policy. Unlike interest or dividends, these bonus payments are not paid to the client directly. Rather the bonus is added to the notional coverage amount, paid in case of death of the client, or, at maturity. The insurance company does not make any express commitment as to whether, and how much, bonus it will offer in the future.

A critical point to be made here is that the bonus is not compounded.[4] Rather, the bonus added is simply the amount of initial coverage, multiplied by the bonus fraction. For example, if the company declares a 7% bonus each year, the amount of coverage offered by the policy will increase by .07*2,500,000=Rs. 175,000 each year. Thus, after 55 years, when the policy matures, its face value will be Rs. 2,500,000 + 55*175,000=Rs. 12,125,000.

If these 7 percent bonuses were in fact compounded, the policy would have a face value of Rs. 2,500,000*1.07ˆ55, or over Rs. 103 million, an amount more than *eight times* larger. Stango and Zinman (2009) describe evidence from psychology and observed consumer behavior that individuals have difficulty understanding exponential growth. Consumers who do not understand compound interest may not appreciate how much more expensive whole life policies are.

A second feature of the two policies may be their relative attractiveness to naive, loss-averse consumers. Agents frequently dismissed term insurance as an option, arguing that the customer was likely to live at least twenty years, hence the premiums would be "lost" or "wasted," while with whole life the purchaser was guaranteed to get at least the nominal premium paid returned.

In Appendix Table 1, we evaluate the whole life insurance product by creating a replicating

[4]It is somewhat surprising that an insurance company has not entered this market and won a substantial amount of business by offering a whole insurance product that does pay compounded bonuses. In fact, there are some whole life products that pay a compounded bonus (i.e. the bonus rate is applied to both the sum assured amount plus all previously accumulated bonus); thus, it is not the case that the insurance industry is unaware that consumers might like these products. Rather, it seems that it is not possible for an insurance company to win substantial amounts of business by aggressively selling whole products that pay compounded bonuses. One explanation for this may be that competition really occurs along the margin of selling effort, as opposed to the quality of the product. In this case, the products that have highest sales incentives will sell, and any particular insurance firm will have an incentive to pay the highest commissions on the highest profit products. We present a formal model along these lines that is consistent with our empirical results later in this paper.

portfolio, which consists of a term insurance policy plus savings in a bank fixed deposit account. Each year, the replicating portfolio provides at least as much coverage (savings plus insurance coverage) as the whole policy, while requiring the exact same stream of cash flows from the client. A 25-year old man seeking coverage of Rs. 2,500,000 would pay Rs. 55,116 per year for whole insurance. If instead he bought a 35-year term policy with Rs. 4,000,000 in coverage, he would pay Rs. 11,996 each year for 35 years. Over that period, he could save the difference (55,116-11,996=43,120); once the term policy expired, the replicating portfolio would save Rs. 55,116 per year. In each year, the death benefit (of term payout, if the policy is active, plus savings) would be greater than the benefit from the whole policy, including the bonuses. The differences are dramatic: the initial coverage of the replicating portfolio is Rs. 4 million, vs Rs. 2.5 million for the whole policy. At age 35, the term plus savings is worth 9% more than the whole payout. By age 55, the replicating portfolio is worth 36% more than the whole payout, and by age 85 the replicating portfolio would be worth Rs. 91 million, compared to Rs. 13 million benefit from the whole policy. The replicating portfolio is almost seven times more valuable.

One argument commonly advanced in favor of whole life insurance is that it provides protection for the individual's whole life, and thus eliminates the need to purchase new term insurance plans in the future. If there is substantial risk that future term insurance premiums might increase due to increases in the probability of death, then term insurance might be seen as more risky than whole insurance. However, this argument does not affect our replication strategy, because the term plus savings plan does *not* require the individual to purchase another term insurance policy 35 years later.[5] The individual has saved up enough in the savings account to provide self-insurance after 25 years, which is greater than the amount of insurance that the whole life policy provides.

But even this comparison understates the difference in value dramatically, for at least two reasons. First, the replicating portfolio builds up a substantial savings balance, which is liquid. Second, if an individual does not pay each premium promptly, the insurance company has the right to declare the policy lapsed. Some estimates suggests lapse rates are high: 6% of outstanding policies lapse in a given year (Kumar, 2009). If the customer lapses after paying premiums for three or more years, the plan guarantees a recovery value of only 30% of premiums paid (less the

[5] Cochrane (1995) discusses this issue in the context of health insurance and proposes an insurance product that also insures against the risk of future premium increases due to changes in risk.

first year's premiums).

Thus, for an equivalent investment, the buyer receives up to six times as much benefit if she purchase term plus savings, relative to whole. We are not aware of many violations of the law of one price that are this dramatic. A benchmark might be the mutual fund industry: $1 invested in a minimal fee S&P500 fund might earn 8% per annum, and therefore be worth $69 after 55 years. If an investor invested $1 in a "high cost" mutual fund that charged 2% in fees, the value after 55 years would be $25, or about one third as large. The life insurance mark-up is thus by this metric twice as large as the mark-up on the highest cost index funds.

2.1 Whole Life Insurance as a Commitment Device

One potential advantage of the whole life policy over term plus savings is that the whole life policy contains committment features that some consumers value (Ashraf et al. (2006)). The structure of whole life plans impose a large cost in the case where premium payments are lapsed, and thus consumers that are sophisticated about their commitment problems may prefer saving in whole life plans versus standard savings accounts where there are no costs imposed when savings are missed. In particular, the LIC Whole Insurance Plan No. 2 discussed in the previous section returns nothing if the policy "lapses' within the first three years.

However, it is not clear that the commitment feature alone is sufficient to explain the popularity of whole life insurance. Ashraf et. al. (2006) finds only 25% of the population exhibit hyperbolic preferences. Moreover, there are other savings products in the Indian context that offer similar commitment device properties but substantially higher returns. Fixed deposit accounts involve penalties for early withdrawal. Public provident fund accounts require a minimum of Rs. 500 per year contribution, and allow the saver no access to the money until at least 7 years after the account is opened. If a saver does not contribute the 500 rupees in a particular year the account is consider discontinued, and the saver has to pay a 50 rupee fine for each defaulting year plus the 500 rupees that were missed as installments.

Finally, there is no reason a financial services provider could not offer commitment savings accounts without an insurance component. The fact that no such product has been developed in India or around the world suggests that this product is not simply satisfying demand for commitment savings.

Nevertheless, we acknowledge that a desire to commit may be relevant for some consumers. Hence, for any shopping visit in which we regard term insurance as the more appropriate product, the mystery shopper clearly told the insurance agent that she or he was seeking risk coverage at a low cost, rather than a savings vehicle.

3 Theoretical Framework

Our empirical work is motivated by recent theoretical work on the provision of advice to potential customers. Our paper tests two types of predictions that arise from this class of models. The first set of predictions concerns the quality of advice provided by commissions motivated agents. These models predict that at least some consumers will receive low quality advice; i.e. they will be encouraged to purchase an advanced product that has higher commissions but no real benefits to them (Inderst and Ottaviani, 2011, Gabaix and Laibson, 2005).[6] We test this by measuring the fraction of agents that recommend customers purchase whole insurance, even in the case where the customer is only seeking insurance for risk protection (i.e. we shut down any commitment savings channel).

The second set of predictions relates to how regulation and market structure affect the quality of advice. We test three predictions from the theoretical literature.

Our first test centers on the role of competition in the provision of advice. Inderst and Ottaviani (2011) and Bolton et. al. (2007) show that increased competition amongst agents who provide products and advice can improve the quality of advice for customers. On the other hand, Gabaix and Laibson (2006) show that increasing competition need not lead firms to unshroud product characteristics that hurt naive consumers. Our auditors vary the level of competition perceived by agents, by reporting whether their information about insurance comes from a friend (low competition), or from another agent from which our auditor is thinking of purchasing insurance (high competition).

Second, a large literature in economics predicts that competition between firms will induce

[6] While the Gabaix and Laibson (2006) paper does not explicitly deal with commissions, it does show that firms will not necessarily have the incentive to unshroud product attributes (such as commissions or low rates of return in our case) because unshrouding these will not necessarily win the firm business. In our case, the analogy would be that life insurance firms do not have the incentive to unshroud these attributes of whole insurance products because they would lose a substantial proportion of business to banks and other financial service providers if individuals move their savings out of life insurance.

firms to disclose all relevant information regarding products (Diamond (1985), Grossman (1989)). In these models, mandatory disclosure enforced by the government does not change consumer decisions and does not improve welfare. However, Inderst and Ottaviani (2011) argue that disclosure requirements can improve the quality of advice by essentially converting unaware customers into customers that are aware of how commissions can bias advice. We test how a disclosure requirement on commissions impacts financial advice by studying a particular type of insurance product, a Unit Linked Insurance Policy (ULIP), where agents were forced to disclose the commissions they earned after July 1, 2010.

Lastly, a key feature of the recent theoretical models in Inderst and Ottaviani (2011) and Gabaix and Laibson (2006) is the presence of two types of agents, with different levels of sophistication. Inderst and Ottaviani (2011) predict that these sophisticated types will receive better advice. We test this prediction by inducing variation in the level of sophistication demonstrated by the agent during the sales visit.

4 Experimental Design

4.1 Setting

In this section we describe the basic experimental setup common to the three separate experiments we ran in this study. All of the auditors used have at least a high school education. Intensive introductory training on life insurance was provided by a former financial products sales manager, and a principal investigator. Subsequently, each auditor was trained in the specific scripts they were to follow when meeting with the agents. Each agent's script was customized to match the agents true life situation (number of children, place of residence, etc.). However, agents were given uniform and consistent language to use when asking about insurance products, and seeking recommendations. Auditors memorized the scripts, as they would be unable to use notes in their meetings with the agents. Following each interview, auditors completed an exit interview form immediately, which was entered and checked for consistency. The auditors and their manager were told neither the purpose of the study, nor the specific hypotheses we sought to test.

Auditors were instructed not to lie during any of the sessions. Upon completion of the study, all auditors were given a cash bonus which they used to purchase a life insurance policy from the agent of their choice. All of our auditors chose to purchase term insurance.

In each experiment, treatments were randomly assigned to auditors, and auditors to agents. Note that because the randomizations were done independently, this means that each auditor did not necessarily do an equivalent number of treatment and control audits for any given variable of interest (i.e. sophistication and/or competition). Table 1 presents the number of audits, number of auditors, and number of life insurance agents for each separate treatment cell in each of our three experiments. Since we were identifying agents as the experiment proceeded, we randomized in daily batches. To ensure treatment fidelity, auditors were assigned to use only one particular treatment script on a given day.

Life insurance agents were identified via a number of different sources, most of which were websites with national listings of life insurance agents.[7] Contact procedures were identical across the treatments. While some agents were visited more than once, care was taken to ensure no auditor visited the same agent twice, and to space any repeat visit at least four weeks apart, both to minimize the burden on the agents, and to reduce the chance the agent would learn of the study.

Table 2 presents summary statistics across the three experiments we report results on in this paper. The Quality of Advice experiment was conducted in one major Indian city, and the Disclosure and Sophistication experiments were conducted in second major Indian city.[8] Across the experiments, between 50-75% of agents visited sold policies underwritten by the Life Insurance Company of India (LIC), a state owned life insurance firm. This fraction is consistent with LIC's market share, which was 66 percent of total premiums collected in 2010.

In terms of the location of the interaction between the auditor and the life insurance agent, one major difference between the Quality of Advice experiment and the Disclosure and Sophistication experiments is that a substantial number of Quality of Advice audits occurred at venues outside the agent's office. These other locations were typically a restaurant, cafe, railway or bus station, or public park. In the Disclosure and Sophistication experiments, the majority of audits took place at the agent's office. On average, each audit lasted about 35 minutes, suggesting these audits do represent substantial interactions between our auditors and the life insurance agents. The length

[7] We also included a small number of agents we found through outdoor advertisements and through a listing of Life Insurance Corporation of India agents.

[8] The Competition experiment was conducted as a sub-treatment within the Quality of Advice experiment, and thus shares the same summary statistics.

of audit did not vary substantially across the different experiments.

Matched pair audit studies used to identify discrimination have been criticized on methodological grounds. These studies, which involve sending, for example, black and white car buyers to purchase a car. Critics argue that even if auditors stick to identical scripts, they may exhibit other differences (apparent education, income, etc.) that could lead sales agents to treat buyers differently for reasons other than the buyer's race or sex (Heckman, 1998). While our study is not subject to this criticism–our treatments were randomized at the auditor level, so we can include auditor fixed effects–we took great care to address other potential threats to internal validity. Outright fraud from our auditors is very unlikely, as they were obliged to hand in business cards of the sales agents. To monitor script compliance, we paid insurance agents within the principal investigators' social network to "audit the auditors"–these agents reported that our auditors adhered to scripts. The outcome we measure, policy recommended, is relatively straightforward, and auditors were instructed to ask the agent for a specific recommendation. To prevent auditor demand effects, we did not inform the auditors of the hypotheses we were interested in testing.

5 Quality of Advice

5.1 Quality of Advice: Catering to Beliefs Versus Needs

In this experiment we test the sensitivity of agents' recommendations to the actual needs of consumers, as well as to consumers potentially incorrect beliefs about which product is most appropriate for them. In particular, one reason agents may recommend whole insurance is a belief that customers will value the commitment savings features. To examine this, we vary the expressed need of the agent, by assigning them one of two treatments. In half of the audits, the auditor signals a need for a whole insurance policy by stating: "I want to save and invest money for the future, and I also want to make sure my wife and children will be taken care of if I die. I do not have the discipline to save on my own." Good advice under this treatment might plausibly constitute the agent recommending whole insurance. In the other half of the audits, the auditor says "I am worried that if I die early, my wife and kids will not be able to live comfortably or meet our financial obligations. I want to cover that risk at an affordable cost." In this case the auditor demonstrates a real need for term insurance. By comparing agent recommendations across these two groups, we can measure whether agent recommendation responds to agents true needs. Appendix Table A2 presents the exact wording of all of the experimental treatments in this study.

We also randomized the customer's stated beliefs about which product was appropriate for him or her. In audits where the auditor was to convey a belief that whole insurance was the correct product for them, the auditor would state "I have heard from [source] that whole insurance may be a good product for me. Maybe we should explore that further?" In the audits where the auditor was to convey a belief that term insurance was the correct product for them, the auditor would state "I have heard from [source] that whole insurance may be a good product for me. Maybe we should explore that further?"

Finally, to understand the role of competition, we also varied the source auditors mentioned when talking about their beliefs. In the low competition treatment, the auditor named a friend as a source of the advice. In the high competition treatment, the auditor said the suggestion had come from another agent from whom the auditor was considering purchasing.

Each of these three treatments (product need, product belief, and source of information) was assigned orthogonally, so this experiment includes eight treatment groups.

Table 3 presents a randomization check to see if there are important differences in the audits that were randomized into different groups. The first two columns compare audits that were randomized such that the auditor had either a bias for term (Column (1)) or a bias for whole (Column (2)). As would be expected given the randomization, there are almost no systematic differences across the two groups. The only significant difference is that audits assigned a bias towards whole were approximately two percentage points more likely to be conducted at the auditor's home. We include audit location fixed effects in our specifications and find they do not substantially change the results.

Columns (3) and (4) present characteristics of audits where the auditor was randomized into having a need for term insurance (Column (3)) or a need for whole insurance (Column (4)). The next two columns present the pre-treatment characteristics of audits where the source of the bias was another agent (Column (5)) or a friend (Column (6)). There are also no statistically significant differences in the pre-audit characteristics across these groups.[9]

[9] Throughout the paper, we use robust standard errors; results and significance levels are virtually identical if we cluster standard errors at the level of randomization, auditor*day.

Before describing the experimental results, we emphasize how poor the quality of advice is: for individuals for whom term is the most suitable product, only 5% of agents recommend purchasing only term insurance, while 74% recommend purchasing only whole. A previous version of this paper documented a range of wildly incorrect statements made by agents, such as "term insurance is not for women;" "term insurance is for government employees only." One even proposed a policy that he described as term insurance, which was in fact whole insurance.

Table 4 presents our main results on how variation in the needs of customers and biases of customers affect the quality of financial advice.[10] Column (1) presents results on whether the agent's final recommendation included a term insurance policy (in about 8% of the cases, agents recommend the consumer purchase multiple products). We find that agents are 10 percentage points more likely to make a final recommendation that includes a term insurance policy if the auditor states that they have heard term insurance is a good product. We also find that agents are 12 percentage points more likely to make a recommendation that includes a term insurance policy if the auditor says they are looking for low-cost risk coverage. Both of these results are statistically significant at the 1 percent level. The interaction of these two variables is statistically insignificant. This suggests that agents are just as likely to cater to beliefs as needs.

In column (2), we add auditor-fixed effects and controls for venue and whether the agent sells policies underwritten by a government-owned insurer. The experimental results are unaffected. Agents from the government owned insurance underwriters (primarily the Life Insurance Corporation of India) are 12 percentage points less likely to recommend a term insurance plan as a part of their recommendation.

Column (3) presents the same exact specification as Column (1), however now the dependent variable takes a value of one if the agent recommended only a term insurance plan. We find much weaker results here. A customer stating that they have heard that term insurance is a good product is only 2 percentage points more likely to receive a recommendation to only purchase term insurance. We find that stating a need for affordable risk coverage only causes a 1.5 percentage point increase in the probability that the agent will recommend exclusively term insurance. This effect is not statistically significant at conventional levels. When the auditor both states that they

[10]In this section we focus on the quality of advice given, and thus report results on how advice responds to a customer's needs versus beliefs. Later, we discuss the impact of the competition treatment when we focus on how quality of advice might be improved.

need risk coverage and they have heard that term is a good product we find an increase of 5.3 percentage points, significant at the ten percent level. Column (4) adds controls.

Thus, comparing Columns (2) and (4) it appears that agents do respond to both the biases and needs of customers, however, they primarily do it by recommending term insurance products as an addition to whole insurance products, rather than recommending the purchase of term.

Overall, the results in Columns (1) - (4) suggest that agents will respond approximately equally to both the needs and pre-existing biases of customers. These results are consistent with the idea that agents maximize the expected revenue from an interaction, and the expected revenue depends both on the probability that the customer will purchase as well as the amount of commission that can be earned. Agents do not seem to attempt to de-bias customers who express perceived needs inconsistent with actual needs; thus, in this context it seems unlikely that commissions motivated agents are effective in undoing behavioral biases customers bring to their insurance purchase decisions.

Columns (5) and (6) shows that stating an initial bias towards term insurance causes the agent to recommend the customer purchase approximately 13 percent more risk coverage, while expressing a need for risk coverage increases the recommended risk coverage by 17 percentage points. Both of these effects are significant at the five percent level, but their interaction is not. Again, these results suggest agents will cater approximately equally to the stated preferences of a customer (even if those preferences are inconsistent with their actual needs), about as much as they cater to the actual stated needs of customers.

Columns (7) and (8) test whether the recommended premium amounts are statistically different across the treatments. We find that the bias and need treatments have small and statistically insignificant effects on the level of premiums the agent recommends that customers pay to purchase insurance. This suggests that although agents are recommending higher coverage levels for those who either have a bias towards term or a need for term (Columns (5) and (6)), customers are not paying higher premiums to obtain this additional coverage. Instead, the increase in risk coverage observed in Columns (5) and (6) is due primarily to the fact that term insurance provides dramatically more risk coverage per Rupee of premium.

Further evidence of this interpretation is obtained from the average amounts of risk coverage and premium amounts when agents recommended term versus whole insurance (not reported). In the case where the auditor sought risk coverage at an affordable cost and said they had heard risk coverage was a good product for them, agents recommending term insurance proposed 2.3 million rupees of risk coverage, with an annual premium cost of approximately 31,000 rupees. Agents recommending whole insurance suggested customers purchase 522,000 rupees of risk coverage, with an annual premium of approximately 28,000 rupees. Our auditors characteristics (income, dependents) are the same no matter what beliefs they express, meaning there is no economic reason to suggest greater coverage levels when the auditor expresses a preference for coverage at low cost. One explanation for this result, consistent with the bad advice hypothesis, is that agents base their recommendations on the amount of premiums customers can pay, as opposed to the amount of risk coverage customers actual need. Our finding here is consistent with anecdotal evidence from discussions with our auditing team: agents typically start the life insurance conversation by estimating how much the individual can afford to put into life insurance per month, rather than determining how much risk coverage the customer needs.

In summary, we find the following. Despite the fact term is an objectively better policy, between 60 and 80 percent of our visits end with a recommendation that the customer purchase whole life insurance. Second, even when customers signal that they are most interested in term insurance and need risk coverage, more than 60 percent of audits result in whole insurance being recommended. Third, we find that agents primarily cater to customers (either their beliefs or needs) by recommending that they purchase term insurance in addition to whole insurance, as opposed to recommending term insurance alone. It is difficult to see how combining term and whole insurance makes sense for someone who is seeking risk coverage.

6 Financial Advice and Market Structure

These previous results are consistent with the models of Inderst and Ottaviani (2011), Gabaix and Laibson (2006) and Bolton et al. (2007) which suggest commissions motivated sales agents will have an incentive to recommend more complicated, but potentially unsuitable, products to customers who are not wary of the agency problems that commissions create (at least under some market structures). In this section we turn to testing theoretical predictions on how advice responds to the regulatory and market structure. As our experimental design allows us to measure the type of advice given, we focus on three predictions. First, the threat of increased competition from another agent will reduce the probability an unsuitable product is recommended. Second, increasing consumers awareness of commissions will reduce the tendency to recommend unsuitable products. Third, agents will provide different advice to sophisticated versus unsophisticated consumers.

6.1 Competition

One way agents may compete with each other is to offer better financial advice. Standard models of information provision suggest that competition amongst advice providers will lead to the optimal advice being given; customers will avoid salesmen who give low quality advice and thus in equilibrium only high quality advice will be given.

In any given interaction between an agent and a customer, it is likely that the agent perceives he has some market power, in that the customer would have to pay additional search costs to purchase from another agent. In this treatment we attempted to experimentally reduce the agent's perceived amount of market power by varying whether the customer mentions that they have already spoken to another agent. Audits randomized into the high competition treatment stated that they heard from another agent term (or whole) might be a good product for them. Audits randomized into the low competition treatment state that they heard from a friend that term (or whole) might be a good product for them.

The audits for which these data are based on are the same as those used in the Quality of Advice experiment. Table 5 presents our results on the impact of greater perceived competition on the quality of advice provided by life insurance agents. The specifications reported here are the same as those in Table 4, but we now introduce a dummy variable that takes the value of 1 if the auditor's bias came from a competing agent, and zero if the bias came from a friend. Columns (1) and (2) show that overall the induced competition does not seem to have an important effect on whether agents recommend term insurance as part of their package recommendation. Columns (5) and (6) show that the competition treatment also did not have an overall increasing effect on whether only a term policy was recommended.

Columns (3) and (4) introduce a set of interaction terms between the bias treatment, the need treatment, and the competition treatment. We are particularly interested in the treatment where the customer is biased towards whole insurance but demonstrates a need for term insurance.

In this setting the agent has the potential to "de-bias" the auditor as their beliefs are inconsistent with their insurance needs. In Columns (3) and (4) we find that the agent is substantially more likely to debias agents when the threat of competition looms. This effect is measured by summing the coefficients on the variables Competition and (Need=Term)*Competition. The sum suggests agents advising customers who need term but are biased towards whole are 10 percent more likely to recommend term insurance if they perceive higher levels of competition. The hypothesis that (Need=Term)*Competition + Competition = 0 can be rejected at the 5% level. This result suggests that if perceived competition is high enough, agents will attempt de-bias customers as a way of winning business.

We do not, however, find that competition increases the possibility that agents will de-bias customers who have a belief that term insurance is a good product but need help with savings. We find that the coefficient on the interaction (Bias=Term)*Competition is small and statistically insignificant.

Columns (7) and (8) report the same specification as those in Columns (3) and (4), however the dependent variable takes the value of one if the agent recommended the customer purchase only term insurance. We do not find any evidence that agents attempt to de-bias consumers by recommending they only purchase term insurance. The coefficient on the interaction term (Need=Term)*Competition is small and insignificant in Columns (7) and (8). We find that the competition treatment is only effective, in this case, when the agent has both a bias and a need towards term insurance. One interpretation of this result is that agents assume that a customer who has the knowledge to know that term insurance is the best product for someone who needs risk coverage is almost surely going to purchase term insurance from the other agent. Thus, the agent in the audit chooses to compete by recommending only a term insurance purchase as well.

6.2 Disclosure

On July 1, 2010, the Indian Insurance Regulator mandated that insurance agents must disclose the commissions they would earn when selling a specific type of whole insurance product called a ULIP. ULIPs are very similar to whole insurance policies, except the savings component is invested in equity instruments with uncertain returns. This regulation was enacted as the Indian insurance regulator faced criticism from the Indian stock market regulator that ULIPs should be regulated in the same was as other equity based investment products. The insurance regulator responded to these criticisms by requiring agents to disclose commissions when selling ULIPs.

There are two specific features of this policy we emphasize before discussing our empirical results. First, it is important to note that the disclosure of commissions required on July 1st is in addition to a disclosure requirement on total charges that came into effect earlier in 2010. In other words, prior to July 1, agents were required to disclose the total charges (i.e. the total costs, including commissions) of the policies they sell, but they were *not* required to disclose how much of those charges went to commissions versus how much went to the life insurance company. Thus, the new legislation requiring the specific disclosure of commissions gives the potential life insurance customer more information on the agency problem between himself and the agent, but does not change the amount of information on total costs. This allows us to interpret our results as the effect of better information about agency, rather than better information about costs more generally.

To focus the visits on ULIPs, agents began by inquiring specifically about ULIP products available. The experimental design here involves two components. First, we conducted audits before and after this legal change to test whether the behavior of agents would change due to the fact that they were forced to disclose commissions. Second, we also randomly assigned each of these audits into two groups, where in one group the auditor conveys knowledge of commissions and in the other group the auditor does not mention commissions. We created these two treatments as we believed only customers who have some awareness of these commissions were likely to be affected by this law change. In one group, we had the auditor explicitly mention that they were knowledgeable about commissions by stating: "Can you give me more information about the commission charges I'll be paying?" In the control group, the auditor did not ask this question about commission charges.

Table 6 presents summary statistics on the disclosure experiment audits. Column (1) pertains to the full sample audits, while (2) and (3) present summary statistics on the audits before and after the regulation went into effect. There are several differences between the pre- and post- audits. In particular, post disclosure change audits were more likely to be conducted with the Life Insurance Company of India, and the meetings took place in different venues. These differences suggest that caution is warranted when comparing the pre- and post- results. Columns (7) and (8) of Table 3 present summary statistics on the randomization of the different levels of knowledge about commissions.

6.3 Did the Disclosure Requirement Change Products Recommended?

We first examine whether audits conducted after the disclosure requirements went into effect were less likely to result in the agent recommending a ULIP policy. Figure 1 shows the weekly average fraction of audits that resulted in a ULIP recommendation. Prior to the commissions disclosure reform, agents recommended ULIPs eighty to ninety percent of the time. Following the reform, there is an immediate and discrete drop in the fraction recommending ULIPs, to between forty and sixty-five percent of audits. The discrete jump suggests the observed differences are driven by the disclosure requirement, rather than being attributable to a steady downtrend trend in the fraction of agents recommending ULIP policies over time.

Table 7 presents the formal empirical results. The dependent variable in all specifications in this table takes a value of one if the agent recommended a ULIP product and zero otherwise. The independent variable Post Disclosure indicates whether or not the audit occurred after the legislation went into effect, July 1st (our earliest post-disclosure audits occurred on July 2nd). The variable Disclosure Knowledge equals one where the client expresses awareness that agents receive commissions and zero otherwise. Finally, we control for whether the agent is from a government underwriter, auditor fixed effects, and the location of the audit.

Column (1) presents a regression without controls. We find that in the post period a ULIP product was 25 percentage points less likely to be recommended. This finding is consistent with the prediction that agents treat customers who are concerned about commissions differently than those who are not, and that disclosure policy can improve customer awareness. We do not find the randomized treatment of the auditor demonstrating knowledge of the commissions significant (Disclosure Knowledge), nor do we find the interaction to be significant.

One potential threat to the validity of our analysis is the change in composition of agents between the pre- and post-period. Perhaps most important is the difference between the fraction of agents selling policies issued by government-owned insurance companies before and after the law change. In Column (2), we control for whether the agent works for a government-run insurance company, as well as location and auditor fixed-effects. The point estimate is slightly smaller, but the effect is still quite sizeable at 19 percentage points.

In columns (3) and (4) we examine agents for government-owned and private insurance companies seperately. Among those selling policies underwritten by government-owned companies, there is a 30 percent decrease in the likelihood of recommending a ULIP policy after the disclosure law becomes effective. Amongst private underwriters, we find a negative point estimate, although the coefficient is not significant at standard levels. The result in Column (3) suggests that the observed reduction in ULIP recommendations in the whole sample is not driven by a compositional shift in the types of agents the auditors meet.

In terms of magnitudes, given the overall percentage of ULIP recommendations in this sample was 71 percent, the approximately 20 percent decrease in ULIP recommendations once disclosure commission became mandatory is an economically large effect. Further analysis (not reported) finds agents were approximately 20 percentage points more likely to recommend whole insurance type products following the law change. There was no change in their propensity to recommend term insurance. Thus, it appears that the ULIP disclosure law change primarily led to substitution away from high commission ULIP products to high commission whole insurance products.

Turning to the experimental treatment, we do not find that audits where our agents showed knowledge of the new disclosure requirements are associated with lower levels of ULIP recommendations. The coefficient on the Disclosure Knowledge variable is small and statistically insignificant in all of the specifications. This treatment does not seem to be affected by the disclosure requirement.

Columns (5) and (6) test whether the commission disclosure requirement had important impacts on the amount of risk coverage and premium payments recommend by agents. We find no statistically significant differences here, suggesting that the types of products recommended were similar in terms of their risk characteristics after the policy change.

6.4 Customer Sophistication

In our final experiment, we manipulated the the level of sophistication about life insurance policies projected by the auditor. Each auditor was randomly assigned to portray either high or low levels of sophistication.

Sophisticated auditors say:

"In the past, I have spent time shopping for the policies, and am perhaps surprisingly somewhat familiar with the different types of policies: ULIPs, term, whole life insurance. However, I am less familiar with the specific policies that your firm offers, so I was hoping you can walk me

through them and recommend a policy specific for my situation."

Unsophisticated agents, on the other hand, state:

"I am aware of the complexities of Life Insurance Products and I don't understand them very much; however I am interested in purchasing a policy. Would you help me with this?"

To ensure clarity of interpretation of the suitability of recommendations, we built into the auditors script several statements that suggest a term policy is a better fit for the client. Specifically, the auditor expressed a desire to maximize risk coverage, and stated that they did not want to use life insurance as an investment vehicle.

We predict that individuals that are sophisticated about life insurance products will be more likely to receive truthful information from life insurance agents; agents internalize that sophisticated agents are not swayed by false claims, and thus presenting dishonest information to sophisticated agents is wasted persuasive effort. In the specific context of our audits this prediction suggests that life insurance agents should be more likely to recommend the term policy to sophisticated agents. Note that we designed our scripts so sophistication here only means that the potential customer is knowledgeable about life insurance products; both sophisticated and unsophisticated agents state that they have the same objective needs in terms of life insurance.

Table 3 presents a randomization check for the Sophistication experiment. The only statistically significant different between the sophisticated and non-sophisticated treatments is that the sophisticated treatments were about eight percentage points less likely to occur at other venues. Overall, the randomization in this experiment appears to be successful. We control for audit location in our results and find this has little impact on the effect of sophistication on recommendations.

The results from the sophistication experiment, reported in Table 8, provide some evidence in support of our prediction that sophisticated customers will receive better advice. We use the same specification as in the previous experiments to analyze this data. In Column (1) the dependent variable takes a value of one if the agent's recommendation included a term insurance plan, and zero otherwise. We find that the sophisticated treatment causes a ten percentage point increase in the likelihood that an agent includes term insurance as a part of their recommendation. This result is statistically significant at the 10 percent confidence level. In Column (2) we include a set of control variables, the point estimate and confidence interval are virtually unchanged. Thus, we do see that agents make some attempt to cater to sophisticated individuals by offering term insurance.

However, in Columns (3) and (4), where the dependent variable takes a value of one if the agent recommended the auditor purchase only a term a insurance plan, we find there is no statistically significant effect of sophistication. Similar to the results in the bias versus needs experiment, it appears that agents attempt to cater to more sophisticated types by including term as a part of a recommendation. However, they do not switch to recommending only term insurance, even to customers who signal sophistication.

In Columns (5) and (6) we look at the impact of sophistication on the amount of coverage recommended by the life insurance agent. Without controls, we find that sophisticated agents receive guidance to purchase approximately 22 percent more insurance coverage (Column (5)). In Columns (7) and (8) we test whether sophisticated agents receive different recommendations in terms of how much premiums they should pay for insurance. We find that signaling sophistication does not have an important impact on the amount of premiums that agents recommend paying, although the confidence interval admits economically meaningful effects of up to 25 percent lower premium costs. Combining the results in Columns (5) - (8), we see that, similar to our results on coverages and premiums in the other experiments, agents seem to recommend approximately the same amount of premiums be paid, regardless of our intervention; they cater to customers primarily by adding a relatively inexpensive term product on top of whole insurance to increase risk coverage without substantially changing premium payments.

7 A Model of Commissions, Bad Advice, and Dominated Products

We, and others, have argued that whole life insurance is dominated by term insurance for individuals who seek insurance mainly for risk coverage. While the goal of this paper is to understand commissions motivated agent behavior (rather than offer a competitive analysis of the Indian insurance industry), it does raise a puzzle: why do the more expensive, dominated, products, such as whole insurance, persist in a setting with competition? We consider here how a dominated product could survive, even in a competitive equilibrium.

We present a simple model, inspired by Gabaix and Laibson (2006), which provides one explanation for how a dominated financial product might exist in competitive equilibrium. The model takes the empirical results found in this paper, that commissions motivated agents appear to provide poor financial advice, and shows how it is possible that if at least some consumers are persuaded by bad advice then it is possible that a dominated product like whole insurance could persist. The model may be particularly relevant for a country like India with a large number of new insurance customers entering the market who are still learning about these products and may be less sensitive to important differences in the long run returns available.

In the model, we focus primarily on the risk coverage offered by the insurance products. The price of term insurance is the premium, while the "price" of whole insurance should be thought of as the premium cost minus any savings value that exists beyond the risk coverage. This is equivalent to assuming whole insurance can be replicated by purchasing term insurance and investing in a savings account. Thus, the model is set up such that buyers should choose whole insurance only if the price is cheaper than term insurance. However, we show that an equilibrium is possible where whole insurance has a higher price than term insurance.

The model has two types of consumers. Sophisticated consumers understand that whole and term insurance are the same product (and thus would always choose the cheaper one), know their own optimal amount of insurance, given prices, and are immune to the persuasive efforts of agents. There is a fixed, exogenous number of sophisticated consumers, s, who want to purchase term insurance, and each has a demand function for term insurance equal to $\alpha - p_t$, where p_t is the price of term insurance.

Unsophisticated consumers, in contrast, can be persuaded to purchase a dominated product if there is an agent that exerts enough effort. In particular, we assume unsophisticated agents demand an amount of insurance $\alpha - p_w$ once they have met with a commissions motivated agent. Agents must exert effort to identify and sell to unsophisticated consumers. We assume that the number of customers they find is equal to the commission on selling insurance set by the insurance company, c. Intuitively, the higher that the insurance firm sets commissions, the more incentive agents have to approach customers and sell insurance. In addition to commissions payments, the insurance firm incurs an underwriting cost of k per unit of either term insurance or whole insurance sold.

The game play is as follows. In period 0, the firm(s) choose whether to offer term, whole, or both insurance products. They also choose the prices p_w and p_t and the commissions they will pay agents to sell whole and term insurance (c_w, c_t). In the second period, agents respond to the incentives set by the insurance companies, and consumers make decisions on how much whole and term insurance to purchase and insurance. An Appendix contains the proofs of all the results discussed here.

7.1 Monopolist Insurance Company

A monopolist insurance firm has three possible options (1) offer only term insurance (2) offer whole and term insurance (3) offer only whole insurance. In the Appendix we show that the monopolist insurance firm will choose to offer both term and whole insurance. The monopolist firm will pay zero commissions for the sale of term insurance (as paying commissions on term insurance does not increase demand) and will charge a price of $\frac{\alpha+k}{2}$ for term insurance. The monopolist firm will pay positive commissions for the sale of whole insurance because demand is increasing in commissions. The firm will set the whole insurance price (p_w) equal to $\frac{1}{3}(2\alpha + k)$ and will pay commissions $\frac{1}{3}(\alpha - k)$. Note that as long as $\alpha > k$ (a condition necessary for there to be positive demand for insurance), that the price of whole insurance will be higher than the price of term insurance.

The intuition for this solution is that offering both term and whole insurance offers the monopolist firm a way to set different commissions and prices for sophisticated versus unsophisticated customers. Sophisticated consumers cannot be persuaded by commissions motivated agents, and thus the firm chooses to set commissions to zero and charge lower prices for term insurance. However, unsophisticated consumers can be persuaded to purchase whole insurance. Thus, the insurance firm chooses to pay higher commissions to encourage agents to persuade consumers to purchase insurance, and then passes these higher commissions onto the consumer in terms of higher prices.

7.2 Two Competing Insurance Companies

We now analyze the impact of competition by considering a Bertrand pricing game where two firms compete by setting term and whole commissions and prices. This game has two players, firm i and firm j. A strategy in this game consists of (1) a choice of which products to offer (term, whole, or both) (2) prices and commissions for each product offered. A firm's payoff function is the profit it earns given its choice of what products, prices, and commissions to offer as well as the other firm's choices.

The payoffs are defined as follows. For term insurance, we use the usual Bertrand pricing game (with homogenous products) assumption that firm i obtains the full market of all s sophisticated consumers if $p_i < p_j$ (and vice versa). For whole insurance, consumers can be influenced to purchase both by higher commissions and lower prices. The number of unsophisticated consumers that firm i sells to given it pays commissions c_i is $c_i - bc_j$. The parameter b, which we assume is always greater than zero, measures the degree to which firm i and j's insurance products compete with each other for customers. If b equals zero then the fact that firm j is paying high commissions does not change the demand for firm i's insurance. If b is large, however, then an increase in commissions by firm j causes a fraction of consumers to switch from firm i's insurance product to firm j's product.

Note, however, that once unsophisticated consumers have been persuaded to purchase from a particular firm because of commissions, the insurance company can charge them the monopoly price. In this sense, competition for unsophisticated consumers happens primarily through commissions, and not through prices. The intuition is that unsophisticated consumers respond strongly to the persuasiveness and effort of agents in choosing what product to buy, but less strongly to the level of prices.

Bertrand competition over prices in the market for term insurance leads to both firms pricing term insurance at marginal cost k. In the Appendix we show that the Nash equilibrium commissions on whole insurance are $c_i^* = c_j^* = \frac{\alpha - k}{3-2b}$, and the Nash equilibrium prices are $p_i^* = p_j^* = \frac{(2-b)\alpha + (1-b)k}{3-2b}$. Note that for commissions and prices to be positive we need $b \leq \frac{3}{2}$.

Even though term and whole insurance are the same product in this model, an equilibrium exists where whole insurance has a higher price than term insurance, and where competition between firms will not eliminate this dominated product. Analogous to the result in Gabaix and

Laibson (2006), a strategy of un-shrouding the whole policy does not work because selling the dominating term policy does not offer the margins necessary to pay large commissions. Thus, it is not profitable for firms to educate consumers on the fact that whole insurance is simply an expensive version of term insurance. In equilibrium, firms sell low commission term insurance to sophisticated consumers, and high commission whole insurance to unsophisticated consumers.

The model also has an interesting prediction on the impact of competition in this market. When paying commissions causes the competitor to lose more business (b increases), competition amongst firms leads to an increase in commissions and prices.[11] Thus, when insurance firms attract customers mainly through commissions, competition can actually lead to higher prices (and commissions), relative to a monopoly provider. The intuition for this result is that as a monopoly provider, paying higher commissions loses more in profits due to higher costs than it gains in extra business. However, when firms compete over commissions, then it becomes necessary to pay higher commissions to win business, and profits for each sale are lower because more commissions have to be paid.

We believe this model is a plausible explanation for why a dominated product like whole insurance can persist in this market. The model fits the basic empirical facts observed in this market: 1) Term insurance and whole insurance co-exist, although whole insurance can be replicated by term insurance and savings accounts 2) Commissions on whole insurance are substantially higher than term insurance 3) Agents provide poor advice (i.e do not try to de-bias consumers towards whole insurance) 4) The industry has multiple, seemingly competitive, insurance providers. Nonetheless, further empirical work is necessary to distinguish the model presented from other potential explanations for the existence of dominated products, such as entry barriers or other market frictions.[12]

8 Conclusion

A critical question facing emerging markets with large swaths of the population entering the formal financial system is how these new clients will receive good information on how to make financial

[11]See appendix for the proof that prices increase.

[12]It is important to note that the Indian insurance industry is characterized by significant barriers to entry, including licensing restrictions and capital requirements, as well as scale economies.

decisions. Clearly, the private sector will be important in educating new investors and providing suitable products. Recent events in developed economies suggest that regulation or improved consumer awareness may be necessary to ensure that the private sector's own incentives do not compromise the quality of financial decisions made by private individuals. This issue is of particular importance in emerging markets where new investors have little experience with formal financial products to begin with.

In this paper, we show that whole life insurance is economically inferior to a combination of investing in savings accounts and purchasing term insurance. Despite the large economic losses associated with investing in whole insurance we find that life insurance agents overwhelmingly encourage the purchase of whole insurance.

We then use an audit study to test two types of predictions emerging from recent theoretical models on commissions and financial advice. The first prediction is that agents will have an incentive to recommend more expensive, less suitable, products to consumers. Throughout our three experimental designs, we find that life insurance agents rarely recommend term insurance. Even in audits where there should be no commitment savings motivation, we still find agents predominantly recommend whole insurance.

We also find that agents cater to customers' pre-conceptions of what the right product is for them as much (if not more) than to objective information about what the right product is. This suggests that, at least in our sample, agents do not actively try to de-bias customers. This result holds even in the case where an agent has an incentive to de-bias the customer because a de-biased customer would purchase a higher commission product. These results suggest that relying on competition to de-bias consumers of their mis-conceptions may not lead to markets that inform consumers.

We find that government underwriters are much more likely to recommend the dominated product. We view the government underwriter result as important. Government ownership is sometimes advanced as a solution to market failures, yet in this setting, agents representing government underwriters, in particular the Life Insurance Company of India, were much less likely to recommend a suitable product.

We then proceed to test predictions on how changes in the regulatory and market structure can affect advice given by financial agents.

We test the theoretical mechanism that competition amongst agents can lead to better advice. As mentioned above, the first order fact seems to be that competition does not suffice to motivate agents to provide good advice in this context. In an experiment,we find that increasing the apparent level of competition does lead to the agent attempting to de-bias the customer by offering term insurance. This also suggests that encouraging customers to shop around when looking for consumer financial products may be a simple way to improve the quality of advice provided by agents.

In another experiment we find that requiring disclosure of commissions on one particular product led to that product being recommended less. This result is interesting in that it suggests that hiding information may be an important part of life insurance agents' sales strategy, and that disclosure requirements can change the optimal strategy of agents. In this case it appears that the disclosure requirement on one product simply had the effect of pushing agents to recommend more opaque products. These results suggest that the disclosure requirements for financial products need to be consistent across the menu of substitutable products.

Lastly we find that agents who signal sophistication by demonstrating some knowledge of insurance products get better advice. Auditors that stated they had a deep understanding of insurance products were 10 percentage points more likely to receive a recommendation that included term insurance. This result suggests that the worst educated may suffer most from commission-motivated sales behavior. Further, it suggests that agents may play an important in helping financial firms discriminate between sophisticated and unsophisticated consumers, which can be valuable if unsophisticated consumers can be persuaded to purchase dominated products.

We present an equilibrium model where a dominated financial product, such as whole insurance, could persist. The key ingredients of this model are the existence of at least some customers who can be persuaded to purchase the dominated product; competition amongst firms leads to agents being paid higher commissions to sell the product, and the higher commissions are passed on to unsophisticated consumers through higher prices. We believe that this type of model may have wider applicability across a range of settings where customers are uninformed about the suitability or value of products.

We believe our study opens some important questions for further research. First, how effective is the persuasive power of agents? How important are behavioral biases such as loss aversion and exponential growth bias in driving demand for a dominated product? In the spirit of Bertrand and

Morse (2011), could consumers be debiased? The answers to these have important implications for optimal regulatory policy and household financial decision-making.

9 References

Ashraf, N., D. Karlan and W. Yin. 2006. Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines." *The Quarterly Journal of Economics* 121(2): 635–672.

Bergstresser, D., J. Chalmers, and P. Tufano. 2009. Assessing the Costs and Benefits of Brokers in the Mutual Fund Industry. *The Review of Financial Studies* 22(10):4129–56.

Bertrand, Marianne, and Adair Morse (2011). "Information Disclosure, Cognitive Biases, and Payday Borrowing," *The Journal of Finance*, 66(6): 1865-1893.

Bolton, P., X. Freixas, and J. Shapiro. 2007. Conflicts of Interest, Information Provision, and Competition in the Financial Services Industry. *Journal of Financial Economics* 85(2):297–330.

Carlin, B. 2009. Strategic Price Complexity in Retail Financial Markets. *Journal of Financial Economics* 91:278–56.

Choi, J.J., D. Laibson, and B.C. Madrian, and A. Metrick. 2009. Reinforcement Learning and Savings Behavior. *Journal of Finance* 64(6): 2515-34.

Choi, J.J., D. Laibson, and B.C. Madrian. 2010. Why Does the Law of One Price Fail? An Experiment on Index Mutual Funds. *Review of Financial Studies* 23(4):1–28.

Cochrane, J. 1995. Time-consistent Health Insurance. *Journal of Political Economy* 103(3):445-473.

Del Guerico, D., J. Reuter, and P. Tkac. 2010. "Broker Incentives and Mutual Fund Market Segmentation." Manuscript, Boston College.

Diamond, D. 1985. Optimal Release of Information by Firms. *Journal of Finance* 40(4): 1071-94.

Gabaix, X., and D. Laibson. 2006. Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets. *Quarterly Journal of Economics* 121(2):505–40.

Grossman, S. 1989. The Informational Role of Warranties and Private Disclosure about Product Quality, In *The Informational Role of Prices*, edited by Sanford J, Grossman, Cambridge, MA: MIT Press.

Heckman, James, 1998. Detecting Discrimination, Journal of Economic Perspectives 12(2):101-116.

Inderst, R., and M. Ottaviani. 2010. Intermediary Commissions and Kickbacks. *American Economic Review*, forthcoming.

Inderst, R., and M. Ottaviani. 2011. How (Not) to Pay for Advice: A Framework for Consumer Financial Protection. mimeo, Northwestern University.

Insurance Regulatory and Development Authority (IRDA) of India. 2010. Annual Report.

Kumar, Jagendra, 2009. "Lapsation of a Life Insurance Policy," Bimaquest, 9(2): 38-44.

Mullainathan, S., M. Noth, and A. Schoar. 2010. The Market for Financial Advice: An Audit Study, mimeo, Harvard University.

Rajagopalan, R. 2010. Comparing Traditional Life Insurance Products in the Indian Market: A Consumer Perspective, mimeo, CORE Centre, India.

Stango, V. and J. Zinman. 2009. Exponential Growth Bias and Household Finance. *Journal of Finance* 64(6): 2807-49.

Woodward, S. 2008. Consumer Confusion in the Mortgage Market, Manuscript, Sand Hill Econometrics.

10 Appendix

11 Model of a Dominated Financial Product

11.1 Monopolist Insurance Company

The monopolist has three possible options. One option is to offer only term insurance. If he chooses this option he chooses prices and commissions to maximize:

$$\max_{\{p_t,c_t\}} s(p_t - c_t - k)(\alpha - p_t) + c_t(p_t - c_t - k)(\alpha - p_t)$$

The first order condition with respect to price p_t is $(s+c_t)(p_t-c_t-k)(-1)+(s+c_t)(\alpha-p_t) = 0$, which simplifies to $p_t = \frac{\alpha+k+c_t}{2}$. The first order condition with respect to c_t is $(s + c_t)(p_t - \alpha) + (\alpha p_t-\alpha k-p_t^2-c_t\alpha+kp_t+c_tp_t) = 0$. Solving this system of equations yields the solution $c_t = \frac{\alpha-k-2s}{3}$ and $p_t = \frac{2\alpha+k-s}{3}$. Note that we need $s \leq \frac{\alpha-k}{2}$ to guarantee that commissions are non-negative (this condition also guarantees that prices are non-negative).[13]

The monopolist's second option is to offer both term and whole insurance. This option essentially constitutes price discrimination, where low prices and zero commissions are associated with term insurance for sophisticated consumers, and high prices and commissions are associated with whole insurance and unsophisticated consumers. The firm will pay zero commissions for the sale of term insurance; paying commissions does not increase demand but it does increase costs. The monopolist firm chooses the term insurance price p_t to maximize $s(p_t - k)(\alpha - p_t)$. The first order condition for p_t is $\alpha - 2p_t + k = 0$. The firm will choose to charge a price $\frac{\alpha+k}{2}$ for term insurance. Total profits from the sale of term insurance will equal $\frac{s(\alpha-k)^2}{4}$.

The firm will pay positive commissions for the sale of whole insurance, because demand is increasing in commissions. The firm maximizes the total profit function from selling whole insurance to unsophisticated customers: $c_w(p_w - k - c_w)(\alpha - p_w)$. The first order condition with respect to price is $c_w\alpha - 2p_wc_w + c_wk + c_w^2 = 0$. The first order condition with respect to the commission level c_w is $c_w(p\alpha - k\alpha - 2c\alpha - p^2 + pk + 2cp) = 0$. Solving these two first order conditions we find that the firm will set the whole insurance price (p_w) equal to $\frac{1}{3}(2\alpha + k)$ and will pay commissions

[13] Intuitively, this condition rules out a situation where there are a large number of sophisticated consumers and thus the firm would choose to pay negative commissions (i.e. force agents to pay the firm for selling to sophisticated consumers). If commissions were negative, agents would have no incentive to sell insurance in this model.

$\frac{1}{3}(\alpha - k)$.

We now show that when both products are offered and prices and commissions are chosen separately for each, that the price of term insurance will be higher than the price of whole insurance:

$$\frac{\alpha + k}{2} < \frac{1}{3}(2\alpha + k)$$

This expression can be simplified to $\alpha > k$, which must be true for their to be any positive demand for either insurance product. Thus, the monopolist will always choose higher prices for the whole insurance product versus the term insurance product. Intuitively, the monopolist pays higher commissions on whole insurance to attract consumers, and then passes on those commissions as higher prices. Total profits from the sale of whole insurance under the price discrimination strategy is $\frac{(\alpha-k)^3}{27}$. Total profits from the strategy of offering both term and whole products is $\frac{s(\alpha-k)^2}{4} + \frac{(\alpha-k)^3}{27}$.

The monopolist's third option is to offer only whole insurance. The sophisticated types never buy this, and the chosen p_w and c_w would be equivalent to those in Case 2. Thus, the firm can always add term insurance paying zero commissions and increase its profits. Thus, the monopolist firm will never offer only whole insurance.

We now show that the monopolist firm will always choose to offer both products as opposed to offering just term insurance. Intuitively, the monopolist can offer term and whole insurance products to price discriminate amongst the two types of consumers. In this case, price discrimination takes the form of offering higher commissions for sales of whole insurance to unsophisticated customers, and commissions equal to zero for sales of term insurance to sophisticated customers. We begin by showing that the profits from term consumers will always be lower when only term insurance is offered versus when both term insurance and whole insurance are offered.

The total profits from selling term insurance when both products are offered is $\frac{s(\alpha-k)^2}{4}$. The total profit from sophisticated consumers when only term insurance is offered is $s[\frac{1}{3}(2\alpha + k - s) - \frac{1}{3}(\alpha - k - 2s)][\alpha - \frac{1}{3}(2\alpha + k - s)]$. We wish to show that:

$$\frac{s(\alpha - k)^2}{4} > s[\frac{1}{3}(2\alpha + k - s) - k - \frac{1}{3}(\alpha - k - 2s)][\alpha - \frac{1}{3}(2\alpha + k - s)]$$

$$\frac{(\alpha - k)^2}{4} > \frac{1}{9}(\alpha - k + s)^2$$

Taking the square root of both sides we have $\frac{\alpha-k}{2} > \frac{1}{3}(\alpha - k + s)$ which simplifies to $\frac{\alpha-k}{2} \geq s$. Note that this is the same condition we needed to guarantee that commissions and prices are positive. Thus, the profits from selling to sophisticated consumers will be higher when both term and whole insurance products are offered, with different commissions and prices, then when term is sold to all customers.

We now show that the profits from unsophisticated consumers are also higher when the price discrimination strategy is followed. The profits on unsophisticated consumers under the price discrimination strategy are $\frac{(\alpha-k)^3}{27}$. The total profits from unsophisticated consumers when only term insurance is offered are $[\frac{1}{3}(\alpha - k - 2s) - \frac{1}{3}(\alpha - 2s)][\alpha - \frac{1}{3}(2\alpha - s)]$. Simplification shows that the price discrimination strategy yields higher profits as long as $3(\alpha - k) + 2s > 0$, which must be true as both $\alpha - k$ and s are non-negative.

Thus, we have shown that a monopolist firm will choose to sell both term and whole insurance, at different prices, to sophisticated and unsophisticated customers respectively. We have also shown that the monopolist will choose higher prices and commissions for whole insurance than for term insurance.

11.2 Two Competing Insurance Companies

The setup of this problem is defined in the Conclusion and Discussion section of the main text. We first solve for firm i's optimal behavior given firm j's possible behavior. Suppose firm j only offers whole insurance paying commission c_j and charging price p_j. In this case firm i will always choose to sell both whole and term insurance. If he chose to sell only one of these products, he could increase his profits by entering the term insurance market as a monopoly provider. Thus, there cannot be an equilibrium where both firms only sell either only term insurance or whole insurance.

Now suppose firm j offers both term and whole insurance. We show that there is one possible equilibrium in this case. Bertrand competition in the market for term insurance gives a Nash equilibrium $p_{i,t} = p_{j,t} = k$. In the term insurance market prices get driven down to marginal cost. Competition in the market for term insurance leads to lower prices, as sophisticated consumers are not persuaded by commissions in their decisions to purchase insurance products.

We now solve for a Nash equilibrium in the market for whole insurance. A price and commissions pair $(c_1^*, p_1^*, c_2^*, p_2^*)$ is a Nash equilibrium in the market for whole insurance if (c_i^*, p_i^*), for each firm i, solves the following problem (we suppress w subscript, but the commission and price term refer to whole insurance):

$$\max_{c_i,p_i}(c_i - bc_j^*)(p_i - k - c_i)(\alpha - p_i)$$

The first order condition with respect to p_i can be simplified to: $\frac{1}{2}(p_i - k + bc_j)$. The first order condition with respect to c_i an be simplified to $c_i^* = \frac{1}{2}(p_i - k + bc_j)$. Solving these two equations in two unknowns we find that firm i's optimal choices given firm j's choices are: $c_i^* = \frac{\alpha - k + 2bc_j}{3}$ and $p_i^* = \frac{1}{3}(2\alpha + k + bc_j)$. In a Nash equilibrium, firm j plays the same best responses given firm i's behavior, and thus we have: $c_j^* = \frac{\alpha - k + 2bc_i^*}{3}$ and $p_j^* = \frac{1}{3}(2\alpha + k + bc_i^*)$.

Solving this system of equations we find that the Nash equilibrium commissions are $c_i^* = c_j^* = \frac{\alpha - k}{3-2b}$, and the Nash equilibrium prices are $p_i^* = p_j^* = \frac{(2-b)\alpha + (1-b)k}{3-2b}$. Note that for commissions and prices to be positive we need $b \leq \frac{3}{2}$.

It is clear from the expression $c_i^* = c_j^* = \frac{\alpha - k}{3-2b}$ that the level of commissions paid will increase in the degree to which the insurance products compete with each other (b). We now show that prices are also increasing in b. We wish to show that the derivative of the expression for equilibrium prices with respect to b is greater than zero:

$$(3 - 2b)^{-1}(-\alpha - k) - (3 - 2b)^{-2}((2 - b)\alpha + (1 - b)k) > 0$$

This expression can be simplified to $\alpha > k$, which must be true for there to be any positive demand for the insurance product.

Figure 1: Fraction of Agents Recommending ULIP (Unit-Linked Life Insurance) Products

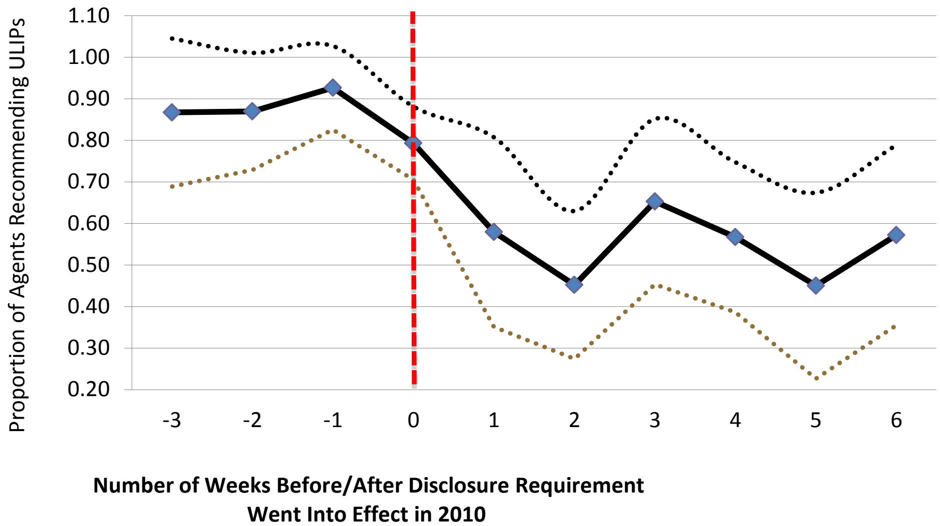


Figure 1 plots the fraction of agents each week recommending ULIP products to our mystery shoppers. The day the reform went into effect, July 1, 2010, is indicated by a red line.

Table 1: Audit Counts

					Number of	
				Audits	Auditors	Agents
Panel A: Competition (City #1)						
By need, belief, and source of beliefs (competition)						
	Need Term	Bias Term	Recommendation from other Agent	61	4	57
	Need Term	Bias Term	Recommendation from friend	65	4	61
	Need Term	Bias Whole	Recommendation from other Agent	57	5	53
	Need Term	Bias Whole	Recommendation from friend	75	4	70
	Need Whole	Bias Term	Recommendation from other Agent	77	4	70
	Need Whole	Bias Term	Recommendation from friend	77	4	71
	Need Whole	Bias Whole	Recommendation from other Agent	68	4	62
	Need Whole	Bias Whole	Recommendation from friend	77	5	73
	Total [a]			557		304
Panel B: Disclosure Experiment (City #2)						
By timing and whether auditor inquired about commission						
	Ask about commission		Pre-Disclosure Requirement	82	4	67
	Ask about commission		Post-Disclosure Requirement	61	3	58
	Do not ask about commission		Pre-Disclosure Requirement	67	4	54
	Do not ask about commission		Post-Disclosure Requirement	47	3	40
	Total [a]			257		198
Panel C: Sophistication Experiment (City #2)						
By level of sophistication						
	Low level of sophistication			114	7	110
	High level of sophistication			103	6	103
	Total [a]			217		209

Table 1 contains audit counts from our three experiments, disaggregated by treatment combinations. The first column provides the total number of audits for each treatment combination, the second column provides the total number of auditors involved for each treatment combination, and the final column provides the number of distinct agents visited for each treatment combination. **Quality of Advice** refers to the experiment where we varied the auditor's needs, beliefs, and the source of their beliefs (competing agent or friend). **Disclosure** refers to the experiment where we varied whether the auditor made a disclosure inquiry, both before and after the mandatory disclosure law, to test the law's effect on agent behavior. **Sophistication** refers to the experiment where we varied the auditors' expressed financial sophistication.
a) Since agents may have been visited by more than one auditor, the number of agents visited is less than the total number of audits.

Table 2: Summary Statistics From Audits

	Quality of Advice	Disclosure	Sophistication
LIC Underwriter	0.73	0.50	0.69
	(0.44)	(0.50)	(0.46)
Audit Location			
Agent Home	0.18	0.14	0.12
	(0.39)	(0.34)	(0.33)
Agent Office	0.12	0.72	0.55
	(0.33)	(0.45)	(0.50)
Auditor Home	0.01	0.06	0.03
	(0.09)	(0.23)	(0.18)
Auditor Office	0.01	0.02	0.18
	(0.12)	(0.12)	(0.39)
Other Venue	0.68	0.07	0.11
	(0.47)	(0.26)	(0.31)
Audit Duration	37.13	37.58	33.22
	(10.22)	(15.88)	(12.58)
Recommendations:			
Only Whole	0.81	0.25	0.75
	(0.39)	(0.43)	(0.43)
Only Term	0.03	0.01	0.14
	(0.17)	(0.09)	(0.35)
Only ULIP	0.08	0.71	0.16
	(0.27)	(0.45)	(0.37)
Any Whole	0.90	0.27	0.82
	(0.30)	(0.44)	(0.38)
Any Term	0.13	0.01	0.22
	(0.33)	(0.11)	(0.42)
Any ULIP	0.10	0.72	0.18
	(0.30)	(0.45)	(0.38)
Observations	557	257	217

Table 2 presents summary statistics from our three experiments. **Quality of Advice** refers to the experiment where we varied the auditor's needs (savings vs. risk), beliefs (whole vs. term) and the source of their beliefs (competing agent or friend). **Disclosure** refers to the experiment where we varied whether the auditor made a disclosure inquiry, both before and after the mandatory disclosure law, to test the law's effect on agent behavior. **Sophistication** refers to the experiment where we varied the auditors' expressed financial sophistication. Note that "LIC" refers to the Life Insurance Corporation of India, a government-owned insurance company that has the largest share of insurers in the country.

Table 3: Tests of Randomization

	Quality of Advice						Disclosure		Sophistication	
	Bias Treatment		*Suitability Treatment*		*Competition Treatment*					
	Term	Whole	Term	Whole	Friend	Agent	Inquiry	No Inquiry	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Government Underwriter	0.82	0.79	0.79	0.82	0.80	0.82	0.50	0.55	0.72	0.71
LIC Underwriter	0.74	0.73	0.71	0.76	0.73	0.74	0.48	0.52	0.68	0.70
Agent is Male	0.84	0.84	0.86	0.83	0.84	0.84	0.88	0.93	0.89	0.93
Agent Dress (1-simple to 5-sophisticated)	4.07	4.03	4.05	4.05	4.11	3.98 **	3.60	3.53		
Physical Quality of Office (1-low to 5-high)	4.18	4.19	4.13	4.23	4.19	4.18	3.57	3.69		
Audit Location										
Agent Home	0.19	0.18	0.17	0.19	0.16	0.21	0.17	0.10 *	0.11	0.14
Agent Office	0.13	0.11	0.12	0.12	0.13	0.11	0.69	0.75	0.53	0.58
Auditor Home	0.00	0.02 **	0.01	0.01	0.01	0.01	0.06	0.05	0.04	0.03
Auditor Office	0.01	0.01	0.01	0.02	0.02	0.01	0.00	0.04 **	0.18	0.18
Other Venue	0.67	0.68	0.69	0.66	0.69	0.66	0.08	0.06	0.15	0.07 *
Audits	280	277	258	299	294	263	143	114	114	103

*** p<0.01, ** p<0.05, * p<0.1

Table 3 presents summary statistics from our three experiments disaggregated by treatment. They are used to perform randomization checks, univariate regressions (with robust standard errors) of the treatment on each independent variable. Significant differences are denoted by asterisks. **Quality of Advice** refers to the experiment where we varied the auditor's needs (*suitability*), beliefs (*bias)* , and the source of their beliefs, competing agent or friend (*competition*). As mentioned in Table 1, **Disclosure** refers to the experiment where we varied whether the auditor made a disclosure inquiry, both before and after the mandatory disclosure law, to test the law's effect on agent behavior. **Sophistication** refers to the experiment where we varied the auditors' expressed financial sophistication. Note that "Government Underwriter" includes LIC, State Bank of India (SBI), United Trust of India (UTI), and the Industrial Development Bank of India (IDBI).

Table 4: Do Agents Cater to Customers Beliefs or Respond to Customer Needs?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Any Term		Only Term		Ln(Coverage)		Ln(Premium)	
Bias=Term	0.096 ***	0.105 ***	0.019 *	0.022 **	0.131 **	0.125 **	-0.013	-0.019
	(0.029)	(0.028)	(0.011)	(0.011)	(0.060)	(0.058)	(0.050)	(0.045)
Need=Term	0.116 ***	0.126 ***	0.015	0.019 *	0.170 **	0.177 **	0.002	-0.005
	(0.032)	(0.031)	(0.011)	(0.011)	(0.075)	(0.075)	(0.051)	(0.048)
(Bias=Term)*(Need=Term)	0.021	0.006	0.053 *	0.049 *	0.055	0.051	0.043	0.038
	(0.057)	(0.055)	(0.030)	(0.028)	(0.128)	(0.127)	(0.065)	(0.060)
Government Underwriter		-0.121 ***		-0.017		-0.222 **		-0.039
		(0.039)		(0.021)		(0.094)		(0.050)
Audit Location								
Agent Home		0.012		-0.021		-0.069		-0.113
		(0.047)		(0.027)		(0.105)		(0.071)
Auditor Home		-0.132		-0.018		-0.499 *		-0.673
		(0.105)		(0.026)		(0.282)		(0.517)
Auditor Office		0.329 **		0.206		0.315		-0.554 ***
		(0.155)		(0.140)		(0.250)		(0.212)
Other Venue		-0.018		-0.018		-0.081		-0.122 **
		(0.041)		(0.022)		(0.089)		(0.052)
Auditor Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	557	557	557	557	538	538	540	540

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table 4 reports regressions where the dependent variables are the (exclusive) presence of term insurance in the agent's recommendation in columns (1) - (4). The dependent variable is the logarithm of risk coverage recommended in Columns (5) and (6) and of premium amount recommended in Columns (7) and (8) . The main independent variables are whether the auditor expressed a bias for term, whether the auditor expressed a genuine need for term, and an interaction between these two variables. The bias for term is expressed through an auditor's explicit stated preference for term, while a need for term is expressed by the auditor mentioning his/her desire to cover risk at an affordable cost (as opposed to the need for whole, which is expressed by wanting to save and invest and not feeling self-disciplined enough to do it on one's own). Dummy variables for venue location (agent office is the omitted category), whether the agent was selling insurance from a government underwriter, and auditor fixed effects are also included in columns (2), (4), (6), and (8). The number of observations in Columns (5) and (6) are less than those in (1) and (2) because agents did not recommend specific levels of coverage in 19 audits.

Table 5: Does the Presence of Competition Improve Agent Advice?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	Recommended Any Term				Recommended Only Term			
Bias=Term	0.105 ***	0.106 ***	0.091 **	0.090 **	0.043 ***	0.045 ***	0.026	0.027
	(0.028)	(0.027)	(0.041)	(0.038)	(0.014)	(0.014)	(0.018)	(0.017)
Need=Term	0.127 ***	0.130 ***	0.067 *	0.068 *	0.042 ***	0.044 ***	0.027	0.029
	(0.028)	(0.027)	(0.038)	(0.035)	(0.015)	(0.014)	(0.019)	(0.020)
Competition	0.024	0.033	-0.011	-0.008	0.010	0.012	0.000	0.001
	(0.028)	(0.027)	(0.023)	(0.024)	(0.014)	(0.014)		(0.006)
(Bias=Term)*Competition			0.011	0.030			-0.013	-0.008
			(0.057)	(0.056)			(0.022)	(0.022)
(Need=Term)*Competition			0.111 *	0.135 **			-0.027	-0.023
			(0.067)	(0.067)			(0.019)	(0.021)
(Bias=Term)*(Need=Term)			0.062	0.075			-0.006	-0.004
			(0.076)	(0.071)			(0.037)	(0.036)
(Bias=Term)*(Need=Term)*Competition			-0.095	-0.158			0.125 **	0.113 **
			(0.115)	(0.113)			(0.059)	(0.055)
Government Underwriter		-0.122 ***		-0.128 ***		-0.020		-0.013
		(0.039)		(0.039)		(0.021)		(0.020)
Audit Location								
Agent Home		0.009		0.002		-0.022		-0.019
		(0.047)		(0.047)		(0.028)		(0.027)
Auditor Home		-0.138		-0.140		-0.018		-0.015
		(0.108)		(0.112)		(0.029)		(0.025)
Auditor Office		0.331 **		0.332 **		0.207		0.202
		(0.156)		(0.158)		(0.139)		(0.137)
Other Venue		-0.020		-0.028		-0.022		-0.016
		(0.040)		(0.040)		(0.023)		(0.022)
Auditor Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	557	557	557	557	557	557	557	557

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table 5 reports regressions where the dependent variables are the (exclusive) presence of term insurance in the agent's recommendation. The main independent variable is competition (the main effect and the interactions with bias and need), which is signaled in an audit in two ways: first, by the auditor mentioning meeting with other providers and second, by the auditor stating a preference based on advice from another agent. Dummy variables for venue location (agent office is the omitted category), whether the agent was selling insurance from a government underwriter, and auditor fixed effects are also included in even-numbered columns

Table 6: Disclosure Experiment Summary Statistics

	Overall	Pre-Regulation	Post-Regulation	Difference
LIC Underwriter	0.50	0.44	0.58	0.15 ***
	(0.50)	(0.50)	(0.50)	(0.06)
Audit Location				
Agent Home	0.14	0.09	0.19	0.10 ***
	(0.34)	(0.29)	(0.40)	(0.05)
Agent Office	0.72	0.75	0.67	-0.09 *
	(0.45)	(0.43)	(0.47)	(0.06)
Auditor Home	0.06	0.07	0.04	-0.04
	(0.23)	(0.26)	(0.19)	(0.03)
Auditor Office	0.02	0.02	0.01	-0.01
	(0.12)	(0.14)	(0.10)	(0.01)
Other Venue	0.07	0.06	0.09	0.03 **
	(0.26)	(0.24)	(0.29)	(0.03)
Audit Duration	37.58	36.14	39.56	3.41 ***
	(15.88)	(14.33)	(17.67)	(2.07)
Recommendations:				
Only Whole	0.25	0.15	0.39	0.24 ***
	(0.43)	(0.36)	(0.49)	(0.06)
Only Term	0.01	0.01	0.00	-0.01
	(0.09)	(0.12)	(0.00)	(0.01)
Only ULIP	0.71	0.83	0.55	-0.29 ***
	(0.45)	(0.37)	(0.50)	(0.06)
Any Whole	0.27	0.15	0.43	0.27 ***
	(0.44)	(0.36)	(0.50)	(0.06)
Any Term	0.01	0.01	0.01	0.00
	(0.11)	(0.12)	(0.10)	(0.01)
Any ULIP	0.72	0.83	0.56	-0.28 ***
	(0.45)	(0.37)	(0.50)	(0.06)
Observations	257	149	108	

Table 6 presents summary statistics from the disclosure experiment disaggregated by timing. They are used to perform a balance check, univariate regressions (with robust standard errors) of the treatment on each independent variable. Significant differences are denoted by asterisks.

Table 7: Disclosure Regulations and Product Recommendations

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	ULIP Recommendation				Ln(Risk Cover)	Ln(Premium)
Sample:	All	All	Government Underwriter	Private Underwriter	All	All
Post Disclosure	-0.25 ***	-0.19 **	-0.30 **	-0.07	0.15	0.03
	(0.09)	(0.08)	(0.12)	(0.08)	(0.13)	(0.07)
Disclosure Inquiry	0.05	0.02	0.07	0.00	0.02	0.00
	(0.06)	(0.06)	(0.13)	(0.05)	(0.11)	(0.06)
Post * (Disclosure Inquiry)	-0.06	-0.02	-0.06	0.07	0.02	-0.01
	(0.12)	(0.10)	(0.17)	(0.11)	(0.17)	(0.09)
Government Underwriter		-0.42 ***			0.29 ***	0.01
		(0.05)			(0.10)	(0.05)
Audit Location						
Agent Home		-0.01	-0.02	0.07 *	0.06	0.04
		(0.08)	(0.10)	(0.04)	(0.12)	(0.08)
Auditor Home		-0.02	-0.25	0.03	0.65 *	0.24
		(0.11)	(0.16)	(0.05)	(0.37)	(0.21)
Auditor Office		0.18	0.65 ***	0.05	0.62 ***	0.30 *
		(0.13)	(0.12)	(0.05)	(0.19)	(0.17)
Other Venue		0.06	0.04	0.06 *	0.07	-0.01
		(0.09)	(0.13)	(0.04)	(0.14)	(0.07)
Auditor Fixed Effects	No	Yes	Yes	Yes	Yes	Yes
Observations	257	257	134	134	214	214

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table 7 reports regressions where the dependent variable is a binary equal to 1 if a ULIP product is recommended for columns (1) -(4). The dependent variable in columns (5) and (6) are, respectively, the logarithm of the risk coverage and premium of the recommended policy. The ULIP product is the product where disclosure of commissions was made mandatory on July 1, 2010. The main independent variables are whether or not the audit occurred after the commissions disclosure law came into effect (*post disclosure*), whether or not the auditor made an explicit commission *disclosure inquiry* , and an interaction between these two variables. Dummy variables for venue location (agent office is omitted), whether the agent is selling insurance from a government-owned insurer, and auditor fixed-effects are included in even-numbered columns.

Table 8: Effect of Sophistication on Quality of Advice

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Recommended Any Term		Recommended Only Term		Ln(Coverage)		Ln(Premium)	
Sophisticated	0.10 *	0.10 *	0.02	0.03	0.22 *	0.21 *	-0.03	-0.06
	(0.06)	(0.06)	(0.05)	(0.05)	(0.12)	(0.12)	(0.09)	(0.10)
Government Underwriter		-0.08		-0.09		-0.25		0.05
		(0.07)		(0.06)		(0.16)		(0.10)
Audit Location								
Agent Home		0.10		-0.01		0.21		-0.21
		(0.10)		(0.06)		(0.18)		(0.18)
Auditor Home		0.02		-0.11 **		0.32		0.03
		(0.14)		(0.05)		(0.29)		(0.14)
Auditor Office		0.13		0.13		0.20		-0.17
		(0.09)		(0.09)		(0.16)		(0.13)
Other Venue		-0.01		0.06		-0.17		-0.28
		(0.09)		(0.09)		(0.24)		(0.19)
Auditor Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	217	217	217	217	209	209	209	209

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table 8 reports regressions where the dependent variables are the (exclusive) presence of term insurance in the agent's recommendation. The main independent variable is whether or not the audit is part of the "sophisticated" treatment group. Sophistication was signaled to the agent by a script in which auditors mentioned how they had been shopping around and were aware of the different types of policies (such as ULIPs, term, etc.) In unsophisticated audits, auditors acknowledged that life insurance was complex but admitted to knowing very little about the types of policies. Dummy variables for auditor identity, venue location, and whether the government purveyed/underwrote the insurance policy are also included in the even-numbered columns.

Appendix Table A1: Comparison of Whole vs. Term Plus Savings

Panel A: Life Insurance Products

	Whole Life Insurance	**Term Life Insurance**
	Government and private insurance companies	Government and private insurance companies
Specific Plan Example	The Whole Life Plan (#2)	Amulya Jeevan (#190)
Firm Offering	Life Insurance Corporation of India (LIC)	Life Insurance Corporation of India (LIC)
Coverage Amount	2,500,000	4,000,000
Premium for 25 year old male	Rs. 55,116	Rs. 11,996
Years client pays	47	35
Years policy pays out	until death of client, no matter the age	35
Historic bonus percentage	7% (non-compounded)	n/a

Panel B: Savings Products

	Promised interest rate
Bank Fixed Deposit	8.75%
Government Provident Fund	8%

Panel C: Comparison of Whole Life vs. Term and Fixed Deposit Savings

		Whole Life Insurance	Term + Savings
Products Purchased		Rs. 2.5m in life insurance at Rs. 55,166 per year for 47 years	Rs. 4m of term life insurance for 35 years, at annual payments of 11,996 per year for 35 years. Savings deposit of Rs 55,166-11,996=43,170 per year for 35 years, earning 8.75% Savings deposit of Rs. 55,166 per year from years 36-47, earning 8.75%
Value Upon Death (Rs.)		Whole Payout	Term Payout (if any) + Savings
Dying at age:	25	2,675,000	4,046,893
	35	4,425,000	4,812,490
	45	6,175,000	6,583,792
	55	7,925,000	10,779,449
	65	9,675,000	16,584,940
	75	11,425,000	39,271,154
	85	13,175,000	91,310,405

Appendix Table A2 Text of Treatments

Quality of Advice Experiment		
Bias treatment	Bias towards term	Bias towards whole
Text of statement	"I have heard from [source] that term insurance is a really good product."	"I have heard from [source] that whole insurance is a really good product."
Needs treatment	Need term	Need whole
Text of Statement	"I am worried that if I die early, my wife and kids will not be able to live comfortably or meet our financial obligations. I want to cover that risk at an affordable cost."	"I want to save and invest money for the future, and I also want to make sure my wife and children will be taken care of if I die. I do not have the discipline to save on my own."
Competition Treatment	High Competition	Low Competition
Competition	"I have already met with some providers, but would like to learn more about the specific products your firm offers so I can make a comparison" [source] in bias statement is "another agent"	"What are the different products that you offer?" [source] in bias statement is "friends"
Disclosure Experiment		
Knowledge treatment	Knowledge of Commissions	No Knowledge
	"Can you give me more information about the commission charges I'll be paying?"	No mention of commission charges
Sophistication Experiment		
Sophistication treatment	Sophisticated	Unsophisticated
	"In the past, I have spent time shopping for the policies, and am perhaps surprisingly somewhat familiar with the different types of policies: ULIPs, term, whole life insurance. However, I am less familiar with the specific policies that your firm offers, so I was hoping you can walk me through them and recommend a policy specific for my situation."	"I am aware of the complexities of Life Insurance Products and I don't understand them very much; however I am interested in purchasing a policy. Would you help me with this?"

Interbank network and bank bailouts: Insurance mechanism for non-insured creditors?[☆]

Tim Eisert[a], Christian Eufinger[b,*]

[a]*Goethe University Frankfurt, Department of Finance and the Ph.D. Program in Law and Economics of Money and Finance, Grueneburgplatz 1 60323 Frankfurt, Germany*
[b]*Goethe University Frankfurt, Department of Economics and Business Administration and the Ph.D. Program in Law and Economics of Money and Finance, Grueneburgplatz 1 60323 Frankfurt, Germany*

April 10, 2012

Abstract

This paper presents a theory that explains why it is beneficial for banks to engage in circular lending activities on the interbank market. Using a simple network structure, it shows that if there is a non-zero bailout probability, banks can significantly increase the expected repayment of uninsured creditors by entering into cyclical liabilities on the interbank market before investing in loan portfolios. Therefore, banks are better able to attract funds from uninsured creditors. Our results show that implicit government guarantees incentivize banks to have large interbank exposures, to be highly interconnected, and to invest in highly correlated, risky portfolios. This can serve as an explanation for the observed high interconnectedness between banks and their investment behavior in the run-up to the subprime mortgage crisis.

Keywords: bailout, cycle flows, cyclical liabilities, interbank network, leverage
JEL: G01, G21, G28, L14

[☆]The authors appreciate helpful comments from Christian Hirsch, Jan Pieter Krahnen, Bjoern Richter, Uwe Walz, Manuel Wiegand and workshop participants at Augsburg University. Furthermore, the authors gratefully acknowledge financial support from the Ph.D. Program in Law and Economics of Money and Finance and the Stiftung Geld und Waehrung

[*]Corresponding author
Email addresses: eisert@finance.uni-frankfurt.de (Tim Eisert), christian.eufinger@hof.uni-frankfurt.de (Christian Eufinger)

1. Introduction

The 2008-2009 financial crisis has prompted many questions about the resilience of the interbank market. Strong growth in the size and density of the interbank network has made concerns such as "too big to fail" and "too interconnected to fail" widespread.[1] However, there is only scarce knowledge of why banks enter into such a high degree of connectivity in the first place, especially since these connections often include cyclical liabilities that could potentially be netted out.

The goal of the present paper is to fill this gap in the literature. We develop a model to show that it can be beneficial for banks to be highly interconnected and even to enter into cyclical liabilities. We claim that this interbank network serves as an insurance mechanism for a bank's creditors if they are not already covered by a deposit insurance (such as, e.g., the FDIC). If a bank failure occurs and there is a non-zero probability that banks will be bailed out by the government, then connections to other banks (e.g., exposures arising from credit default swap (CDS) contracts, bonds, and interbank lending), particularly cyclical liabilities, can actually increase the expected repayment of uninsured creditors. This can be best understood by considering the option pricing approach to explicit and implicit loan guarantees. Merton (1977) shows that the value of these guarantees is akin to a put option. In the case of a bank failure such circular lending activities increase the benefit from this option. This incentivizes banks to be highly interconnected, which implies that many cyclical liabilities occur.[2] We also show that, due to the high interconnectedness, banks are incentivized to invest in correlated assets, thereby increasing the likelihood of a joint default. Banks' risk-shifting incentives increase with their interbank exposure and interconnectedness as well. Therefore, our model helps explain why banks invested in risky correlated investments (e.g., US subprime loans) in the run-up to the financial crisis.

Due to the high interconnectedness and resulting cycle flows (i.e. cyclical liabilities), banks are lending to and borrowing from each other large amounts leading to an increased leverage of each bank, without necessarily altering the aggregate relationship between the banking sector and the ultimate creditors or depositors (Shin (2009); Adrian and Shin (2011)), and high systemic risk. However, systemic risk is not only arising from the interconnectedness of banks but can also result from a "joint failure risk arising from the correlation of returns on the asset side of bank balance sheets" (Acharya (2009, p. 225)). We show that the mechanism presented in this paper provides an incentive for banks to increase both types of systemic risk. Moreover, we investigate the interaction between these two sources of systemic risk and show that they cannot be considered individually.

The rest of the paper is organized as follows. Section 2 provides an overview of the related literature. Using a simple example, Section 3 presents our main argument, that due to a positive bailout probability, cyclical liabilities lead to higher expected repayments for uninsured creditors. Section 4 develops our main model and provides implications for

[1] See Minoiu and Reyes (2011), who explore the properties of the global banking network during 1978-2010 and assess its dynamics during financial crises.

[2] Among others, Takács (1988) proves that the expected number of cycle flows increases with the density of the network

the investment behavior of banks. Section 5 shows that interbank connections can lead to risk shifting. Section 6 provides two extensions to our main model. First, we extend our model to a three-region economy and compare different network structures. Second, we introduce risk aversion and show that our main results are not affected. Section 7 concludes.

2. Related Literature

Several empirical papers find that the global banking network has a very high density and a high degree of concentration. Using locational statistics from the Bank for International Settlements on exchange-rate adjusted changes in cross-border bank claims, Minoiu and Reyes (2011) analyze the global banking network and find that, besides a high network density, there exists a positive correlation between network density and the circularity of liabilities (measured by the network's clustering coefficient). Kubelec and Sá (2010) use a cross-country panel dataset of 18 countries to investigate the development of the global financial network over time. They show that the interconnectivity of the global financial network has increased significantly over the past two decades. In line with our results, they find that the global financial network is characterized by a large number of small links and a small number of large links and that the network has become more clustered.

Using micro-level data from Loan Analytics, Hale (2011) shows that in the years 2002-2006 (i.e., before the crisis) the global banking network was characterized by an increasing number of banks, an increasing number of connections between banks, and an increasing number of countries in which banks participate in the global banking network. Moreover, the author finds that this network expansion was mainly driven by a higher interconnectedness of existing banks rather than the entrance of new banks into the global network. This supports our idea that banks were highly connected across countries in the run-up to the financial crisis. Similar evidence can be found for national interbank markets (Wells (2004); Mueller (2006); May et al. (2008)).

Furthermore, there is also a very high interconnectedness in other interbank markets besides the traditional interbank lending market. For example, BIS 2011 shows that banks also have very high cross-exposures due to derivative contracts (mainly CDSs), since banks that sell CDSs in turn also purchase them to hedge their risk. This reduces their net exposure but increases the amount of cyclical liabilities substantially. The extent of these cyclical liabilities can easily be seen from exposure data provided by the International Swaps and Derivatives Association.[3] Comparing the net and gross notional amounts of outstanding CDSs on European sovereign debt shows that the gross is often more than 10 times larger than the net amount. These hedging activities, which in turn entail enormous levels of gross exposure, build up huge counterparty risks. Hence, as the default of AIG demonstrates, as soon as the chain of bilateral netting breaks down, gross exposure becomes net exposure.

[3] See `http://www.isdacdsmarketplace.com/exposures_and_activity/top_10_cds_positions`

Our paper is also related to several strands of the theoretical literature. First, it adds to the literature on liquidity and interbank markets. Pioneering work in this area has been accomplished by Bhattacharya and Gale (1987), who show that banks can co-insure each other through an interbank market against liquidity shocks as long as these shocks are not perfectly correlated. This theme has been taken on by many other papers. For example, Freixas and Holthausen (2005) analyze the scope for international interbank market integration when cross-border information about banks is less precise than home country information. Here, banks can cope with these shocks by investing in a storage technology or can use the interbank market to channel liquidity. Allen et al. (2009) show that the interbank market is characterized by excessive price volatility if there is a lack of opportunities for banks to hedge aggregate and idiosyncratic liquidity shocks. A recent paper by Castiglionesi et al. (2011) shows that there exists a negative relation between a bank's activity in the interbank market and its bank capital because it is optimal for banks to postpone payouts to investors when they are hit by liquidity shocks that cannot be co-insured in the interbank market, in which case interbank activity is low.

In addition, our paper is related to the literature on financial contagion. In Section 6.2 we incorporate our modeling idea into a model setup originally proposed by Allen and Gale (2000). This framework is used by many papers (e.g. Brusco and Castiglionesi (2007), Leitner (2005) and Freixas et al. (2000)). Therefore, we show that the results we find in our main model under the assumption of risk neutrality remain valid when incorporated into a setup of the type proposed by Allen and Gale (2000) and Brusco and Castiglionesi (2007). Similar to these papers, we see the interbank market as an insurance mechanism. In these previous studies, the interbank market is supposed to insure banks against liquidity shocks that result from depositors already withdrawing their money in an intermediate period. In our setting an additional insurance mechanism results from the fact that if a bank is connected to other banks, the expected repayment to uninsured creditors increases in case the bank defaults. This is because even if this specific bank is not bailed out, there nevertheless exists a positive probability that the next bank in the chain will be. If markets have reached a high network density with high capital flows, implying that many and large cycle flows exist, then ultimately the failing bank will receive funds from banks it is connected to if they are bailed out.

Similar to our model, Castiglionesi and Navaro (2010) use a banking network with core and periphery banks (uninsured creditors) that differ with respect to their interconnectedness (and investment risk) and establish conditions under which fragility is an optimal feature of financial networks. Cukierman and Izhakian (2011) develop a microfounded general equilibrium model of the financial system composed of ultimate borrowers, ultimate lenders, and financial intermediaries and investigate the impact of bailout uncertainty on leverage, interest rates, the volume of defaults, and the real economy. Our approach differs from theirs in that we start by assuming a fixed bailout probability and investigate how it affects the expected repayment uninsured creditors receive from a bank (and hence the interest rate the bank is able to pay uninsured creditors) under different network structures. David and Lehar (2011) also present a mechanism that incentivizes banks to create cyclical liabilities. In the case where banks have perfect information about the interbank network and the liabilities of all banks, cyclical liabilities can act as a commitment device to facilitate mutual private sector bailouts. In contrast, we investigate the effect of possible government bailouts on the incentives of banks to create such liabilities.

Lastly, our paper relates to the literature on bank bailouts. Acharya and Yorulmazer (2007) focus on whether governments have an incentive to bail out banks ex post if they engaged in herding behavior ex ante. Diamond and Rajan (2002) show that bailouts alter available liquidity in the economy and distinguish between well-targeted bailouts (which can be beneficial) and poorly targeted ones that can lead to a systemic crisis. Gorton and Huang (2004) argue that there is a potential role for governments to provide liquidity through, for example, bank bailouts to reduce the problem of agents hoarding liquidity inefficiently.

3. Main Idea

To illustrate our main idea, we use a very simple framework similar to that of Rotemberg (2011). We assume that the interbank market consists of a few core banks and some uninsured creditors (e.g., mutual funds, bondholders, regional banks). One of the core banks has an investment project that costs one unit in the first period and generates a return $R > 1$ in the second period with probability λ and a return of zero otherwise. The only source of capital to fund this project is to borrow from the uninsured creditors. In return for the initial funding, the bank must repay R_D to its uninsured creditor. All parties are risk-neutral.

We develop the intuition of our model in two steps. First, we discuss a situation without network connections to other core banks. At $t = 0$ the bank (B_A) borrows one unit from the uninsured creditor (C) and invests in a project (P). In the second period, the cash flow from the project is realized. If the project is successful, the bank receives an amount R and is able to fully repay its uninsured creditor. If the project fails and the bank is not bailed out, the uninsured creditor receives no repayment. Conversely, if the government bails out the bank (i.e., takes over the bank and settles all its liabilities), the creditor again receives his full repayment (see Figure 1).

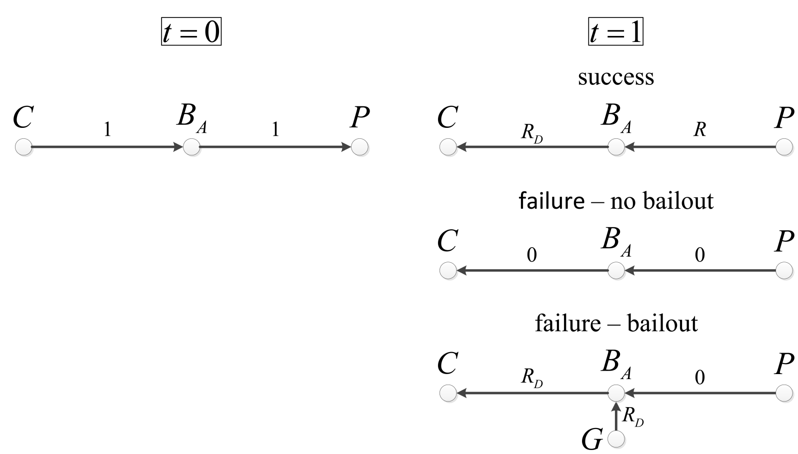


Figure 1: Capital flows without interbank market and zero bailout probability

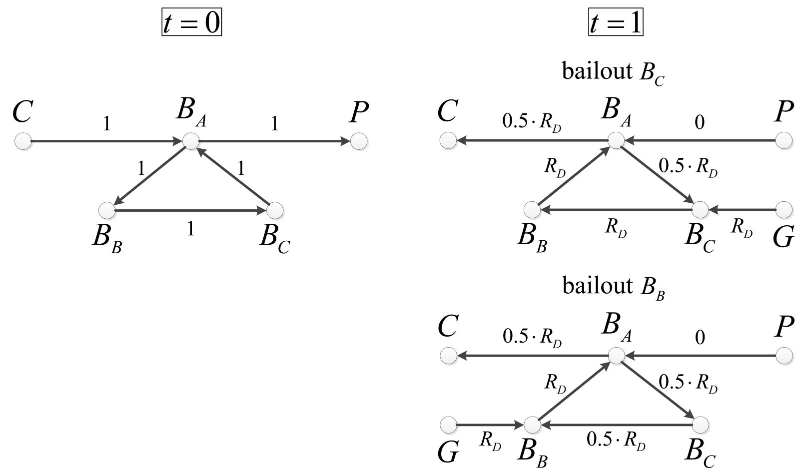


Figure 2: Capital flows with interbank market and bailout

In a second step we allow the bank to establish an interbank network at $t = 0$ by lending one unit of capital that, for example B_A receives from its uninsured creditor in a circular way. To be precise, bank B_A lends one unit of capital to bank B_B, which in turn lends it to bank B_C, from which the capital flows back to B_A and is then invested into the project. For now, we assume that banks B_B and B_C do not have any other investments. We relax this assumption in the next sections. Moreover, for ease of illustration, we assume that the gross interest rate on the interbank market is R_D as well. If the project is successful, B_A receives the project return R and uses it to settle its liabilities with B_C.[4] After receiving the payment from B_B it repays its uninsured creditor. If the project fails, bank B_A defaults since it cannot repay its creditors. If the government steps in and bails out bank B_A, both the uninsured creditor of B_A and bank B_C receive their full repayment R_D, implying that all claims are settled in this case. If the government refuses to bail out B_A, B_C defaults as well. Now the government (not necessarily the same one as in the case of B_A, since B_C may be established in another country) must decide whether to bail out B_C. If it does, it takes over B_C and settles its liabilities. Therefore B_B receives R_D from B_C and hence B_B can pay back its debt to B_A. However, B_A has total liabilities of $2R_D$ and is therefore still unable to meet all its obligations. Consequently, the funds B_A received from B_B must be divided among the creditors of B_A, that is, the uninsured creditor of B_A, on the one hand, and B_C, on the other hand.

The common procedure in bankruptcy proceedings is for debt to be paid back on a pro rata basis once a default occurs. Therefore, each creditor receives $\frac{1}{2}R_D$. Since the government takes over B_C, it receives this amount. However, it has to pay R_D to bail out the bank and hence records a loss of $\frac{1}{2}R_D$. The case where B_C is not bailed out but

[4]Throughout the paper we assume that, as soon as there exists a clearing payment vector, the banks use this vector to settle all liabilities in the network. This no longer holds if the sequence of payments is chosen in a less sophisticated manner. In this case, banks can still default, even though there is enough liquidity in the system to settle all claims. However, this would only reinforce our mechanism, since it would increase the value of the government's implicit guarantee.

B_B is can be described analogously. The corresponding cash flows (in case the project fails and one of the other banks is bailed out) are presented in Figure 2. Hence, in case there is a positive probability of a government bailout if a bank defaults, the bank can considerably increase the expected repayment of its uninsured creditor by first channeling funds through the interbank market and only lending them out to the ultimate borrower afterwards. This is because the uninsured creditor receives a positive repayment as soon as at least one of the banks is bailed out.

If the bank has the bargaining power, creditors will demand a lower interest rate (risk premium) given the existence of an interbank network (the participation constraint of uninsured creditors is already binding for lower values of R_D) which considerably reduces the bank's borrowing cost. This in turn leads to higher profits for the bank, which can help explain the comparatively high return-on-equity ratios of banks. If, on the other hand, the uninsured creditor has the bargaining power, he will increase his expected repayment by increasing R_D until the participation constraint of the owners of the bank is just binding. Furthermore, creditors will only deposit money in banks that are part of an interbank network, since the expected repayment in this case is higher than when the bank is not connected to others via an interbank market.

Note that the described mechanism can be reinforced by channeling more than one unit of capital through the interbank market. For example, this can be realized by repeating the circular lending procedure a couple of times (e.g., K repetitions lead to an interbank network exposure of K). This increases the expected repayment to the uninsured creditor even further. Moreover, it is easy to see that the expected repayment to the uninsured creditor can also be increased by increasing the number of banks in the interbank network.

4. The Main Model

Having described the main mechanism we now formalize our idea and develop our main model. We consider an economy that consists of two dates $t = 0$ and $t = 1$ and two different regions, A and B (which can be interpreted as, e.g., two different countries). Each region is comprised of a continuum of identical banks. We assume that, due to competition, all banks adopt the same behavior and can thus be described by a representative bank (protected by limited liability). The representative bank in region A (B) is denoted by B_A (B_B). In line with Allen and Gale (2000), these banks can establish an interbank market (network) by exchanging an arbitrary amount of interbank deposits K at $t = 0$ in return for a payment of KR_D at $t = 1$. This is a simplified approach to model the cycle flows that otherwise result from a high degree of market density.[5]

Furthermore, we assume that there exists an uninsured creditor (endowed with c units of capital at date $t = 0$) and one investor who provides equity financing to the bank in each region. Creditors are denoted C_A and C_B in regions A and B, respectively. This

[5]Note, however, that there exists some anecdotal evidence from German Landesbanks that even this kind of bilateral circular lending exists on the interbank market. For example, a 2006 report by Fitch describes that after the abolition of the explicit state guarantee, Landesbanks bought bonds from each other in large amounts, thereby creating "cyclical liabilities" bilaterally.

contract takes the form of a standard debt contract; that is, it cannot be made contingent on either the realization of the investment or the realization of the state of nature. Lastly, we consider a government in each region. All actors are risk neutral.

We consider a situation where each bank has access to two investment possibilities in two different industries (denoted 1 and 2), as in Acharya and Yorulmazer (2007). Both investments need an initial amount of capital o which is normalized to one. One can think of these investment opportunities as portfolios of loans to firms in one of the two industries. More precisely, bank B_A (B_B) can lend to firms in industry A_1 or A_2 (B_1 and B_2). If in equilibrium banks decide to lend to firms in the same industry, that is, they either lend to A_1 and B_1 or to A_2 and B_2, then the returns of their loan portfolios are assumed to be perfectly correlated ($\rho = 1$). However, if they decide to invest in different industries, we assume that the returns are uncorrelated ($\rho = 0$).

The investment opportunities are only available at date $t = 0$. Both portfolios generate a return of R with probability λ or a return of zero with probability $(1 - \lambda)$ at $t = 1$. Note that we assume that the investment opportunity has a positive net present value (NPV), that is, $\lambda R > 1$, and that $\lambda \geq 1/2$. The latter can be motivated by considering the Value at Risk constraint of the Basel Accord, which states that banks must choose a minimum quality for their loan portfolio to limit their default probability. Consequently, the decision in which industry to invest only affects the correlation of returns, but not their magnitude. This structure allows us to determine whether interbank connections incentivize banks to invest in correlated investments.

Finally, to model risk-neutral investors we follow Allen and Gale (2005) and Brusco and Castiglionesi (2007) in that we assume that the equity investor I_A (I_B) in region A (B) is endowed with e units of capital at $t = 0$ and has no endowment at date $t = 1$. He can use his endowment for either consumption or to buy bank shares. In the latter case the investor is entitled to receive dividends at $t = 1$ (denoted by d_1). His utility is then given by

$$u(d_0, d_1) = d_0 \lambda R + d_1$$

Since an investor can obtain a utility of $e\lambda R$ by immediately consuming his initial endowment (consumption at $t = 0$ is denoted by d_0), he has to earn an expected return of at least λR on his invested capital in order to give up consumption at date $t = 0$. By investing an amount e_0 at $t = 0$, the equity investor obtains a lifetime utility of $(e - e_0)\lambda R + d_1$. Hence, he will only buy bank shares if the expected utility from doing so is higher than the utility he would get from immediately consuming his endowment, that is, if $(e - e_0)\lambda R + E[d_1] \geq e\lambda R$. This leads to the following participation constraint for investors:

$$E[d_1] \geq \lambda e_0 R$$

Under the assumption of perfect competition in the banking market (i.e., creditors have all the bargaining power), this constraint will be binding. Hence, the total amount of funds provided to the bank is given by $c + e = o = 1$. Due to the prevailing capital structure of banks, we assume that $c > e$, that is, that the bank has more debt than equity.

The timing of our model is as follows:

$t = 0$ — $t = 1$

Investors provide equity e_o
Banks exchange K
Banks raise debt capital c
Banks invest in loan portfolio

Cash flows are realized
Governments decide on bailouts

If both investments are successful, the banks are able to settle their interbank claims, repay the uninsured creditors, and pay the investors a positive dividend. If, however, the investment of one or both banks fails, either one or both banks may not be able to meet their liabilities and will consequently default. In case of a default we assume that there is a positive probability α that the government of the respective country will step in and bail out the bank, that is, take over the bank and repay all its liabilities.[6] It would be reasonable to assume that α is initially increasing in the interconnectedness of the bank (too interconnected to fail), its balance sheet size (too big to fail) and the number of failing banks (too many to fail). However, as soon as the bank reaches a critical size, it becomes "too big to save" and therefore its bailout becomes impossible and α drops to zero. Since we want to isolate the direct effect that cycle flows have on the expected repayment of uninsured creditors, we assume that the bailout probability is not increasing in either the balance sheet size of the bank or its interconnectedness or the number of failing banks. Making the bailout probability increasing with one of these factors would reinforce our results, since this gives banks an incentive to increase their interconnectedness even further. However, we capture the argument of being too big to save by assuming that the bailout probability becomes zero as soon as a bank's balance sheet exceeds a critical threshold $\overline{L} >> R$.[7] If the bank's size reaches this threshold, the government will no longer be able to provide enough capital to bail it out. Therefore, α becomes:

$$\alpha = \begin{cases} \alpha_B & \text{if } (c+K)R_D \leq \overline{L} \\ 0 & \text{if } (c+K)R_D > \overline{L} \end{cases}$$

Consequently, the payments to the uninsured creditors and investors depend on the performance of the loan portfolio and on whether a bank is bailed out if a default occurs. As described in the previous section, we can derive our results no matter which party (i.e., creditors or banks) has the bargaining power. To ensure consistency with our

[6]This is a simplification, since the bailout probability for different banks is probably correlated. However, for our mechanism to work, it is sufficient that the bailout probabilities are not perfectly correlated. This is certainly true if the banks are established in different countries. Furthermore, the recent crisis, the bailout of Bear Stearns, and the default of Lehman Brothers show that bailout decisions are also not perfectly correlated within the same country.

[7]This assumption is supported by the findings of Acharya et al. (2011). These authors show that financial sector bailouts and sovereign credit risk are linked. On the day of the announcement of large bailouts, the CDS spreads on government bonds rose significantly. If a government has to spend very high amounts to rescue a bank, it becomes virtually impossible to obtain funding for this bailout at acceptable terms. Thus, once a bank is too large, it can no longer be rescued.

extension that considers risk-averse creditors, we assume here that the creditors have all the bargaining power. Due to perfect competition in the banking sector, this implies that banks seek to maximize the repayment of uninsured creditors by choosing the parameters R_D, ρ, and K. Having described the setup, we now return to our main questions in this section: Which level of interconnectedness do banks choose and do they prefer to invest in correlated assets?

Both aspects are important to consider, since they both increase systemic risk. On the one hand, interconnectedness leads to systemic risk resulting from spillover effects that are transmitted through the interbank market (even without correlation on the asset side of the banks' balance sheet). On the other hand, even without being interconnected, correlation increases systemic risk due to possible joined bank failures. The following analysis investigates the interaction between these two sources of systemic risk and determines how interconnectedness influences the banks' investment decision, that is, whether they invest in correlated loan portfolios. To analyze this issue we derive the highest expected repayment banks can achieve with an investment correlation of zero and one, respectively. Then we compare the resulting repayments to determine which of the two yields a higher return for uninsured creditors.

4.1. Positively Correlated Investments

Consider first the situation where bank investments are perfectly positively correlated, that is, $\rho = 1$. In this case there are five different outcomes (depending on the success of the investments and whether the banks are bailed out or not), depicted in Table 1.

$\rho = 1$	Prob.	L_A	L_B	B_A	B_B	C_A	C_B	I_A	I_B
S_1	λ	S	S	N	N	cR_D	cR_D	$R - cR_D$	$R - cR_D$
S_2	$(1-\lambda)\alpha^2$	F	F	B	B	cR_D	cR_D	0	0
S_3	$(1-\lambda)(1-\alpha)\alpha$	F	F	B	N	cR_D	$cR_D \frac{K}{c+K}$	0	0
S_4	$(1-\lambda)(1-\alpha)\alpha$	F	F	N	B	$cR_D \frac{K}{c+K}$	cR_D	0	0
S_5	$(1-\lambda)(1-\alpha)^2$	F	F	N	N	0	0	0	0

Table 1: Capital flows for investment correlation of $\rho = 1$

Column 1 presents the five different states, while column 2 presents the probability of each given state occurring. Columns L_A and L_B show whether the investments of banks B_A and B_B are successful (S) or not (F). Columns B_A and B_B show whether banks B_A and B_B are bailed out by the government (B) or not (N). The columns C_A and C_B show the repayment of uninsured creditors, while columns I_A and I_B show the dividends the equity holders receive. To understand the cash flows presented in Table 1, first note that if either both investments are successful (S_1) or both banks are bailed out (S_2), the uninsured creditors of both banks will receive their full repayment. These states only differ with respect to the dividend paid to the investor, since in the case of a bailout the government takes over the bank and thus has the residual claim. If only one bank is bailed out (S_3 and S_4), then the creditor of this bank will receive his full repayment whereas the creditor of the other bank will receive only a fraction $\frac{K}{c+K}$ of his claim cR_D. Since the model is symmetric, it is sufficient to focus on the optimization problem of one of the banks. Hence, we only analyze the behavior of bank B_A. Due to perfect competition, bank B_A wants to maximize the expected repayment to its uninsured creditor C_A. Thus, its optimization problem becomes:

$$\max_{R_D,K} U_1 = \lambda c R_D + (1-\lambda)\left[\alpha c R_D + \alpha(1-\alpha) c R_D \frac{K}{c+K}\right] \quad (1)$$

subject to

$$E[d_1] \geq \lambda e R$$

The objective function consists of the following parts: With probability λ the investment of the bank is successful and creditors receive their contractually specified repayment cR_D. With probability $(1-\lambda)$ the investment fails. In this case the return of the creditors depends on whether the banks are bailed out or not. Specifically, if bank B_A is bailed out (which happens with probability α), the government repays all liabilities and hence its creditors again receive the full repayment. If, however, the government decides not to bail out bank B_A, the repayment depends on whether bank B_B is bailed out (remember that since investment outcomes are perfectly correlated, bank B_B is in default as well). If bank B_B is not bailed out either, the repayment is clearly zero. However, if bank B_B is bailed out, then the government injects funds of $R_D(c+K)$. This allows bank B_B to settle all its claims. Therefore, B_A receives $R_D K$ and has to split these proceeds on a pro rata basis (it owes money to its uninsured creditor C_A and bank B_B). Therefore, the uninsured creditors of bank B_A will receive a share $\frac{c}{c+K}$ of the funds bank B_A received from B_B. Furthermore, the binding participation constraint of the equity holder implies

$$E[d_1] = e\lambda R \Rightarrow \lambda(R - cR_D) = e_0 \lambda R \Rightarrow R_D = R$$

Inserting $R_D = R$ into (1) yields the following maximization problem:

$$\max_{K} U_1 = \lambda c R + (1-\lambda)\left[\alpha c R + \alpha(1-\alpha) c R \frac{K}{c+K}\right] \quad (2)$$

Since R and c are given, it will depend on K whether the government will be able to fully repay the bank's liabilities in case of a bailout. Let $\overline{K_1}$ denote the interbank exposure where the government is just able to repay all liabilities; this will be given by $\overline{K_1} = \frac{\overline{L}}{R} - c$. In the following we split the amount of interbank deposits into two intervals. For $K \in [0, \overline{K_1}]$ (government will be able to repay all liabilities and $\alpha = \alpha_B$) the first-order condition of the objective function becomes

$$\frac{\partial U_1}{\partial K} = R \frac{\alpha_B (1-\alpha_B)(1-\lambda) c^2}{(c+K)^2} > 0 \quad (3)$$

If, on the other hand, banks increase their exposure to an even higher level, that is, $K \in (\overline{K_1}, \infty]$, then the government will not be able to provide enough funds to settle all the liabilities of the failed bank and the bailout probability α drops to zero. Hence, the expected repayment of C_A drops to $\lambda c R$.

Thus, the expected utility of the uninsured creditors is increasing in K as long as

$R(c+K) < \overline{L}$. This implies that banks will choose an amount of interbank deposits $K = \overline{K_1}$ such that $R(c+K) = \overline{L}$.[8] Increasing cross-exposure on the interbank market beyond this threshold decreases the expected repayment of the uninsured creditor. Therefore, the highest expected utility for the creditor that can be achieved when choosing a correlation $\rho = 1$ is given by

$$\overline{U_1} = \lambda cR + (1-\lambda)\left[\alpha_B cR + \alpha_B(1-\alpha_B)\overline{L}\frac{c\overline{K_1}}{(c+\overline{K_1})^2}\right] \tag{4}$$

Our findings can be summarized in the following corollary.

Corollary 4.1. *If banks choose perfectly correlated investments (given a positive bailout probability), they will increase their interbank exposure up to the threshold $K = \overline{K_1}$, such that their total liabilities equal $\overline{L}$, that is, to a level that makes it just possible to bail them out in case of default.*

Proof The proof follows from the previous discussion. QED

To understand why it makes sense intuitively to choose such a high level of interbank deposits, one must consider two opposing effects. On the one hand, higher exposure increases the funds injected by the government in case of a bailout and hence increases the funds that can be split among a bank's creditors. On the other hand, a higher amount of interbank deposits decreases the fraction that the uninsured creditor of the bank that is not bailed out receives, since $\frac{c}{c+K}$ decreases in K. Since the first effect outweighs the second effect, banks choose the highest possible liabilities $\overline{L}$.

4.2. Uncorrelated Investments

We next turn to the case where banks decide to invest in different industries, that is, $\rho = 0$. Here, two scenarios must be considered. On the one hand, the interbank exposure can be chosen such that even if the one bank's investment is successful but the other bank's investment fails, the first bank will be unable to repay its obligations and hence financial contagion will occur. On the other hand, if the exposure is low enough, a successful bank will stay solvent no matter what happens to the other bank. Let K^* denote the "switching point", that is, the level of interbank exposure where a successful bank will just stay solvent, even if the other bank fails (see the Appendix for the derivation of K^*). The different possibilities for the cash flows are presented in Tables 2 and 3, where the notation is as described before. It is crucial to note that the interest rate R_D differs between the two possibilities, since the participation constraint of the equity investors differs. Table 2 presents the cash flows for $K \leq K^*$.

[8] Due to minimum capital requirements, banks must often back interbank loans with equity capital. Hence, depending on the risk weights of interbank loans and the banks' amounts of equity, there may be an individual upper limit for the banks' interbank exposure K that prevents banks from increasing their exposure to $K = \overline{K_1}$.

$\rho=0$	Prob.	L_A	L_B	B_A	B_B	C_A	C_B	I_A	I_B
S_1	λ^2	S	S	N	N	cR_D^1	cR_D^1	$R-cR_D^1$	$R-cR_D^1$
S_2	$(1-\lambda)^2\alpha^2$	F	F	B	B	cR_D^1	cR_D^1	0	0
S_3	$(1-\lambda)^2(1-\alpha)\alpha$	F	F	B	N	cR_D^1	$cR_D^1\frac{K}{c+K}$	0	0
S_4	$(1-\lambda)^2(1-\alpha)\alpha$	F	F	N	B	$cR_D^1\frac{K}{c+K}$	cR_D^1	0	0
S_5	$(1-\lambda)^2(1-\alpha)^2$	F	F	N	N	0	0	0	0
S_6	$\lambda(1-\lambda)\alpha$	S	F	N	B	cR_D^1	cR_D^1	$R-cR_D^1$	0
S_7	$\lambda(1-\lambda)\alpha$	F	S	B	N	cR_D^1	cR_D^1	0	$R-cR_D^1$
S_8	$\lambda(1-\lambda)(1-\alpha)$	S	F	N	N	cR_D^1	$cR_D^1\frac{K}{c+K}$	X_0	0
S_9	$\lambda(1-\lambda)(1-\alpha)$	F	S	N	N	$cR_D^1\frac{K}{c+K}$	cR_D^1	0	X_0

Table 2: Outcomes for $K \leq K^*$, where $X_0 = R - cR_D^1 - cR_D^1\frac{K}{c+K}$ - No contagion

States $S_1 - S_5$ parallel the respective outcomes in Table 1. Things differ from the results of Table 1 if only one investment fails, depending on whether the successful bank stays solvent (no contagion; see Table 2) or also becomes insolvent (see Table 3). If the interbank exposure is low enough ($K \leq K^*$) such that there is no contagion, then the successful bank can always fully repay its uninsured creditor, whereas the creditor of the unsuccessful bank will only receive the full amount if this bank is bailed out (S_6 and S_7 in Table 2). Otherwise, he will get just a fraction of his repayment (S_8 and S_9 in Table 2). If, on the other hand, the interbank exposure is higher than the threshold K^*, the successful bank will not be able to settle its interbank liabilities and, on top of that, will be unable to fully repay its creditor. Depending on which bank (if any) is bailed out, the creditors of both the successful and the failed bank receive either their full repayment or just a fraction ($S_6 - S_{13}$ in Table 3).

$\rho=0$	Prob.	L_A	L_B	B_A	B_B	C_A	C_B	I_A	I_B
S_1	λ^2	S	S	N	N	cR_D^2	cR_D^2	$R-cR_D^2$	$R-cR_D^2$
S_2	$(1-\lambda)^2\alpha^2$	F	F	B	B	cR_D^2	cR_D^2	0	0
S_3	$(1-\lambda)^2(1-\alpha)\alpha$	F	F	B	N	cR_D^2	$cR_D^2\frac{K}{c+K}$	0	0
S_4	$(1-\lambda)^2(1-\alpha)\alpha$	F	F	N	B	$cR_D^2\frac{K}{c+K}$	cR_D^2	0	0
S_5	$(1-\lambda)^2(1-\alpha)^2$	F	F	N	N	0	0	0	0
S_6	$\lambda(1-\lambda)\alpha$	S	F	N	B	cR_D^2	cR_D^2	$R-cR_D^2$	0
S_8	$\lambda(1-\lambda)(1-\alpha)\alpha$	S	F	B	N	cR_D^2	$cR_D^2\frac{K}{c+K}$	0	0
S_9	$\lambda(1-\lambda)(1-\alpha)^2$	S	F	N	N	$R\frac{c+K}{c+2K}$	$R\frac{K}{c+2K}$	0	0
S_{10}	$\lambda(1-\lambda)\alpha$	F	S	B	N	cR_D^2	cR_D^2	0	$R-cR_D^2$
S_{12}	$\lambda(1-\lambda)(1-\alpha)\alpha$	F	S	N	B	$cR_D^2\frac{K}{c+K}$	cR_D^2	0	0
S_{13}	$\lambda(1-\lambda)(1-\alpha)^2$	F	S	N	N	$R\frac{K}{c+2K}$	$R\frac{c+K}{c+2K}$	0	0

Table 3: Outcomes for $K > K^*$ - Contagion

In a next step we must compare the expected repayments of the uninsured creditor in these two scenarios, that is, $K \leq K^*$ and $K > K^*$. To do so, we first derive the precise values of R_D^1 and R_D^2 from the binding participation constraint of the equity holder. If $K \leq K^*$, we obtain from the constraint $E[d_1] \geq e\lambda R$

$$\lambda^2(R - cR_D^1) + \lambda(1-\lambda)\left[\alpha(R - cR_D^1) + (1-\alpha)\left(R - cR_D^1 - cR_D^1\tfrac{K}{c+K}\right)\right] \geq (1-c)\lambda R$$
$$\Rightarrow R_D^1 = R\tfrac{c+K}{c+2K-K[\lambda+(1-\lambda)\alpha]}$$

The interest rate R_D^1 is decreasing in the interbank exposure K since, due to an increased K, a higher fraction of the investment return is paid from bank B_A to B_B and thus creditor

C_B receives a higher repayment. This reduces the dividend payment of I_A. Hence, to satisfy the investor's participation constraint, R_D^1 must be reduced. Furthermore, R_D^1 is increasing in the success probability λ, the bailout probability α, and the investment return R. The success probability λ increases the probability that the equity investor will receive a dividend payment. Hence, a lower dividend payment is sufficient to satisfy his participation constraint. An increase in the bailout probability α makes the bailout of the other bank more likely in case of default. This increases the probability that the investor will receive the full dividend payment $R - cR_D^1$.

Furthermore, the interest rate R_D^1 depends on the amount of debt borrowed from the uninsured creditors. Consider, for example, bank B_B. If this bank increases c, C_B is entitled to a higher fraction of the bank's liquidation value. Hence, the fraction paid back into the interbank market is lower. This reduces the dividend payment equity investor I_A receives and thus the interest rate R_D^1 must be reduced. If, on the other hand, bank B_A increases c, investors have to invest less equity and the interest rate R_D^1 can be increased. Conversely, if $K > K^*$, we obtain

$$\begin{aligned} &\lambda^2(R - cR_D^2) + \lambda(1-\lambda)\alpha(R - cR_D^2) \geq (1-c)\lambda R \\ &\Rightarrow R_D^2 = R\left(\frac{\lambda + (1-\lambda)\alpha - (1-c)}{c[\lambda + (1-\lambda)\alpha]}\right) \end{aligned} \tag{5}$$

Therefore, as soon as $K > K^*$, a change in K does not alter the dividend payment to I_A and hence no longer changes the interest rate R_D^2. For the same reasons as for R_D^1, R_D^2 is increasing in the success probability λ, the bailout probability α, the debt amount c, and the investment return R. Given our assumptions on λ, c, and e, we can make sure that $0 < R_D^2 < R$. Plugging the value of R_D^1 (since we approach K^* from below) into the formula for K^* in equation (33) (see the Appendix) yields

$$K^* = \frac{c(1-c)}{\lambda + (1-\lambda)\alpha - 2(1-c)}$$

Hence, to obtain a positive interbank exposure K for which the successful bank stays solvent (in case one bank is successful and the other is not), it must hold that $\lambda + (1-\lambda)\alpha - 2(1-c) > 0$. Otherwise, we can restrict our analysis to the case $K > K^*$. Therefore, if the investment correlation is zero, the overall utility of the uninsured creditors (depending on the amount of interbank deposits) is

$$\begin{aligned} U_0(K \leq K^*) &= [\lambda + (1-\lambda)\alpha]cR_D^1 + (1-\lambda)(1-\alpha)[\lambda + (1-\lambda)\alpha]cR_D^1\frac{K}{c+K} \\ &= [\lambda + (1-\lambda)\alpha]cR \\ U_0(K > K^*) &= \left[\alpha(1+\lambda) + \lambda^2(1-2\alpha) - \alpha^2\lambda(1-\lambda)\right]cR_D^2 + \lambda(1-\lambda)(1-\alpha)^2 R \\ &+ \alpha(1-\lambda)(1-\alpha)cR_D^2\frac{K}{c+K} \end{aligned}$$

We now have to compare the utility of the creditors for the different levels of interbank deposits. In the Appendix, we formally show that banks have an incentive to choose a level of interbank deposits $\overline{K_0} = \frac{\overline{L}}{R_D^2} - c$ in case $(c + K^*)R_D^1 < \overline{L}$. If, on the other hand, $(c + K^*)R_D^1 \geq \overline{L}$, banks will be indifferent between all possible interbank exposures in the interval $K = [0, \overline{K_0}]$. Hence, if $(c + K^*)R_D^1 < \overline{L}$, the highest expected utility for the non-insured creditor that can be achieved when choosing a correlation of $\rho = 0$ is given by

$$\begin{aligned}\overline{U_0} &= \left[\alpha(1+\lambda) + \lambda^2(1-2\alpha) - \alpha^2\lambda(1-\lambda)\right] cR_D^2 + \lambda(1-\lambda)(1-\alpha)^2 R \\ &+ \alpha(1-\lambda)(1-\alpha)\overline{L}\frac{c\overline{K_0}}{(c+\overline{K_0})^2}\end{aligned}$$

Furthermore, if $(c + K^*)R_D^1 \geq \overline{L}$, the maximal expected utility becomes

$$\overline{U_0} = \lambda cR + (1-\lambda)\alpha cR$$

This finding can be summarized in the following corollary:

Corollary 4.2. *If banks choose uncorrelated investments (given a positive bailout probability), two scenarios must be considered:*

a) *If $(c + K^*)R_D^1 < \overline{L}$, banks will increase their interbank exposure up to the threshold $K = \overline{K_0}$,*

b) *If $(c + K^*)R_D^1 \geq \overline{L}$, banks will be indifferent between all possible interbank exposures in the interval $K = [0, \overline{K_0}]$.*

Proof See the Appendix. QED

Hence, intuitively, two cases must be distinguished. On the one hand, the level of interbank exposure above which contagion occurs may be low enough so that the bank can be bailed out ($(c + K^*)R_D^1 < \overline{L}$). Then it is always optimal to increase the interbank exposure K to a level that just enables the government to bail out the bank ($K = \overline{K_0}$), implying that contagion can occur. This is due to the fact that as soon as the interbank exposure K exceeds the contagion threshold K^*, a change in K no longer alters the interest rate R_D^2. Therefore, the only downside for the bank's creditor in choosing a higher K is due to the states where only his own bank is successful. In such cases a higher interbank exposure implies that a higher fraction of the return generated by that bank is transferred to the creditor of the other bank. However, the creditor benefits in the same way in case his own bank fails while the other bank is successful. The benefits and costs of the respective states add up to zero. An additional upside of a higher K results from the state where both banks fail and the other bank is bailed out (as described in Section 4.1). Taken together, these effects incentivize banks to increase their interbank exposure up to $\overline{K_0}$.

If, on the other hand $(c + K^*)R_D^1 \geq \overline{L}$, the banks are unable to choose an interbank exposure that leads to contagion and at the same time allows the government to bail out the bank ($K \leq \overline{K_0} < K^*$). In this interval for K, equity investors receive a dividend payment whenever their own bank is successful. This payment, however, only depends on the interbank exposure if the other bank fails and is not bailed out. In this state, if K is increased, R_D^1 must be reduced. Consequently, this effect will reduce the creditor's expected repayment. However, there is also a countervailing effect if the bank increases K. Due to the effect previously described, an increase in K increases the expected repayment to the creditor in cases where his own bank fails but the other bank is either successful or bailed out. These two effects offset each other such that the expected repayment to the uninsured creditor is not influenced by the choice of the interbank exposure in the interval $K \leq \overline{K_0} < K^*$.

4.3. Comparison of Correlated and Uncorrelated Investments

What remains is to show under which correlation structure uninsured creditors receive a higher expected repayment. In the Appendix we formally prove that $\overline{U_1} > \overline{U_0}$ will always hold, implying that banks will always choose investments that are perfectly correlated. This main finding can be summarized in the following proposition.

Proposition 4.3. *If banks are connected via an interbank market and there is a non-zero bailout probability, it is optimal for them to invest in correlated assets. Moreover, they have an incentive to increase their interbank exposure until their total liabilities equal $\overline{L}$, that is, the highest amount that still allows the bank to be bailed out.*

Proof See the Appendix. QED

To understand why this result holds, first note that the contractually specified repayment R_D^2 is lower if investments are uncorrelated (since $R_D^2 < R$). This is because the investor has a lower probability of receiving a dividend payment if investments are independent. Hence, a higher dividend amount (and thus a lower contractually specified repayment for the creditor) must be paid to satisfy the investor's participation constraint. Moreover, the probability of receiving the full repayment is lower as well, if investments are uncorrelated. Furthermore, the lowest positive repayment the creditors can receive is higher with perfectly correlated investments (this lowest repayment occurs with the same probability under either correlation structure). Finally, if investments are uncorrelated, there is an additional "intermediate" state in which the creditor receives only a fraction of his repayment (again as in the case of a zero bailout probability). These three effects together more than offset the fact that a zero repayment is more likely (occurs with probability $(1-\lambda)(1-\alpha)^2$ compared to $(1-\lambda)^2(1-\alpha)^2$) if investments are perfectly correlated. Hence, these three effects dominate the diversification effect that results from investing in different industries, which is why banks prefer to invest in correlated loan portfolios.

In this section we demonstrate that banks always have an incentive to increase the interbank exposure until the government is just able to bail them out. The benefit of being connected to other banks can be further enhanced by choosing correlated assets. This gives banks an incentive to herd. We can thus provide an additional explanation for the herding behavior of banks besides the effect discussed by Acharya and Yorulmazer (2007). In their paper correlated investments increase the bailout probability of each bank. Even if we abstract from the fact that correlated investments increase the bailout probability, we find an additional incentive for herding behavior. Hence, the mechanism described in this paper leads to an overall increase in systemic risk that results from both interconnectedness as well as herding behavior.

Given that it is optimal for banks to invest in correlated portfolios to maximize their creditors' repayment, we henceforth restrict our analysis to the case where banks invest in correlated investment portfolios.

5. The Interbank Network and Risk Shifting

After showing that it is optimal for banks to invest in correlated investments, we now use this finding and consider the impact of interbank connections on the incentive of banks to engage in risk shifting. To model the riskiness of the investment decision, we consider two assets: a risk-free storage technology that transfers one unit of wealth today into one unit of wealth tomorrow, and a risky negative NPV investment that generates a return $R_N > 1$ with probability $\lambda_N < 1$ such that $\lambda_N R_N < 1$. As in the previous section, banks get c from uninsured creditors and e from equity holders such that $c + e = 1$. Depending on the asset the bank invests in and given that there is no bailout possibility, it can offer creditors either a repayment of c (if it invests in the safe asset) or cR_D^N with probability λ_N and $R_D^N \leq R_N$ if it invests in the risky negative NPV asset. The promised repayment R_D^N results from the binding participation constraint of the equity holder. We assume that the outside option of the equity holder is now given by the risk-free storage technology. Therefore the participation constraint becomes

$$E[d_1] = e \Rightarrow \lambda_N(R_N - cR_D^N) = e \Rightarrow R_D^N = \frac{\lambda_N R_N - (1-c)}{c\lambda_N}$$

Furthermore, since $\lambda_N R_N < 1$

$$R_D^N = \frac{\lambda_N R_N - (1-c)}{c\lambda_N} < \frac{1-(1-c)}{c\lambda_N} = \frac{1}{\lambda_N} \Rightarrow \lambda_N R_D^N < 1$$

We first consider a scenario without a bailout possibility and no interbank network. Here, it can be easily seen that the expected repayment of the creditors is higher if the bank invests in the safe asset since

$$c > \lambda_N c R_D^N \tag{6}$$

Hence, without the possibility of a bailout, banks will always choose the safe investment. Next we consider the case where the bank has a positive probability of being bailed out by the government but still no connections to other banks. Now it can become profitable to switch to the negative NPV investment if the bailout probability is high enough. More precisely, a bank will switch to the negative NPV investment if the expected repayment of creditors for this investment is higher than for the safe repayment c. This yields the condition

$$\lambda_N c R_D^N + (1-\lambda_N)\alpha c R_D^N > c \tag{7}$$

Besides the state of nature where the investment is successful, creditors now also receive the higher return R_D^N when the bank is bailed out by the government. The critical α, that is, the bailout probability where the bank is indifferent between the two investments is given by

$$\alpha^* = \frac{1 - \lambda_N R_D^N}{(1-\lambda_N) R_D^N} < 1$$

Hence, for $\alpha > \alpha^*$ it is always profitable to switch to the negative NPV investment. Now we again allow the bank to exchange funds with the bank in the other region. As before, the banks exchange funds K in period 0 in return for a payment of KR_D^N at $t = 1$. Whether banks will switch to the negative NPV investment again depends on α. Whenever the expected repayment of the uninsured creditor from investing in the negative NPV investment opportunity is higher, banks will shift away from the risk-free investment. Formally, the following condition must be satisfied:

$$\lambda_N c R_D^N + (1-\lambda_N)\left[\alpha c R_D^N + \alpha(1-\alpha) c R_D^N \frac{K}{c+K}\right] > c \tag{8}$$

Solving this equation for α yields the critical threshold:

$$\alpha^{**} = \frac{c+2K}{2K} - \sqrt{\frac{(c+2K)^2}{4K^2} - \frac{(c+K)(R_D^N\lambda_N - 1)}{KR_D^N(\lambda_N - 1)}} \tag{9}$$

We show in the Appendix that the critical α is strictly smaller if a bank is connected (i.e., $K > 0$) to another bank on the interbank market, that is, $\alpha^* > \alpha^{**}$. Hence, the critical threshold α is lower once a bank enters into connections with other banks. Put differently, a lower bailout probability is sufficient to make the bank switch to the negative NPV investment. The positive bailout probability can turn a negative NPV investment into a positive NPV investment from the perspective of the uninsured creditors since they will receive the high repayment with a higher probability. This effect is reinforced once the bank is connected to another bank if this other bank has a positive bailout probability as well. Our results are summarized in the following proposition.

Proposition 5.1. *The more interconnected a bank becomes, the lower the critical bailout probability that makes it profitable for the bank to engage in risk shifting, that is, to switch to negative NPV investments*

Proof See the appendix. QED

Risk shifting thus becomes more attractive for banks since the downside risk is limited by two factors. First, the downside risk is limited by the positive bailout probability because creditors receive their full repayment after the bank is bailed out. Second, the interbank connection further reduces the downside risk, since it adds an additional state where the creditor receives a positive repayment. These two effects turn a negative NPV investment into a positive NPV investment (from the perspective of the uninsured creditors).

Taking the results of Sections 4 and 5 together may help explain why many banks invested in highly correlated low quality assets in the run-up to the financial crisis (e.g., subprime loans). Section 4 shows that interbank connections incentivize banks to invest in highly correlated portfolios because they benefit from defaulting in states where the banks they are connected to default as well. This section additionally shows that, given that banks prefer correlated investment projects, interbank connections make risk shifting (i.e., investing in low quality assets rather than safe assets) more attractive (as long as there is a positive probability that defaulting banks are bailed out). Hence, one reason for the observed investment behavior prior to the financial crisis may be that the high interconnectedness of large banks incentivized them to invest in highly correlated low quality assets.

6. Extensions

6.1. Three Region Economy

So far we have assumed that the economy consists of only two regions. This gave banks an incentive to increase the funds exchanged, K, in period 0 up to $\overline{K_1}$. Now we want to focus on whether the benefits from taking advantage of the bailout possibility have an influence on the interbank network size and structure. In particular, we analyze the change in the expected utility of the creditors after an additional bank is added to the interbank network. Furthermore, we analyze whether the creditors derive a higher utility if the network is directed or bidirected (see Figure 3).

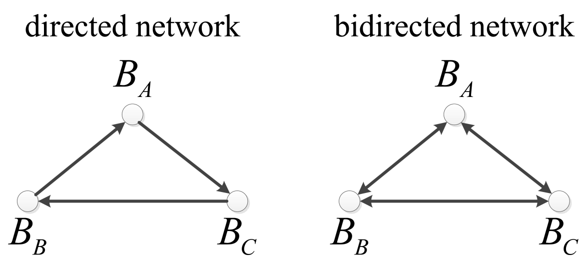


Figure 3: Interbank network structures

Afterwards we investigate how the desired network structure changes if we relax the assumption that the governments in each region (country) can provide exactly the same amount of bailout funds. To derive these results, we extend our model to a three region economy (A, B, and C) and start the analysis by checking whether this improves the expected repayment of the uninsured creditors. First, we examine a directed interbank network. In this case, banks deposit funds K in a neighboring region and receive funds from another neighboring region in return for a payment of KR_D at $t = 1$. Since the model is still symmetric, the expected utility of all uninsured creditors is the same. Hence, it is sufficient to consider only one specific bank and its creditor. In this setup, the expected repayment (U_{DI}) of the uninsured creditors in $t = 1$ becomes

$$U_{DI} = \lambda c R_D + (1-\lambda)\left[\ \alpha c R_D + (1-\alpha)\alpha c R_D \tfrac{K}{c+K} + (1-\alpha)^2 \alpha c R_D \tfrac{K^2}{(c+K)^2}\ \right] \qquad (10)$$

To fully capture the respective repayments in the different default states, consider the view of a creditor of bank B_A. If bank B_A is bailed out, the creditor receives the full repayment. If bank B_A is not bailed out, the repayment of the creditor depends on what happens to the other banks. If bank B_B is bailed out, the creditor receives a fraction $\frac{K}{c+K}$ of his promised repayment. If bank B_B is not rescued but bank B_C is, then the creditor receives a fraction $\frac{K^2}{(c+K)^2}$. Due to the perfect correlation of the banks' investments, the binding participation constraint of the equity holders is again $E[d_1] = \lambda Re$, implying that $R_D = R$. We maintain the assumption of perfect competition, implying that banks must still maximize the expected repayment of their creditors. Hence, the maximization problem for a specific bank becomes

$$\max_K U_{DI} = \lambda cR + (1-\lambda)\left[\ \alpha cR + (1-\alpha)\alpha cR\tfrac{K}{c+K} + (1-\alpha)^2\alpha cR\tfrac{K^2}{(c+K)^2}\ \right] \tag{11}$$

Again, we split the amount of interbank deposits into two intervals. In the interval $K \in [0, \overline{K_1}]$ the government will be able to bail out the bank and repay all liabilities. Hence, for this interval, $\alpha = \alpha_B$ and the derivative of the objective function becomes

$$\frac{\partial U_{DI}}{\partial K} = (1-\lambda)(1-\alpha_B)\alpha_B cR\left[\frac{c}{(c+K)^2} + (1-\alpha_B)\frac{2cK}{(c+K)^3}\right] > 0 \tag{12}$$

Thus, increasing K again enhances the expected utility of the creditor in this interval. If, on the other hand, banks increase their exposure even more, that is, $K \in (\overline{K_1}, \infty]$, the bailout probability α drops to zero. Hence, the expected repayment to C_A drops again to λcR. Thus, in the three region case with a directed interbank network, the expected utility of the uninsured creditors is increasing in K as well, as long as $R(c+K) < \overline{L}$. This implies that banks will choose the same amount of interbank deposits $K = \overline{K_1}$ as in the two region case. Therefore, the highest expected utility that can be achieved is

$$\overline{U_{DI}} = \lambda cR + (1-\lambda)\begin{bmatrix} \alpha_B cR + (1-\alpha_B)\alpha_B cR\frac{\overline{K_1}}{c+\overline{K_1}} \\ +(1-\alpha_B)^2\alpha_B cR\frac{\overline{K_1}^2}{(c+\overline{K_1})^2} \end{bmatrix} \tag{13}$$

$$= \lambda cR + (1-\lambda)\begin{bmatrix} \alpha_B cR + (1-\alpha_B)\alpha_B \overline{L}\frac{c\overline{K_1}}{(c+\overline{K_1})^2} \\ +(1-\alpha_B)^2\alpha_B \overline{L}\frac{c\overline{K_1}^2}{(c+\overline{K_1})^3} \end{bmatrix} \tag{14}$$

Comparing the maximal expected utility of the creditor in a three bank interbank market (see equation (14)) with the two bank case, where the bank in region A is only connected to one other region (see equation (4)), one can easily see that the expected utility increases if the bank is connected to more banks. Since in the three region case each bank is now linked to two other banks (instead of only one other bank) the expected repayment of the uninsured creditors increases. Moreover, the repayment of creditors is again increasing in the interbank exposure K. Therefore, banks will prefer to be connected to two banks instead of only one.

We now consider a bidirected interbank network structure, that is, a structure where each bank has bilateral exposure to all other banks. Since the model is still symmetric, we again restrict our analysis to bank B_A and its creditor. Table 4 summarizes the possible states for this network structure. If the investments are successful, all banks are able to settle their liabilities and no default occurs (S_1). Hence, the uninsured creditor receives cR_D and the investor receives a dividend $R-R_D$. If the investment fails, the repayment to the uninsured creditors depends on whether the banks are bailed out or not. Several cases must be considered here. If B_A is bailed out by the government (states S_2 to S_4 and state S_6), creditor C_A receives the full repayment. If, however, only one or both of the other banks (B_B and B_C) are rescued, the creditor of bank B_A will receive only a fraction of the contractually specified repayment. In case both other banks are bailed out, each receives an amount $(c+2K)R_D$ from its respective government. Therefore, they are able to fully repay their creditors and settle their interbank claims. Hence, bank B_A receives KR_D from B_B and B_C, respectively, that is, $2KR_D$ in total. Since the bank's total liabilities are $(c+2K)R_D > 2KR_D$, it must split these funds on a pro rata basis among its creditors. Consequently, the uninsured creditor of bank B_A who holds a fraction $\frac{c}{c+2K}$ of the total liabilities receives a total payment of $cR_D\frac{2K}{c+2K}$. The remaining funds are paid back to the other banks.

$\rho = 1$	Prob.	L	B_A	B_B	B_C	C_A	I_A
S_1	λ	S	N	N	N	cR_D	$R-cR_D$
S_2	$(1-\lambda)\alpha^3$	F	B	B	B	cR_D	0
S_3	$(1-\lambda)(1-\alpha)\alpha^2$	F	B	B	N	cR_D	0
S_4	$(1-\lambda)(1-\alpha)\alpha^2$	F	B	N	B	cR_D	0
S_5	$(1-\lambda)(1-\alpha)\alpha^2$	F	N	B	B	$cR_D\frac{2K}{c+2K}$	0
S_6	$(1-\lambda)(1-\alpha)^2\alpha$	F	B	N	N	cR_D	0
S_7	$(1-\lambda)(1-\alpha)^2\alpha$	F	N	B	N	$cR_D\frac{K}{c+K}$	0
S_8	$(1-\lambda)(1-\alpha)^2\alpha$	F	N	N	B	$cR_D\frac{K}{c+K}$	0
S_9	$(1-\lambda)(1-\alpha)^3$	F	N	N	N	0	0

Table 4: Capital flows in a bidirected connected interbank network

We now discuss the states where only one bank receives funds from its government, that is, states S_7 and S_8. The symmetry of our model framework allows us to focus on state S_7, since the cash flows in S_8 can be derived analogously. To derive the exact repayment the uninsured creditor of bank B_A receives, we proceed in several steps. First, we determine the total amount of funds channeled through bank B_A during the repayment process. Since bank B_A is in default and funds are again split on a pro rata basis, the uninsured creditor receives a fraction of $\frac{c}{c+2K}$ of every unit of capital that arrives at bank B_A. The solution strategy is thus as follows: We start by tracking all funds injected into the financial system by the governments and follow these funds until they arrive at bank B_A for the first time. In a next step, we examine the funds that are paid back into the financial system and arrive again at bank B_A. The last step is necessary since capital flows are exchanged continuously between banks B_A and B_C. Note that since bank B_B is bailed out, all its liabilities are settled and hence all funds that arrive at bank B_B stay there.

Next we return to a detailed description of state S_7. In state S_7 only bank B_B is bailed out and thus receives funds of $(c+2K)R_D$, which is sufficient to settle all liabilities, implying that banks B_A and B_C both receive KR_D. A fraction $\frac{K}{c+2K}$ of these funds KR_D that bank B_C receives are passed on to bank B_A. Hence, bank B_A receives an amount $KR_D(1+\frac{K}{c+2K})$ in the first round. As described above, a fraction $\frac{c}{c+2K}$ is directly paid to the uninsured creditor, whereas each of the other banks receives a fraction $\frac{K}{c+2K}$. However, a fraction of the funds that go to bank B_C flows back to bank B_A. This implies that a fraction $\frac{K^2}{(c+2K)^2}$ is returned to bank B_A after the next cycle flow. After these funds arrive at bank B_A, the same flows occur again. This yields a capital flow to creditor C_A, that can be expressed as a geometric series:

$$\begin{aligned} KR_D\left(1+\frac{K}{c+2K}\right)\sum_{i=0}^{\infty}\left(\frac{K}{c+2K}\right)^{2i}\frac{c}{c+2K} &= KR_D\left(1+\frac{K}{c+2K}\right)\frac{c}{c+2K} \\ &= cR_D\frac{K}{c+K} \end{aligned} \quad (15)$$

As already discussed, state S_8 can be described analogously, implying that the creditor of bank B_A receives the same repayment in this state. Therefore, the expected repayment (U_{BI}) of the uninsured creditors in $t = 1$ can be written as

$$U_{BI} = \lambda cR_D + (1-\lambda)\left[\alpha cR_D + (1-\alpha)\alpha^2 cR_D\frac{2K}{c+2K} + 2(1-\alpha)^2\alpha cR_D\frac{K}{c+K}\right] \quad (16)$$

Again, the participation constraint of the investors implies that $R_D = R$. Due to the fact that U_{BI} is increasing in K until the total liabilities of the bank are equal to $\overline{L}$, the banks will again choose $K = \overline{K_{BI}}$, where $\overline{K_{BI}} = \frac{\overline{L}}{2R_D} - \frac{1}{2}c$ such that $(c + 2K)R_D = \overline{L}$. Hence, the maximal expected utility for the uninsured creditor in a bidirected interbank market is

$$\overline{U_{BI}} = \lambda cR + (1-\lambda)\left[\alpha_B cR + (1-\alpha_B)\alpha_B^2 cR\frac{2\overline{K_{BI}}}{c+2\overline{K_{BI}}} + 2(1-\alpha_B)^2\alpha_B cR\frac{\overline{K_{BI}}}{c\overline{K_{BI}}}\right] \quad (17)$$

Now we can compare the highest possible expected utility for creditors in a directed versus a bidirected interbank network. Comparing equations (13) and (17) shows that banks can maximize the expected repayment of their non-insured creditors by trying to establish large directed cycle flows within the interbank market instead of just creating bilateral exposure with other banks. This result can be summarized in the following proposition.

Proposition 6.1. *If all governments can spend equally high amounts for a bailout program, banks in a three region economy are incentivized to create large directed cycle flows instead of bilateral exposures.*

Proof See the Appendix. QED

This result also makes sense intuitively. To make as much use as possible of the bailout possibility, banks prefer being part of a long cycle flow instead of lending money only bilaterally. Thereby, they can benefit to a larger extent from the bailout of any of the banks that are part of the cycle. However, this mechanism only works if a bank can be sure that the other banks will continue to create this large cycle and not start to exchange funds bilaterally.

In a next step, we relax the assumption that the governments in the respective regions can provide the same amount of bailout funds and show how this influences the utility maximizing network structure. Therefore, we assume from now on that there is a different critical threshold $\overline{L}$ for each government (due to different country sizes) where banks become too big to save and therefore the bailout probability decreases to zero. Without loss of generality, we assume that country A can provide more bailout funds than country B, which in turn can provide more than country C. Hence, in the following we assume that $\overline{L_A} > \overline{L_B} > \overline{L_C}$.

In the beginning of this section, we show that the expected repayment of the uninsured creditor is maximized if banks establish a directed interbank network. However, here a directed interbank network is only utility enhancing until bank B_C reaches a balance sheet size of $\overline{L_C}$, which happens at an interbank exposure of K^C where $K_C = \frac{\overline{L_C}}{2R} - \frac{1}{2}c$. Exceeding this threshold would reduce the bailout probability of bank B_C to zero. Hence, if B_C's balance sheet exceeds $\overline{L_C}$, the expected utility for creditor C_C becomes

$$U_{DI}^{C}(K > K^{C}) = \lambda cR + (1-\lambda)\left[(1-\alpha)\alpha cR\frac{K}{c+K} + (1-\alpha)^2\alpha cR\frac{K^2}{(c+K)^2}\right] \tag{18}$$

Note that this is only true as long as the other two banks are still not too big to save. One can see directly from (18) that the expected repayment of C_C is smaller for $K > K^C$ than for an interbank exposure of $K = K^C$. Therefore, bank B_C does not have an incentive to accept additional funds from other banks as soon as it reaches an interbank exposure of K^C. However, at this point banks B_A and B_B would still be able to increase their interbank exposure to a certain extent without immediately becoming too big to save. Since B_C is not willing to borrow any additional funds on the interbank market, the only option to increase the interbank exposure of B_A and B_B is to lend and borrow bilaterally. Now we must check whether this enhances the expected repayment of creditors C_A and C_B.

Since an additional bilateral interbank exposure between B_A and B_B does not alter the cash flows that are induced by the directed interbank network created by banks B_A, B_B, and B_C, we can consider the bilateral exposure between B_A and B_B in isolation. This added value of bilateral exposure was already discussed in Section 4. Therefore, we can conclude that banks B_A and B_B lend to and borrow from each other until bank B_B becomes too big to save as well. Hence, if governments differ in their ability to bail out banks, banks have an incentive to first establish a connected directed interbank network that includes all banks. As soon as some banks become too big to save they stop their borrowing and lending activities on the interbank market. The remaining banks (which are not yet too big to save) then continue to increase their interbank exposure by establishing directed capital flows between each other. This leads to an interbank network of very high density where the degree centrality of banks is increasing in their size, that is, bigger banks are more connected than smaller banks. Furthermore, our model predicts that larger banks tend to be established in countries with higher bailout possibilities.

Proposition 6.2. *If governments differ in their ability to bail out banks, the density of the interbank network will become very high and the degree centrality of banks will increase in their balance sheet size. Furthermore, large banks will be mainly established in countries with higher bailout possibilities.*

Proof Omitted.

Finally, we relax the assumption that the bailout probability α is not increasing in the interconnectedness of the bank (too interconnected to fail) or in its balance sheet size (too big to fail). As already noted, this reinforces the incentive for banks to have a high interbank exposure K. Furthermore, banks now have the incentive to channel funds through banks that have a very high probability of being bailed out in case of default, which in turn increases the bailout probability of these banks even more. This mechanism may lead to the core bank system that is present in almost all countries and often accounts for the large majority of the total interbank lending.

6.2. Risk Averse Creditors

From now on we allow uninsured creditors to be risk averse (in line with the literature on interbank networks and financial contagion, e.g., Allen and Gale (2000) and Brusco and Castiglionesi (2007)). Here, the interbank market not only is present for the reasons discussed in the previous sections, but also allows banks to co-insure against regional liquidity shocks as in Allen and Gale (2000). We show that even if the interbank market has a different reason to exist, our main mechanism is still present. Specifically, we show that banks have an incentive to increase their interbank exposure beyond the level that would be sufficient to perfectly co-insure against liquidity shocks. Our economy in this section now consists of three dates $t = 0, 1, 2$ and, again, two regions A and B, each with a continuum of identical banks that all adopt the same behavior and can thus be described by a representative bank (protected by limited liability). Furthermore, there are now n ex ante identical uninsured creditors and again one risk-neutral investor. Creditors have Diamond-Dybvig (1983) preferences, that is,

$$U(c_1, c_2) = \begin{cases} u(c_1) \text{ with probability } \omega^i \text{ (early creditors)} \\ u(c_2) \text{ with probability } 1 - \omega^i \text{ (late creditors)} \end{cases}$$

where the utility function $u(\cdot)$ is defined for non-negative numbers, strictly increasing, strictly concave, and twice continuously differentiable and satisfies Inada conditions. Each creditor is endowed with one unit of capital at date $t = 0$. Of the n creditors in each region there are n_e^i early creditors and n_l^i late creditors. Thus $\omega^i \equiv \frac{n_e^i}{n}$ represents the fraction of early creditors, where ω^i can be either high or low ($\omega_H > \omega_L$). There are two equally likely states S_1 and S_2. At date $t = 1$ state-dependent liquidity preferences are revealed (see Table 5).

Each region has the same ex ante probability of facing a high liquidity shock. A creditor's type is private information and the proportion of early creditors in the whole economy is given by $\gamma = \frac{\omega_H + \omega_L}{2}$. Thus, there is no aggregate uncertainty. At $t = 1$ all liquidity-related uncertainty is resolved and creditors learn their type.

There are two types of investment opportunities: a risk-free, liquid type and a risky, illiquid one. The risk-free asset is a storage technology that transfers one unit of capital at a certain period into one unit of capital in the following period. The illiquid asset is only available at date $t = 0$ and generates a return of either $R > 1$ with probability λ or zero with probability $(1 - \lambda)$ at date $t = 2$ for each unit of capital invested. Note that we again assume that the illiquid asset has a positive NPV, that is, $\lambda R > 1$, and that investment outcomes are again perfectly positively correlated across regions.

	A	B
S_1	ω_H	ω_L
S_2	ω_L	ω_H

Table 5: Liquidity shocks

Since our model now has three dates, the equity investors are entitled to receive dividends at $t = 1$ and $t = 2$. Hence, the investor's utility is now

$$u(d_0, d_1, d_2) = \lambda R d_0 + d_1 + d_2$$

As before, since investors can obtain a utility of λRe by immediately consuming the initial endowment, they must earn an expected return of at least λR on their invested money to give up consumption at date $t = 0$. This leads to the following participation constraint for investors:

$$E[d_1 + d_2] \geq e_0 \lambda R$$

Central Planner Economy

In this economy the Pareto-efficient allocation can be characterized as the solution to the problem of a planner maximizing the creditors' expected utility. By pooling resources the planner can overcome the problem of the regions' asymmetric liquidity needs. Let y and x denote the per capita amounts invested in the risk-free and risky assets, respectively. Furthermore, let c and cR_D denote the amounts creditors can withdraw to satisfy their liquidity needs at date $t = 1$ and date $t = 2$, respectively. In this context, R_D can be understood as the interest rate creditors earn by not withdrawing their funds for an additional period. The planner's problem can then be written as

$$\max_{x,y,c,R_D} U = \gamma u(c) + (1 - \gamma)\, \lambda u(cR_D)$$

subject to

$$x + y \leq n,\ \gamma\, 2nc \leq 2y,\ (1 - \gamma)\, 2ncR_D \leq 2xR,$$

$$x \geq 0,\ y \geq 0,\ c \geq 0,\ R_D \geq 0.$$

The first set of constraints represents budget constraints for periods 0, 1 and 2. Since optimality requires that the constraints be binding, the optimization problem can be rewritten as

$$\max_{y} \gamma u\left(\frac{y}{\gamma n}\right) + (1 - \gamma)\ \lambda u\left(\frac{R(n - y)}{(1 - \gamma)n}\right)$$

Given the utility function's properties this optimization problem has a unique interior solution. The optimal value $y^* \in (0,1)$ can be obtained from the first-order condition

$$u'\left(\frac{y^*}{\gamma n}\right) = \lambda R u'\left(\frac{R(n-y^*)}{(1-\gamma)n}\right)$$

Once y^* has been determined, we can use the remaining constraints to determine the optimal values of the other variables. Hence, we obtain

$$c^* = \frac{y^*}{\gamma n},\ R_D^* = \frac{R(n-y^*)}{(1-\gamma)nc^*},\ \text{and } x^* = n - y^*$$

Since $\lambda R > 1$, we can conclude that $u'(c) > u'(cR_D)$ and hence $R_D > 1$, implying that consumption is higher at $t = 2$ than at $t = 1$. Consequently, late creditors have no incentive to mimic early creditors. We denote the first-best allocation as $\delta^* = (y^*, x^*, c^*, R_D^*)$.

Decentralized Economy with an Interbank Market and No Bailout Possibility

Allen and Gale (2000) show that this first-best allocation can be achieved by allowing banks in a decentralized economy to co-insure against liquidity shocks. Co-insurance is possible since the liquidity needs of the two regions are negatively correlated. In contrast to Allen and Gale (2000), we again allow banks to exchange an arbitrary amount of deposits K, and not only the amount necessary to achieve first-best. However, we show that exchanging funds above the level of the first best solution does not increase the utility of uninsured creditors if there is no bailout possibility. Let k denote the amount of interbank deposits that is withdrawn by the bank that faces a high liquidity shock at date $t = 1$.

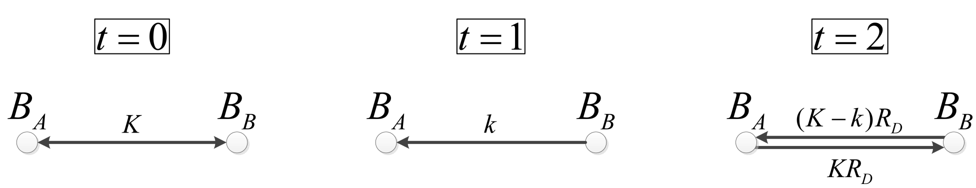


Figure 4: Capital flows in the two region economy

The capital flows are depicted in Figure 4. At $t = 0$ the two banks exchange deposits K. At $t = 1$ the bank with the high liquidity shock (B_A in Figure 4) withdraws an amount k from the other bank to satisfy the liquidity needs of its creditors. In the final period bank B_A receives its remaining deposits $(K-k)$ from bank B_B and pays back the deposits that bank B_B deposited in bank B_A. Additionally, both banks earn a rate of return R_D on these remaining deposits. Furthermore, we assume that contracts again take the form of a standard debt contract, that is, they cannot be made contingent on either the realization of the risky asset or the realization of the state of nature. Hence, each bank can offer a contract $\delta = (y, x, c, R_D, K)$ to its creditors and the bank in the other region. Now R_D additionally represents the gross return paid on interbank deposits held from $t = 1$ until $t = 2$. With perfect competition in the banking sector, the banks will offer their creditors a contract that replicates the first-best outcome. The optimization problem of a bank can then be written as

$$\max_{x,y,c,R_D,K,k} U = \frac{1}{2}[\omega_H u(c) + (1-\omega_H)\lambda u(cR_D)] + \frac{1}{2}[\omega_L u(c) + (1-\omega_L)\lambda u(cR_D)] \quad (19)$$

subject to

$$\omega_H nc + d_1 \leq y + k \quad (20)$$

$$\omega_L nc + d_1 + k \leq y \quad (21)$$

$$(1-\omega_H)ncR_D + d_2 + KR_D \leq Rx + (K-k)R_D \quad (22)$$

$$(1-\omega_L)ncR_D + d_2 + (K-k)R_D \leq Rx + KR_D \quad (23)$$

$$x \geq 0,\ y \geq 0,\ c \geq 0,\ R_D \geq 0,\ x+y \leq 1+e_0,\ E[d_1+d_2] \geq \lambda Re_0,\ k \leq K$$

Constraints (20) and (21) represent budget constraints at date $t = 1$ and constraints (22) and (23) represent budget constraints at date $t = 2$. As shown by Allen and Gale (2000), optimality requires that $k^* = (\omega_H - \gamma)cn$. As long as there is no positive bailout probability, the actual amount of funds exchanged, K, does not alter the utility of the creditors as long as $K \geq k^*$. This leads to the following proposition.

Proposition 6.3. *If there is no possibility for banks to be bailed out and the two representative banks exchange an amount K of deposits, then the first-best allocation δ^* can be implemented by a decentralized banking system offering standard deposit contracts. Moreover, banks have no incentive to exchange more funds than required to achieve first-best, that is, they will only exchange $k^* = (\omega_H - \gamma)cn$.*

Proof For the proof of the first part of the proposition we refer to the proof of Proposition 3 of Brusco and Castiglionesi (2007). To see why the second part holds true, that is, why banks do not exchange more than necessary to achieve first-best, note that optimality again requires the constraints to be binding. Then the amount of funds actually exchanged, K, drops out of the optimization problem. Hence, the amount that is actually exchanged does not influence the utility of the creditors. Therefore, banks have no incentive to exchange more funds than necessary to achieve first-best, which implies that $K = k^* = (\omega_H - \gamma)cn$. QED

This result reconfirms the findings of the previous articles by Allen and Gale (2000) and Brusco and Castiglionesi (2007).

Decentralized Economy with an Interbank Market and Positive Bailout Probability

So far we have assumed that after a bank failure occurs, creditors receive no repayment in period 2. Now we investigate how the results change if there is the possibility that a bank will be bailed out by the government after a default. As before, we assume this happens with probability α. Therefore, the optimization problem becomes

$$\begin{aligned}\max_{x,y,c,R_D,K,k} U &= \frac{1}{2}\left[\omega_H u(c) + (1-\omega_H)\left[\begin{array}{l}\lambda u(cR_D) + (1-\lambda)[(1-\alpha)^2 u(0)\\ +\alpha(1-\alpha)u(cR_D)\\ +\alpha(1-\alpha)u\left(cR_D\theta_1\right) + \alpha^2 u(cR_D)]\end{array}\right]\right]\\ &+ \frac{1}{2}\left[\omega_L u(c) + (1-\omega_L)\left[\begin{array}{l}\lambda u(cR_D) + (1-\lambda)[(1-\alpha)^2 u(0)\\ +\alpha(1-\alpha)u(cR_D)\\ +\alpha(1-\alpha)u\left(cR_D\theta_2\right) + \alpha^2 u(cR_D)]\end{array}\right]\right] \quad (24)\end{aligned}$$

with

$$\theta_1 = \frac{K-k}{(1-\omega_H)nc + K} \; and \; \theta_2 = \frac{K}{(1-\omega_L)nc + (K-k)}$$

subject to

$$\omega_H nc + d_1 \leq y + k \quad (25)$$

$$\omega_L nc + d_1 + k \leq y \quad (26)$$

$$(1-\omega_H)ncR_D + d_2 + KR_D \leq Rx + (K-k)R_D \quad (27)$$

$$(1-\omega_L)ncR_D + d_2 + (K-k)R_D \leq Rx + KR_D \quad (28)$$

$$x \geq 0,\ y \geq 0,\ c \geq 0,\ R_D \geq 0;\ x + y \leq 1 + e_0;\ E[d_1 + d_2] \geq \lambda Re_0;\ k \leq K$$

Equation (24) is the objective function of the optimization problem of the representative bank in region i. The bank in region i is equally likely to face a high or a low liquidity shock. If a high liquidity shock occurs in, for example, region A, a fraction ω_H of the creditors will withdraw their funds at $t = 1$ and the remaining creditors will demand repayment in $t = 2$. At $t = 2$ several cases must be considered. The risky asset yields a positive return R with probability λ and creditors receive their promised repayment cR_D. If the risky asset yields a zero payoff, the return of the creditor depends on whether the banks are bailed out or not. If neither of the two banks is bailed out, creditors receive no payment. If the bank in region A is bailed out, the government steps in and creditors receive their full repayment cR_D. If only the bank in region B is bailed out, bank B_A receives the funds still owed to it by B_B (see Figure 4). Since B_A has already withdrawn an amount k at date $t = 1$ it receives the remaining funds $(K-k)R_D$. Since B_A has two creditors, namely, its uninsured creditor and bank B_B, funds are again split on a pro rata basis. Hence, creditors receive a fraction θ_1 of their promised repayment. Finally, if both banks are bailed out, then creditors again receive the full amount. The second case (where B_A faces a low liquidity shock) can be described analogously.

All constraints are as in the previous section. By examining the optimization problem, it becomes obvious that the amount of funds exchanged, K, now has an influence on the utility of the creditors. Although K again drops out of the constraints (optimality again requires the constraints to be binding), it now also enters the objective function directly because it determines the amount that creditors receive in the case of a default if only one bank (here this would be bank B_B) is bailed out. Before the repayment in this state of nature was zero.

Again, optimality requires that banks choose first-best, that is, $k^{**} = (\omega_H - \gamma)cn$. Note, however, that the optimal consumption of creditors (c) changes. Compared to the case without bailout, creditors now consume less in period 1 and increase their consumption in period 2 (we formally show this in the Appendix). This also implies that the optimal amount of funds withdrawn in period 1 is now smaller than in the situation without bailout. Hence, we obtain the following first-order condition for K:

$$\begin{aligned}\frac{\partial U}{\partial K} &= \frac{1}{2}(1-\omega_H)(1-\lambda)\alpha(1-\alpha)c^2 nR_D\frac{(1-\gamma)}{(K+cn(1-\omega_H))^2}u'\left(cR_D\frac{K-cn(\omega_H-\gamma)}{K+cn(1-\omega_H)}\right)\\ &+ \frac{1}{2}(1-\omega_L)(1-\lambda)\alpha(1-\alpha)c^2 nR_D\frac{(1-\gamma)}{(K+cn(1-\gamma))^2}u'\left(cR_D\frac{K}{K+cn(1-\gamma)}\right)\\ &> 0 \end{aligned} \tag{29}$$

As we can see from the first-order condition, the utility of the creditor is now increasing in K (i.e., the funds exchanged in period 0), since K increases the amount that the creditor receives in case of default of the risky asset (although the amount needed to satisfy the consumption needs of creditors is now actually smaller, banks have an incentive to increase their interbank exposure). Therefore, banks have an incentive to increase the amount of interbank deposits and hence their connectivity to a level that exceeds the first-best solution derived before.

Proposition 6.4. *Given a positive bailout probability, banks have an incentive to increase their interbank exposure beyond the first-best level.*

Proof First note that the constraints are the same as in the previous section, where we excluded the possibility of a bailout. Again, optimality requires that the constraints be binding, which implies that K drops out of the constraints. Hence, we only have to examine the objective function. The results follow from the positive derivative of the creditors' utility function with respect to K. QED

Hence, even if the interbank market does not exist only as an insurance for non-insured creditors but also to co-insure against regional liquidity shocks, as in Allen and Gale (2000), the main mechanism is still present. Therefore, banks are still incentivized to increase their interbank exposure as long as they are not too big to save (given that there is a positive bailout probability).

7. Conclusion

This paper sheds light on the puzzle why banks have an incentive to be highly interconnected on the interbank market and why it can be rational to engage in circular lending activities, although this considerably increases systemic risk and leverage without altering the aggregate relation with the real economy. We show that banks create these cyclical liabilities because it enables them to make use of the implicit government bailout guarantees. Such guarantees shift the probability distribution of the returns of risky investments and thereby increase the expected repayment of uninsured creditors. Furthermore, the mechanism we derive in this paper is able to explain why banks choose correlated investments. Hence, the presented mechanism leads to an overall increase in systemic risk that results from both interconnectedness as well as herding behavior. Moreover, we show that interconnectedness incentivizes banks to engage in risk shifting. Therefore, our model helps explain why banks invested in risky correlated investments (e.g., US subprime loans) in the run-up to the financial crisis. Finally, we show that the optimal network structure depends on the amount of funds that is available to bail out banks in different countries. Our results continue to hold even if we allow creditors to be risk averse.

Several policy implications can be derived from our results. Generally, each of these policy implications aims at reducing the banks' incentive to create high interbank exposures by entering into cyclical liabilities. One of the key topics in the current discussion in the European Union is the introduction of a financial transaction tax in order to limit speculative trading activities. Since interconnectedness can not only be created via interbank loans, but also by using derivatives like e.g. CDS, such a tax may be a potential mechanism to reduce the high interconnectedness and therefore mitigate the systemic risk problems that result from investing in highly correlated low-quality assets. Similarly, one can think about increasing the risk weights for interbank loans under the Basel accord and thereby increase the amount of equity necessary to satisfy minimum capital requirements. Currently banks do not have to hold high amounts of capital for most of their interbank exposure. If interbank loans get a higher risk weight, it may incentivize banks to reduce their circular lending activities and hence reduce systemic risk in the interbank market. A third possibility to mitigate the incentives to create large cycle flows would be the introduction of the widely discussed bank levy. Charging banks with large balance sheets (that can very well result from high amounts of cyclical liabilities) higher taxes for their systemic risk can potentially mitigate the incentive to create these large cycle flows in the first place.

Appendix

Switching point K^* in Section 4.2

Here, we will formally derive the critical threshold of interbank deposits K^* (from Section 4.2) that just allows a successful bank to stay solvent if the bank it is connected to defaults and is not bailed out. The critical cases to derive this threshold are those where only one investment fails and neither of the banks is bailed out, i.e. S_9 and S_{13}. Here, the bank with the successful investment will pay the following amount to the bank with the failed investment:

$$\min\left\{KR_D, R\frac{K(c+K)}{c^2+2cK}\right\}$$

The first term represents the amount the successful bank owes to the failed bank and the second term results from:

$$\sum_{i=0}^{\infty} R\left(\frac{K}{c+K}\right)^{(1+2i)} = R\frac{K}{c+K}\frac{1}{1-\frac{K^2}{(c+K)^2}} = R\frac{K(c+K)}{c^2+2cK}$$

Hence, the failing bank receives either its full repayment (if there are enough funds available to settle all claims), i.e. $KR_D \leq R\frac{K(c+K)}{c^2+2cK}$ or receives a payment of $R\frac{K(c+K)}{c^2+2cK}$. The critical threshold up to which the bank receives its full repayment can be written as:

$$K_1^* R_D = R\frac{K_1^*(c+K_1^*)}{c^2+2cK_1^*} \Rightarrow K_1^* = \frac{c(R-cR_D)}{2cR_D-R} \tag{30}$$

>From (30) we can see that the successful bank can always pay back its liabilities to the unsuccessful bank as long as $R > 2cR_D$. Thus, it will never default in this case. In what follows we will focus on the more interesting case where a default is possible depending on the level of K. Hence, from now on we will assume that $R < 2cR_D$. We next consider the repayment the uninsured creditor gets from the successful bank. This is given by:

$$\min\left\{cR_D, R\frac{(c+K)}{c+2K}\right\}$$

where the fist term is the total amount owed to the uninsured creditor and the second term comes from:

$$\sum_{i=0}^{\infty} R\frac{c}{c+K}\left(\frac{K}{c+K}\right)^{2i} = R\frac{c}{c+K}\frac{1}{1-\frac{K^2}{(c+K)^2}} = R\frac{(c+K)}{c+2K}$$

Hence, as long as K is small enough such that $cR_D \leq R\frac{(c+K)}{c+2K}$ the successful bank can fully repay its uninsured creditor. However if K exceeds a critical threshold, the bank is unable to settle all its claims and can only repay $R\frac{(c+K)}{c+2K}$ to its creditor. The critical threshold is given by:

$$cR_D = R\frac{(c+K_2^*)}{c+2K_2^*} \Rightarrow K_2^* = \frac{c(R-cR_D)}{2cR_D - R} \tag{31}$$

As can be seen from (30) and (31), the thresholds K_1^* and K_2^* are the same. We now turn to the repayment of the uninsured creditor of the failed bank. This is given by:

$$\min\left\{cR_D, KR_D\frac{c}{c+K}, R\frac{K}{c+2K}\right\} = \min\left\{cR_D\frac{K}{c+K}, R\frac{K}{c+2K}\right\}$$

where the first term is again the total amount owed to the uninsured creditor, the second term is the maximal payment from the bank with the successful investment to the bank with the failed investment times the fraction the insured creditor gets from this payment, and the last term comes from:

$$\sum_{i=0}^{\infty} R\frac{c}{c+K}\left(\frac{K}{c+K}\right)^{(1+2i)} = R\frac{cK}{(c+K)^2}\frac{1}{1-\frac{K^2}{(c+K)^2}} = R\frac{K}{c+2K}$$

One can immediately see that the unsuccessful bank can never fully repay its uninsured creditors. Furthermore, as long as K is small enough such that $cR_D\frac{K}{c+K} \leq R\frac{K}{c+2K}$, the payment of the unsuccessful bank to its uninsured creditors is $cR_D\frac{K}{c+K}$. If K is too high, the payment is $R\frac{K}{c+2K}$. The critical threshold where this switches is given by

$$cR_D\frac{K_3^*}{c+K_3^*} \leq R\frac{K_3^*}{c+2K_3^*} \Rightarrow K_3^* = \frac{c(R-cR_D)}{2cR_D - R} \tag{32}$$

Hence, all three thresholds are the same, which is why we will denote them in the following by

$$K^* \equiv K_1^* = K_2^* = K_3^*. \tag{33}$$

Therefore, if a specific bank has a successful investment, it is able to settle all its liabilities, even if the other bank fails, as long as its interbank exposure is $K \leq K^*$. This completes the derivation of K^*.

Proof of Corollary 4.2

We now we have to check whether the expected utility for the uninsured creditor is maximized by choosing $K \leq K^*$ or by choosing $K > K^*$. For the interval $K \in [0, K^*]$ we know that:

$$U_0(K \leq K^*) = \lambda cR + (1-\lambda)\alpha cR$$

Therefore, the expected utility of non-insured creditors does not depend on the interbank exposure K. This makes the bank indifferent with regard to the choice of K. For the interval $K \in [K, \overline{K_0}]$ with $\overline{K_0} = \frac{\overline{L}}{R_D^2} - c$ we know that:

$$\begin{aligned}
U_0(K = K^*) &= \lambda cR + (1-\lambda)\alpha cR \\
\frac{\partial U_0}{\partial K}(K^* \leq K \leq \overline{K_0}) &= R\frac{\alpha(1-\alpha)(1-\lambda)c[\lambda + (1-\lambda)\alpha - (1-c)]}{(c+K)^2(\lambda + (1-\lambda)\alpha)} > 0
\end{aligned}$$

Hence, if $(c+K^*)R_D^1 < \overline{L}$, the bank will increase the interbank exposure K until $K = \overline{K_0}$. As soon as this threshold is hit the bailout probability α drops to zero and the expected utility for the uninsured creditors decreases to $\lambda^2 cR_D^2 + \lambda(1-\lambda)R$. If, on the other hand, $(c+K^*)R_D^1 \geq \overline{L}$, the bank will be indifferent about the choice of K in the interval $K = [0, \overline{K_0}]$. Therefore, if $(c+K^*)R_D^1 < \overline{L}$, the bank chooses $K = \overline{K_0}$ in order to maximize the expected utility of its uninsured creditor:

$$\begin{aligned}
\overline{U_0} &= \left[\alpha(1+\lambda) + \lambda^2(1-2\alpha) - \alpha^2\lambda(1-\lambda)\right] cR_D^2 \\
&+ \lambda(1-\lambda)(1-\alpha)^2 R + \alpha(1-\lambda)(1-\alpha)cR_D^2\frac{\overline{K_0}}{c+\overline{K_0}}
\end{aligned}$$

In case $(c+K^*)R_D^1 \geq \overline{L}$ the maximal expected utility of its uninsured creditor becomes:

$$\overline{U_0} = \lambda cR + (1-\lambda)\alpha cR$$

This completes the derivation of the expected utility of uninsured creditors in the case of a correlation of zero and the proof of Corollary 4.2.

Proof of Proposition 4.3

To determine whether banks prefer correlated investments, we compare the utility of the uninsured creditors for both types of investment correlations (i.e. a correlation of one and zero) and for the latter case the situations where $(c+K^*)R_D^1 < \overline{L}$ and $(c+K^*)R_D^1 \geq \overline{L}$. First, we consider the case that $(c+K^*)R_D^1 < \overline{L}$:

$$\begin{aligned}
\overline{U_1} &> \overline{U_0} \\
\lambda cR + (1-\lambda)\left[\alpha cR + \alpha(1-\alpha)\overline{L}\frac{c\overline{K_1}}{(c+\overline{K_1})^2}\right] &> \left[\begin{array}{l}\left[\alpha(1+\lambda) + \lambda^2(1-2\alpha) - \alpha^2\lambda(1-\lambda)\right] cR_D^2 \\ +\lambda(1-\lambda)(1-\alpha)^2 R + \alpha(1-\lambda)(1-\alpha)\overline{L}\frac{c\overline{K_0}}{(c+\overline{K_0})^2}\end{array}\right]
\end{aligned}$$

After inserting the expression in equation (5) for R_D^2, we can simplify the right hand side and the inequality becomes:

$$\begin{aligned}
\lambda cR + (1-\lambda)\left[\alpha cR + \alpha(1-\alpha)\overline{L}\frac{c\overline{K_1}}{(c+\overline{K_1})^2}\right] &> \left[\begin{array}{l}\frac{\alpha[\alpha-(1-c)]+\alpha\lambda(1-c)(1-\alpha)+\lambda^2 c[1-\alpha(2-\alpha)]}{\alpha+\lambda(1-\alpha)}R \\ +\alpha(1-\lambda)(1-\alpha)\overline{L}\frac{c\overline{K_0}}{(c+\overline{K_0})^2}\end{array}\right] \\
(1-\lambda)\alpha(1-\alpha)\overline{L}\frac{c\overline{K_1}}{(c+\overline{K_1})^2} &> \alpha(1-\lambda)(1-\alpha)\left[\overline{L}\frac{c\overline{K_0}}{(c+\overline{K_0})^2} - R\frac{(1-c)}{\alpha+\lambda(1-\alpha)}\right] \\
R\frac{1-c}{\alpha+\lambda(1-\alpha)} + \overline{L}\frac{c\overline{K_1}}{(c+\overline{K_1})^2} &> \overline{L}\frac{c\overline{K_0}}{(c+\overline{K_0})^2}
\end{aligned}$$

Since the first term on the left hand side is positive and $\frac{cK}{(c+K)^2}$ is decreasing in K as well as $\overline{K_0} > \overline{K_1}$, it follows that $\overline{U_1} > \overline{U_0}$. Next, we consider the case that $(c+K^*)R_D^1 \geq \overline{L}$:

$$\begin{aligned}
\overline{U_1} &> \overline{U_0} \\
\lambda cR + (1-\lambda)\left[\alpha cR + \alpha(1-\alpha)\overline{L}\frac{c\overline{K_1}}{(c+\overline{K_1})^2}\right] &> \lambda cR + (1-\lambda)\alpha cR \\
(1-\lambda)\alpha(1-\alpha)\overline{L}\frac{c\overline{K_1}}{(c+\overline{K_1})^2} &> 0
\end{aligned}$$

Hence, $\overline{U_1}$ is always larger than $\overline{U_0}$, irrespective of whether $(c+K^*)R_D^1 < \overline{L}$ or $(c+K^*)R_D^1 \geq \overline{L}$. Therefore, the bank always chooses $\rho = 1$. This completes the proof.

Proof of Proposition 5.1

In the following we compare the critical bailout probabilities for the case without (α^*) and with interbank network (α^{**}). By plugging in the critical values derived in Section 5, one can see that:

$$\begin{aligned}
\alpha^* &> \alpha^{**} \\
\frac{1-\lambda_N R_D^N}{(1-\lambda_N)R_D^N} &> \frac{c+2K}{2K} - \sqrt{\frac{(c+2K)^2}{4K^2} - \frac{(c+K)(R_D^N\lambda_N - 1)}{KR_D^N(\lambda_N - 1)}} \\
\sqrt{\frac{(c+2K)^2}{4K^2} - \frac{(c+K)(R_D^N\lambda_N - 1)}{KR_D^N(\lambda_N - 1)}} &> \frac{c+2K}{2K} - \frac{1-\lambda_N R_D^N}{(1-\lambda_N)R_D^N} \\
\frac{(c+2K)^2}{4K^2} - \frac{(c+K)(R_D^N\lambda_N - 1)}{KR_D^N(\lambda_N - 1)} &> \left(\frac{c+2K}{2K} - \frac{1-\lambda_N R_D^N}{(1-\lambda_N)R_D^N}\right)^2 \\
\frac{(c+2K)(1-\lambda_N R_D^N)}{K(1-\lambda_N)R_D^N} - \frac{(c+K)(R_D^N\lambda_N - 1)}{KR_D^N(\lambda_N - 1)} &> \left(\frac{1-\lambda_N R_D^N}{(1-\lambda_N)R_D^N}\right)^2 \\
\frac{K(1-\lambda_N R_D^N)}{K(1-\lambda_N)R_D^N} &> \left(\frac{1-\lambda_N R_D^N}{(1-\lambda_N)R_D^N}\right)^2 \\
R_D^N - \lambda_N R_D^N &> 1 - \lambda_N R_D^N \\
R_D^N &> 1
\end{aligned}$$

This last inequality is always true. This completes the proof.

Proof of Proposition 6.1

In order to show that $\overline{U_{DI}} > \overline{U_{BI}}$ holds, it is sufficient to compare the respective cash flows in case the investments fail, since the success state is equal for both cases. Hence, we have to show that

$$\begin{bmatrix} \alpha_B cR + (1-\alpha_B)\alpha_B cR\frac{\overline{K_1}}{c+\overline{K_1}} \\ +(1-\alpha_B)^2\alpha_B cR\frac{\overline{K_1}^2}{(c+\overline{K_1})^2} \end{bmatrix} > \begin{bmatrix} \alpha_B cR + (1-\alpha_B)\alpha_B^2 cR\frac{2\overline{K_{BI}}}{c+2\overline{K_{BI}}} \\ +2(1-\alpha_B)^2\alpha_B cR\frac{\overline{K_{BI}}}{c\overline{K_{BI}}} \end{bmatrix}$$

After subtracting $\alpha_B cR$ and canceling out $(1-\alpha_B)\alpha_B$ the inequality becomes

$$\frac{\overline{K_1}}{c+\overline{K_1}} + (1-\alpha_B)\frac{\overline{K_1}^2}{(c+\overline{K_1})^2} > \alpha_B\frac{2\overline{K_{BI}}}{c+2\overline{K_{BI}}} + 2(1-\alpha_B)\frac{\overline{K_{BI}}}{c\overline{K_{BI}}}$$

Then we use the information that $\overline{K_{BI}} = \frac{1}{2}\overline{K_1}$ to get

$$\begin{aligned}
\frac{\overline{K_1}}{c+\overline{K_1}} + (1-\alpha_B)\frac{\overline{K_1}^2}{(c+\overline{K_1})^2} &> \alpha_B\frac{\overline{K_1}}{c+\overline{K_1}} + (1-\alpha_B)\frac{\overline{K_1}}{c+\frac{1}{2}\overline{K_1}} \\
\frac{\overline{K_1}^2 c(1-\alpha)}{(c+\overline{K_1})^2(2c+\overline{K_1})} &> 0
\end{aligned}$$

Since in the last line all terms on the left hand side are always positive, it holds that $\overline{U_{DI}} > \overline{U_{BI}}$. This completes the proof.

Discussion of optimal consumption with risk-averse creditors and positive bailout probability

To understand why the optimal consumption decreases in $t = 1$ if a bailout is possible first note that a bailout simply changes the probability distribution of the investment. Without bailout creditors receive funds for consumption only with probability λ in period 2. Now if the investment fails there is still a positive probability that creditors receive (at least parts of) their funds. To fully capture the optimal consumption decision we look at a situation where the investment returns and respective probabilities match exactly those of the risky investment considered in the paper when there is a positive bailout probability.

$$\max_{x,y,c,R_D} U = \gamma u(c) + (1-\gamma)\left[\lambda u(cR_D) + (1-\lambda)[\alpha u(cR_D) + \alpha(1-\alpha)u(\theta c R_D)]\right] \tag{34}$$

subject to

$$x + y \leq n,\ \gamma\, 2nc \leq 2y,\ (1-\gamma)\, 2ncR_D \leq 2xR,$$
$$x \geq 0,\ y \geq 0,\ c \geq 0,\ R_D \geq 0.$$

Since the constraints in the respective periods again have to be binding we can solve them for c and R_D, respectively and can plug these values into the objective function. This yields:

$$\max_y U = \gamma u\left(\frac{y}{\gamma n}\right) + (1-\gamma)\left[\begin{array}{l} \lambda u\left(\frac{R(n-y)}{(1-\gamma)n}\right) \\ +(1-\lambda)\left[\alpha u\left(\frac{R(n-y)}{(1-\gamma)n}\right) + \alpha(1-\alpha)u\left(\theta\frac{R(n-y)}{(1-\gamma)n}\right)\right] \end{array}\right]$$

The first order condition with respect to y then yields:

$$u'\left(\frac{y}{\gamma n}\right) = u'\left(\frac{R(n-y)}{(1-\gamma)n}\right)[\lambda R + (1-\lambda\alpha R)] + (1-\lambda)\theta R\alpha u\left(\theta\frac{R(n-y)}{(1-\gamma)n}\right) \tag{35}$$

Looking at this first order condition one can see that the marginal utility of consumption in period 1 is higher now, implying that consumption is lower. Hence, if it is more likely to get the higher repayment at $t = 2$ creditors want to shift more consumption to this later period. This completes the discussion of the optimal consumption allocation with risk-averse creditors and positive bailout probability.

References

[1] Acharya, V. (2009): A theory of systemic risk and design of prudential bank regulation, Journal of Financial Stability, 5(3), 224-255

[2] Acharya, V., I. Drechsler and P. Schnabl (2011): A Pyrrhic Victory? - Bank Bailouts and Sovereign Credit Risk, Working Paper, available at SSRN

[3] Acharya, V. V. and T. Yorulmazer (2007): Too many to fail - An analysis of time-inconsistency in bank closure policies, Journal of Financial Intermediation, 16, 1-31

[4] Adrian, T. and H. S. Shin (2011): Financial Intermediary Balance Sheet Management, Working Paper, available at SSRN

[5] Allen, F., and D. Gale (2000): Financial Contagion, Journal of Political Economy, 108(1), 1-33

[6] Allen, F., and D. Gale (2005): Systemic Risk and Regulation, in M. Carey and R. M. Stulz (eds.), The Risks of Financial Institutions, 341-375.

[7] Allen, F., E. Carletti and D. Gale (2009). Interbank Market Liquidity and Central Bank Intervention, Journal of Monetary Economics, 56(5), 639-652 Journal of Monetary Economics

[8] Bhattacharya, S. and D. Gale (1987): Preference Shocks, Liquidity and Central Bank Policy, in W. Barnett and K. Singleton (eds.), New Approaches to Monetary Economics, Cambridge: Cambridge University Press, 69-88

[9] Bank for International Settlements (2011): Detailed tables on preliminary locational and consolidated banking statistics at end-June 2011, Basel, Switzerland

[10] Brusco S. and F. Castiglionesi (2007), Liquidity Coinsurance, Moral Hazard and Financial Contagion, Journal of Finance, 62, 2275-2302

[11] Castiglionesi, F., F. Feriozzi and G. Lóránth (2011): Liquidity Coinsurance and Bank Capital, Working Paper, available at SSRN

[12] Castiglionesi, F. and N. Navarro (2010): Optimal Fragile Financial Networks, Working Paper, available at SSRN

[13] Cukierman, Y. and Y. Izhakian (2011): Bailout Uncertainty in a Microfounded General Equilibrium Model of the Financial System, Working Paper, available at SSRN

[14] David, A. and A. Lehar (2011): Why are Banks Highly Interconnected?, Working Paper, available at SSRN

[15] Diamond, D. and P. Dybvig (1983): Bunk runs, deposit insurance and liquidity, Journal of Political Economy, 91, 401419

[16] Diamond, D.W. and R. G. Rajan (2002): Bank Bailouts and Aggregate Liquidity, American Ecnomic Review (Papers and Proceedings), 92(2), 38-41

[17] Fitch Ratings Special Report (2006): One Year After the Abolition of the Landesbank State Guarantee Uplift for Short-Term Ratings Still Justified, London, United Kingdom

[18] Freixas, X. and C. Holthausen (2005): Interbank Market Integration under Asymmetric Information, Review of Financial Studies, 18, 459-490

[19] Freixas, X., B. Parigi, and J. Rochet (2000). Systemic Risk, Interbank Relations and Liquidity Provision by the Central Bank, Journal of Money, Credit & Banking, 32, 611-38

[20] Gorton, G. and L. Huang (2004): Liquidity, Efficency and Bank Bailouts, American Economic Review, 94(3), 455-483

[21] Hale, G. (2011): Bank Relationships, Business Cycles, and Financial Crises, Working Paper

[22] Kubelec, C. and F. Sá (2010): The geographical composition of national external balance sheets: 19802005, Bank of England Working Paper No. 384

[23] Leitner, Y. (2005): Financial Networks: Contagion, Commitment, and Private Sector Bailouts, Journal of Finance, 60, 2925-2953

[24] May, R., S. Levin and G. Sugihara (2008): Ecology for bankers, nature, 451, 893-895

[25] Merton, R. C. (1977): An Analytic Derivation of the Cost of Deposit Insurance and Loan Guarantees, Journal of Baking and Finance, 1, 3-11

[26] Minoiu, C. and J. A. Reyes (2011): A Network Analysis of Global Banking: 1978 2010, Working Paper, available at SSRN

[27] Mueller, J. (2006): Interbank Credit Lines as a Channel of Contagion, Journal of Financial Services Research 29(1), 37-60

[28] Rotemberg, J. (2011): Minimal Settlement Assets in Economies with Interconnected Financial Obligations, Journal of Money, Credit and Banking, 43(1), 81-108

[29] Shin, H. S. (2009): Securitisation and Financial Stability, The Economic Journal, 119, 309332

[30] Soramäki, K., M.L. Bech, J. Arnold, R.J. Glass and W. E. Beyeler (2006): The Topology of Interbank Payment Flows, Federal Reserve Bank of New York Staff Reports, No. 243

[31] Takács, L. (1988): On the Limit Distribution of the Number of Cycles in a Random Graph, Journal of Applied Probability, 25, 359-376

[32] Wells, S. (2004): U.K. Interbank Exposures: Systemic Risk Implications, Journal of Monetary Economics, 2(1), 66-77.