

Collateral-Motivated Financial Innovation*

Ji Shen
London School of Economics

Hongjun Yan
Yale School of Management

Jinfan Zhang
Yale School of Management

August 8, 2012

*We thank Darrell Duffie, John Geanakoplos, Robert Marquez, Zhiguo He, Arvind Krishnamurthy, Andrew Metrick, Konstantin Milbradt, Adriano Rampini, Wei Xiong, Dimitri Vayanos, and seminar participants at LSE, MIT, NBER, and Yale, as well as Econometric Society Meetings, University of Minnesota Finance Conference, WFA for helpful comments. Please direct all correspondence to Hongjun Yan, Email: hongjun.yan@yale.edu. The latest version of the paper is available at <http://faculty.som.yale.edu/hongjunyan/>.

Collateral-Motivated Financial Innovation

Abstract

The paper proposes a collateral view of financial innovation: Many innovations are partly motivated by alleviating collateral/margin constraints for trading (speculation or hedging). We analyze a model of investors with disagreement. The trading need motivates investors to introduce derivatives, which are endogenously determined in equilibrium. In the presence of a collateral friction in cross-netting, the “optimal” security is the one that isolates the variable with disagreement. It is optimal in the sense that alternative derivatives cannot generate any trading. With an arbitrarily small trading cost, the optimal security is “unfunded”—i.e., has a zero initial value. The endogenous difference in collateral requirements across assets leads to a basis—a spread between the prices of an asset and its replicating portfolio. This basis reflects the shadow value of collateral, leading to a number of novel predictions. Our model also shows that financial innovations diminish the impact of policies that target spot markets only (e.g., Regulation T, Tobin tax). More broadly, our analysis highlights the common theme behind a variety of innovations with strikingly different appearances: the invention of securities (e.g., futures, swaps), legal practice (e.g., the superseniority of repos and derivatives), legal entities (e.g., special purpose vehicles), as well as the efforts in improving cross-netting.

JEL Classification Numbers: G11, G23.

Keywords: Financial Innovation, Collateral, Margin, Leverage, Asset Pricing, Derivative, Basis.

1 Introduction

The past half century has witnessed a tremendous amount of financial innovations. What are the motives behind them? Existing theories emphasize the role of risk sharing (e.g., Allen and Gale (1994)), transaction costs and regulatory constraints (Benston and Smith (1976), Miller (1986)), and information asymmetry (Gorton and Pennacchi (1990), DeMarzo and Duffie (1999)).

This paper proposes an alternative view: Many successful financial innovations are partly motivated by mitigating collateral (or margin) constraints for trading. Suppose, for example, two traders have different expectations for the future value of a security—say, a corporate bond. If their disagreement is about the company’s default probability, rather than the future movements of riskless interest rates, then it is natural that the traders prefer to take positions in credit default swaps (CDSs), rather than the corporate bond. This is because by isolating the default probability, the variable that traders are interested in betting on, a CDS on the underlying corporation requires least collateral and is efficient in facilitating their speculation.¹ This collateral motivation is not limited to speculative trading: Suppose, for instance, a risk manager of a corporation needs to hedge a certain exposure, and can trade two financial instruments with the same hedging quality. To the extent that raising capital is costly, the risk manager clearly has a preference for the instrument with a lower collateral requirement.

Motivated by the above intuition, we analyze an equilibrium model of investors with disagreement about a portion of the cash flow from an asset. The disagreement motivates investors to trade this asset, and possibly to introduce new derivatives to facilitate their trading. Casual intuition suggests that investors would introduce derivatives that are linked to the disagreement. However, the impact of this innovation on other markets is far less clear. Would investors try to complete the markets? Which markets would thrive, and which would disappear? What is the notion of “optimal” innovation in this context?

To understand these issues, first consider a benchmark case without collateral frictions. In this case, if an investor defaults on his promise (e.g., debt or a short position in an Arrow security), his counterparty can seize the collateral the investor has posted for the trade and the defaulting investor faces no further penalty.² The benchmark case features a frictionless collateralization procedure, in which an investor can use (any part of) his overall portfolio as collateral. For convenience, we refer to this as *portfolio margin*. It is easy to see that the collateral constraint under portfolio margin is equivalent to a nonnegative wealth constraint. Moreover, if investors introduce financial assets to complete markets, the resulting equilibrium is Pareto optimal. This benchmark case highlights the benefit of market completeness but does not have sharp predictions on financial innovation.

Our main analysis is focused on a collateral friction. In particular, we note that portfolio margin

¹A vivid example is documented in Michael Lewis’s book *Big Short*. During 2004–2006, a number of investors were convinced that the subprime mortgage market would soon collapse, and wanted to bet on it. However, they found that existing instruments (e.g., the stocks of home building companies) can offer only an “indirect” bet. The book tells a detailed story of how those investors push investment banks to create the market of CDS contracts on subprime mortgage bonds, which provides a more “direct” bet on the subprime mortgage market.

²This lack of further penalty assumption is perhaps most suitable for the case of security trading, where many positions are set up for hedging, speculation, or short-term financing purposes. When an investor defaults, the top priority for his counterparties is perhaps to get compensated quickly to reestablish those positions with other investors, rather than going through a lengthy bankruptcy procedure to liquidate the defaulting investor’s other assets.

is often impractical. For example, if an investor holds a portfolio in which individual asset returns offset each other, under portfolio margin, the collateral requirement for the whole portfolio can be much lower than that for one asset alone. In practice, however, this cross-asset netting is far from perfect. For instance, if one asset in the portfolio is exchange-traded while the other is over-the-counter or traded at a different exchange, then the investor has to post collateral for both assets separately, even if these two positions largely offset each other.³ Moreover, it may be difficult or too costly for a dealer to precisely estimate the correlation among securities to determine the collateral for the whole portfolio. Or, a trader may prefer not to reveal his whole portfolio to his dealer and thus has multiple dealers, which is a common practice among hedge funds. Finally, different parts of the portfolio may be governed by different jurisdictions, and regulations may also impose various constraints on collateralization, making cross-netting imperfect.

Our key assumption, motivated by these frictions, is that investors have to post collateral for each security in their portfolios separately, which we refer to as *individual security margin*. The essential point of this assumption is that cross-asset netting is imperfect rather than impossible. Our model’s main implications are the following.

First, this collateral friction determines the financial innovation in equilibrium. Intuitively, due to the collateral friction, investors prefer to trade a derivative that isolates the portion of the cash flow with disagreement, rather than trading the underlying asset. This is because the cash flow from the underlying asset has two portions, but investors are only interested in trading one. To the extent that the “unwanted” portion, the portion without disagreement, increases the collateral requirement for trading the underlying asset, it makes the underlying asset less appealing than the derivative. Consequently, the derivative that completely carves out the unwanted cash flow is “optimal” in the sense that its existence would drive out any other derivative markets: if one introduced any other derivatives, those markets would not generate any trading. It is important to note the difference between our result and the standard intuition of completing markets. We demonstrate that in a world with N states, investors may choose to introduce more than N securities but the markets are still incomplete—i.e., the securities do not span the whole state space.

Second, the optimal derivative tends to be “unfunded”—i.e., the initial value of the derivative is designed to be zero. The reason is that for a security that facilitates speculation or hedging, its essential role is to transfer wealth across states of nature rather than across time. Hence, if its price is not zero, one party gets paid initially, but there is a chance for him to pay this amount back in the future. Making the security unfunded avoids this potential “round trip” in wealth transfer. To the extent there is an infinitesimal cost of transferring funds, the unfunded security would be strictly preferred. This perhaps explains why many derivatives, such as futures and swaps, are designed to be unfunded.

Third, due to the high collateral requirement of the underlying asset, its price can be lower than the price of its replicating portfolio. This is consistent with the empirical evidence on the so-called corporate bond–CDS basis: the price of a corporate bond is often lower than the price of the portfolio of a CDS and a Treasury bond that replicates the corporate bond’s cash flow (e.g., Mitchell

³One famous example is that Metallgesellschaft AG, a German conglomerate, had a large short position in oil forward contracts and an offsetting long position in oil futures in the early 1990s, but eventually ran into a liquidity crisis when the collateral requirement became excessive (see Culp and Miller (1995)).

and Pulvoni (2011), Garleanu and Pedersen (2011)). More recently, Fleckenstein, Longstaff, and Lustig (2010) find that the prices of Treasury Inflation-Protected Securities (TIPS) are lower than that of their replicating portfolios that consist of inflation swaps and nominal Treasury bonds. These phenomena are consistent with our model: It requires more collateral to take a long position in the underlying assets (i.e., corporate bonds or TIPS). In contrast, to take an equivalent long position through those unfunded derivatives (i.e., CDS and inflation swaps), one needs only to post collateral to cover the daily movements in the mark-to-the-market value, and so the collateral requirement is much smaller. This difference in collateral requirement makes the derivatives more appealing and leads to a basis.

Fourth, the above intuition implies that the basis reflects the shadow value of collateral, leading to a number of time-series and cross-sectional implications on the basis. For example, when investors face tighter funding constraints, saving collateral becomes more valuable, leading to a larger basis. Our model also implies that the basis is higher if the unwanted cash flow, the portion of the cash flow without disagreement, is more volatile. This is because the volatility of the unwanted cash flow determines the amount of collateral that can be saved by trading the derivative. Moreover, when investors' funding liquidity dries up, the basis increases more for assets with more volatile unwanted cash flows. These implications are consistent with the empirical evidence in Mitchell and Pulvoni (2011) and Fleckenstein, Longstaff, and Lustig (2010) and are parallel to the implications in Garleanu and Pedersen (2011), where the collateral requirements are exogenous. In contrast, our model endogenizes both financial innovation and the collateral requirements. Moreover, our model also has a number of novel predictions. It implies that the basis should increase with the disagreement among investors, because a larger disagreement implies a stronger speculation motive and so a higher shadow value of collateral. Another implication is that the basis increases upon a positive supply shock to the underlying asset (e.g., a failing institution selling a large amount of the underlying asset), and this impact is stronger for assets with more volatile unwanted cash flows.

Fifth, we conduct a "policy experiment" in our model to evaluate the impact of margin policy. The Securities and Exchange Act of 1934 empowers the Federal Reserve Board (Fed) to set the margin requirements for trading stocks and bonds. The main goal of this policy is to contain the impact of speculation on financial markets. Although the Fed made 23 adjustments on stock margin requirements during the first 40 years, it has been inactive since 1974. A large number of empirical studies examine the impact of this margin policy. While earlier studies often drew opposite conclusions, many recent ones generally conclude that the change in margin requirements does not have a observable impact on the stock market volatility.⁴ To evaluate this policy, we impose a margin requirement on the underlying asset in our model. We find that when this margin policy gets tighter (i.e., investors can borrow less with the underlying asset as collateral), the wealth fluctuation caused by speculation generally reduces. However, financial innovations can significantly diminish the impact of this policy. When the margin policy becomes tighter, investors can no longer take large speculative positions in the underlying asset, but they can set up similar positions in derivative markets. It is interesting to note the coincidence that the Fed stopped actively adjusting the margin policy in 1974 and the derivative markets started taking off at around that time. One possibility is that, as our model suggests, the existence of active derivative markets makes the

⁴See Alexander et al. (2004) for a recent review.

margin policy less effective. Indeed, conversations with practitioners appear to suggest that active investors such as hedge funds barely view this margin policy as a meaningful constraint. More generally, this analysis suggests that financial innovations diminish the impact of policies that target spot markets only. Tobin tax is another example. In 1972, James Tobin proposed the idea of taxing transactions in currency markets to contain speculations. Clearly, this transaction tax makes speculations in the spot market less appealing. However, the impact of this tax on speculation is likely to be limited since investors can move their positions to derivative markets, where this tax is less effective because derivatives are usually, as suggested by our model, unfunded and so the tax is zero if it is levied as a fraction of transaction prices.

Finally, and more broadly, this collateral view of financial innovation highlights the common theme behind a variety of innovations, such as the invention of new securities, legal practice, legal entities, as well as policy changes. As noted earlier, many derivatives, such as swaps and futures, allow investors to take large positions with very little collateral. Another example of collateral-motivated innovation is the emerging legal practice of the so-called superseniority of derivatives and repos. When a firm goes bankrupt, its repo and derivative counterparties can simply seize the collateral posted in the transactions up to the amount the firm owes them, instead of going through a lengthy and costly bankruptcy procedure.⁵ In the context of our analysis, this practice can also be viewed as carving unwanted cash flows out of derivative and repo transactions: Suppose an investor enters an interest rate swap to hedge or speculate on interest rate risk. Without superseniority, even if his counterparty posts a large amount of collateral, the investor is still not well protected since he would have to go through automatic stay when his counterparty defaults. With superseniority, however, the investor can immediately seize the collateral upon default, and so can be better protected even by a smaller amount of collateral. From this investor's perspective, his counterparty's assets, apart from the posted collateral, are unwanted cash flows, which are carved out of the swap transaction by superseniority. Similarly, financial innovations may take the form of legal entities. For example, special purpose vehicles (SPVs), have become prevalent in recent decades with the rise of securitization. Again, in the context of our model, we can view creating an SPV as carving out unwanted cash flows (i.e., the SPV sponsor's assets other than those allocated to the SPV). Note that the collateral friction in our model arises from the limitations on cross-netting in posting collateral. In practice, it is becoming increasingly possible to have more cross-netting. For example, on December 12, 2006, the Securities and Exchange Commission (SEC) approved a rule change that made partial cross-netting available to some investors in the exchange-traded options market.⁶ There have also been efforts from brokers and hedge funds that attempt to get around the regulation-induced margin requirements (see, e.g., Brunnermeier and Pedersen (2009)). One can view these continuing efforts by regulators and market participants in modifying the margin procedure as a form of collateral-motivated financial innovation. Their goal is simply to satisfy the demand from market participants to alleviate their collateral constraints.

⁵This exceptional treatment accorded derivatives and repos in bankruptcy is recent. It was formalized by the introduction of "Act to Amend Title 11, United States Code, to Correct Technical Errors, and to Clarify and Make Substantive Changes, with Respect to Securities and Commodities" to the bankruptcy code as of July 27, 1982 (Pub. L. 97-222 (HR 4935)). There have been numerous revisions over the years. A recent example is the Financial Netting Improvements Act of 2006 (Pub. L. 109-390).

⁶For more details see the *Customer Portfolio Margin User Guide*, available from the website of The Options Clearing Corporation, www.optionsclearing.com.

2 Literature Review

There is an extensive literature on financial innovation. Recent surveys, such as Allen and Gale (1994) and Duffie and Rahi (1995), emphasize the risk-sharing role of financial instruments, Tufano (2003) also discusses the roles of regulatory constraints, agency concerns, transaction costs and technology (Ross (1989), Benston and Smith (1976), Merton (1989), White (2000)). More recent studies explore the role of rent seeking (Biais, Rochet, and Woolley (2010)) and neglected risk (Gennaioli, Shleifer, and Vishny (2011)). These studies generally abstract away from collateral constraint, which is the focus of this paper. One exception is Santos and Scheinkman (2001), which analyzes a model where exchanges set margin levels to screen traders with different credit qualities. Also related are the studies that analyze the impact of financial innovation in models with heterogeneous beliefs or preferences (Zapatero (1998), Bhamra and Uppal (2009), Simsek (2012), Banerjee and Graveline (2011)). These studies focus on the insight that innovation leads to more speculation and higher volatility, while our analysis focuses on the collateral friction and its implications on endogenous financial innovation and asset prices.

The role of collateral has been analyzed in various contexts, such as the macroeconomy (e.g., Kiyotaki and Moore (1997)), corporate debt capacity (e.g., Rampini and Viswanathan (2010)), arbitrageurs' portfolio choices (e.g., Liu and Longstaff (2004)), asset prices and welfare (Basak and Croitoru (2000), Gromb and Vayanos (2002)). Our analysis of collateral requirement builds on Geanakoplos (1997, 2003), which has been extended to study leverage cycle (Fostel and Geanakoplos (2008), Geanakoplos (2009)), speculative bubble (Simsek (2011)), and debt maturity (He and Xiong (2010)). Our analysis on leverage is also related to the studies of financial products that help constrained investors to take leverage (Frazzini and Pedersen (2011), Jiang and Yan (2012)). Finally, our paper is related to Garleanu and Pedersen (2011), who analyze the impact of collateral on the violation of the law of one price. One major difference is that the endogenous financial market structure and collateral requirements are the focus of our paper, but are exogenously given in their study. Our model also has a number of novel predictions—e.g., the impact of disagreement and supply shocks on basis.

The rest of the paper is as follows. Section 3 presents the model and its equilibrium. Section 4 analyzes the violation of the law of one price. Some further analysis of the model is presented in Section 5. Section 6 studies the impact of financial innovation on the economy. Some extensions of the model are analyzed in Section 7. Section 8 provides some general discussions, and Section 9 concludes. All proofs are summarized in the Appendix.

3 A Model of Financial Innovation

We consider a two-period economy, $t = 0, 1$, that is populated by a continuum of investors. The total population is normalized to 1. Investors make investment decisions at $t = 0$ and consume all their wealth at $t = 1$. All investors are risk neutral, and their objective is to maximize their expected consumption at $t = 1$. There is a riskless storage technology with a return of 0. All investors have the same endowment, and the aggregate endowment is e ($e \geq 0$) dollars in cash and β ($\beta \geq 0$) units of asset A , which is a claim to a random cash flow \tilde{A} at $t = 1$. Investors

have different beliefs about the distribution of the cash flow, and the disagreement is focused on a portion of it. More precisely, we denote the cash flow as

$$\tilde{A} = \tilde{V} + \tilde{U}, \quad (1)$$

and investors disagree on the distribution of \tilde{V} but share the same belief about the distribution of \tilde{U} . We assume \tilde{V} has a binary distribution. There are two types of investors, optimists o and pessimists p . Investor i , $i \in \{o, p\}$, believes the distribution of \tilde{V} is

$$\tilde{V} = \begin{cases} V_u & \text{with a probability } h_i, \\ V_d & \text{otherwise,} \end{cases} \quad (2)$$

with $V_u > V_d$ and $h_o > h_p$. We use α_o and α_p to denote the population sizes of optimists and pessimists, respectively, and $\alpha_o + \alpha_p = 1$. Without loss of generality, we assume \tilde{U} has a mean of zero. In addition, we have the following simplifying assumptions: First, \tilde{U} has a bounded support on $[-\Delta, \Delta]$, with $\Delta > 0$ and $V_u - \Delta > V_d + \Delta$. Second, $V_d - \Delta \geq 0$, i.e., \tilde{A} is nonnegative. Third, \tilde{U} is independent of \tilde{V} . We use $F(\cdot)$ to denote the cumulative distribution function of \tilde{U} , and assume that $F(\cdot)$ is differentiable. It is straightforward to generalize all these assumptions, and the analysis remains similar but becomes more tedious.

3.1 Financial Innovation

The disagreement among investors motivates them to trade. Naturally, optimists want to buy asset A , and pessimists want to sell. Moreover, the trading need may also induce financial innovation. In particular, we assume that investors can introduce any financial derivatives. More formally, we define the probability space spanned by \tilde{V} and \tilde{U} as $\mathcal{H} \equiv (\{V_d, V_u\} \times [-\Delta, \Delta], \mathcal{F}, \mathcal{P}_o, \mathcal{P}_p)$, where $\{V_d, V_u\} \times [-\Delta, \Delta]$ is the sample space, \mathcal{F} is the sigma-algebra generated by $\{V_d, V_u\} \times [-\Delta, \Delta]$, and \mathcal{P}_o and \mathcal{P}_p are the probability measures for optimists and pessimists, respectively. Any security in this economy can be described as a claim to a cash flow, which can be described by a random variable in \mathcal{H} . For convenience, if a derivative is a claim to a random cash flow \tilde{K} , we will refer to it as “asset K ”.

Note that we assume that the investors can directly design derivatives. Hence, there is no separate role for financial intermediation in this section. One interpretation is that financial intermediaries simply respond to the demand from investors when designing derivatives. Alternatively, one can model financial intermediaries’ other objectives. For example, in one of the extensions in Section 7, we introduce a financial intermediary into our economy. We assume that only the financial intermediary can design securities. Investors take securities as given and pay a fee when they trade the securities. The objective of the financial intermediary is to maximize the fees from investors’ trading.

3.2 Speculation vs. Hedging

In the above discussion, investors’ trading is motivated by speculation. However, one can reinterpret the model so that the trading is motivated by hedging. For example, one interpretation is that all investors have rational expectations, $\Pr(\tilde{V} = V_u) = \Pr(\tilde{V} = V_d) = 0.5$, but have different

preferences. The utility function of investor i , $i \in \{o, p\}$, is given by

$$u_i(c) = \begin{cases} h_i c & \text{if } \tilde{V} = V_u, \\ (1 - h_i)c & \text{if } \tilde{V} = V_d. \end{cases} \quad (3)$$

One can interpret this specification as investor o having hedging need at the “up state” $\tilde{V} = V_u$. For example, investor o may incur some extra cost (e.g., the cost of financial distress) at the up state, and so has an incentive to hedge this risk. That is, he has a relatively high marginal utility, h_o , at the up state. Likewise, investor p has a relatively high marginal utility at the “down state.” This modeling device is similar to that in models in which some investors prefer “early” consumptions while others prefer “late” ones: Specification (3) implies that some investors prefer consumption at the up state, while others prefer consumption at the down state. Note that, in our model, the speculation and hedging interpretations are mathematically equivalent. In our later discussions, we will mostly adopt the speculation interpretation, and it is straightforward to restate the results under the hedging interpretation.

3.3 Default

Following Geanakoplos (1997, 2003), we assume that, upon default, the debt holder (or derivative counterparty) can seize the collateral posted in the trade, but the defaulting investor faces no further penalty. This assumption can be broadly interpreted as limited enforcement.⁷ Essentially, our assumption implies that when an investor defaults, his counterparty can seize only the collateral posted for this trade, and finds it too costly to get further compensation by penalizing the defaulting investor (e.g., seizing other assets). Therefore, our analysis is perhaps best suitable for security trading, where, in the event of default, the top priority for creditors is perhaps to get compensated quickly. For example, in the Lehman bankruptcy case, 80% of Lehman’s derivative counterparties terminated their contracts within weeks of bankruptcy. That is, if the derivative position is in-the-money for a counterparty, it can immediately seize the collateral in Lehman’s margin account for that trade. If the collateral value is less than the amount owed to the counterparty, however, it would be very costly to get further compensation through the lengthy bankruptcy procedure as an unsecured debt holder. For example, the final settlement plan for Lehman bankruptcy was approved more than three years later in November 2011, and the senior bondholders get only 21.1 cents on the dollar. Even more seriously, many of the more than 906,000 derivatives transactions with Lehman were likely to be for hedging, speculation, or short-term financing purposes. For Lehman’s counterparties, the failure to get compensated quickly to reestablish those positions with other counterparties is likely to be much more costly.⁸

This lack of penalty upon default implies that investors need to post collateral to back up their promises. The focus of our analysis is how the collateralization process—more precisely the friction in it—determines the financial innovation in equilibrium. Before introducing the collateral friction, we first consider a frictionless benchmark next.

⁷See Kehoe and Levine (1993) for an early contribution. This idea has lately been applied to asset pricing; see, e.g., Alvarez and Jermann (2000), Chien and Lustig (2010).

⁸All the numbers for Lehman bankruptcy are based on Bala Dharan’s speech at the NYU Stern Five-Star Conference on Research in Finance on December 2, 2011.

3.4 Benchmark Case: Portfolio Margin

Let's first consider the case with a perfect collateralization procedure. Specifically, an investor can use any part of his asset as collateral. For convenience, we refer to this frictionless collateralization procedure as "portfolio margin" since an investor only needs to post collateral for his overall portfolio. It is straightforward to define the *collateral equilibrium with portfolio margin* as the prices of asset A and all derivatives introduced, each investor's positions in the riskless technology, asset A and derivatives, such that all investors' positions satisfy the portfolio margin constraint; all investors maximize their expected utility; and all financial markets clear: the aggregate holding in asset A is β and the aggregate holding in each derivative is zero.

Collateral constraints put restrictions on investors' trading. For example, investors cannot borrow without collateral. How does this constraint affect equilibrium prices? With this perfect collateralization procedure, it is easy to see that the collateral constraint is equivalent to the constraint that an investor's wealth has to be nonnegative for all possible states at $t = 1$. That is, this collateral constraint only rules out "empty" promises. The collateral equilibrium is identical to the equilibrium without collateral constraints but investors face a non-negative wealth constraint at $t = 1$.

Without collateral frictions, investors can simply invent a complete set of Arrow securities in this economy, so they can achieve the same equilibrium allocation and prices as in the traditional complete market equilibrium, which is Pareto optimal. However, this benchmark case does not have sharp predictions on which market will be developed in equilibrium. Casual intuition suggests that investors would introduce derivatives that are linked to the disagreement. However, it is not clear what would happen to other derivative markets. Would investors try to complete the markets? Which markets would thrive, and which would disappear? What is the notion of "optimal" innovation in this context? To shed light on these issues, we need to take seriously the frictions in the collateralization procedure.

3.5 Individual Security Margin

The key collateral friction analyzed in this paper is that an investor has to post collateral for each position in his portfolio separately, which we refer to as "individual security margin." Note that the collateral requirement under "portfolio margin" can be much smaller than that under "individual security margin." For example, if an investor holds a portfolio in which individual asset returns offset each other, under portfolio margin, the collateral requirements for the whole portfolio can be much lower than that for one individual asset alone.

In practice, however, this cross-asset netting is far from perfect. For example, if one asset in the portfolio is exchange-traded while the other is over-the-counter or traded at a different exchange, then the investor has to post collateral for both assets separately, even if these two positions largely offset each other. One famous example is that Metallgesellschaft AG, a German conglomerate, had a large short forward position in oil and an offsetting long position in oil futures in the early 1990s, but eventually ran into a liquidity crisis when the collateral requirement became excessive (see Culp and Miller (1995)). Moreover, it may be difficult or too costly for a dealer to precisely estimate the correlation among securities to determine the collateral for the whole portfolio. Or, a trader may

prefer not to reveal his whole portfolio to his dealer by having multiple dealers, which is a common practice among hedge funds. Finally, different parts of the portfolio may be governed by different jurisdictions, and regulations may also put various constraints on collateralization. For example, the Board of Governors of the Federal Reserve System has a number of regulations on the initial margin requirements (Regulations T, U, X) for various institutions.

Our individual security margin assumption captures these frictions by ruling out cross-netting, which manifests itself on the following occasions: When investors buy a risky asset, they can use the asset itself as collateral to borrow to finance the purchase. However, they cannot use one risky asset as collateral to finance the purchase of another risky asset. Similarly, when an investor shorts a risky asset, he needs to put the proceedings as well as some of his own cash as collateral (i.e., investors cannot use a risky asset as collateral to short another risky asset). It is easy to see that cross-netting is necessary if investors use a risky asset as collateral to long or short another risky asset, or issue state-contingent debts to finance the purchase of a risky asset. Our assumption rules out these transactions for simplicity. Its essence is that cross-netting is imperfect rather than impossible. This assumption reflects the practice in reality. For example, to purchase securities on margin is to use those securities themselves as collateral, and margin loans are generally not state contingent. The securities are placed in the margin account in “street name”—i.e., the broker-dealers are the legal owners and can liquidate those positions when investors fail to maintain certain margin requirements (see, e.g., Fortune (2000)). On the short side, as noted by Geczy, Musto, and Reed (2002), the collateral for equity loans is almost always cash, and the standard collateral for U.S. equities is 102% of the shares’ value.

This assumption also rules out asset-backed securities, whereby the security designer acquires some underlying assets and uses those assets as collateral to issue new securities, which are claims to certain tranches of the cash flows from the underlying assets. The designer’s objective is to maximize the proceeds from the new securities. This problem has been studied by Allen and Gale (1987) and DeMarzo and Duffie (1999), among others. While the asset-backed security market is very important, a major part of the derivative market, such as the futures and swaps, takes a different form. Usually, investors take both the long and the short sides of the derivatives and use liquid assets (e.g., cash or Treasury securities) rather than the underlying assets themselves as collateral. More importantly, those positions are often much larger than the underlying asset, which makes it difficult, if not impossible, to arrange the market in the form of asset-backed securities.⁹ Hence, in our analysis below, asset-backed securities are ruled out by the individual security margin assumption. However, our model implications remain similar if investors are allowed to issue asset-backed securities, but the procedure is costly.

Finally, one might think that long and short positions are treated asymmetrically in our model: one can only use cash as collateral for short positions, but there is no such constraint for long positions (e.g., investors can buy risky assets on margin). This may appear problematic for derivative contracts, where the distinction between long and short positions is immaterial. For example, a long position in a derivative that is a claim to a cash flow $-\tilde{A}$ is the same as a short position in a derivative that is a claim to \tilde{A} . However, this relabeling long and short positions has no real im-

⁹For example, according to *Financial Times*, Lehman had around \$130 billion bond outstanding before it collapsed, but the total credit derivatives linked to Lehman were around \$400 billion. “Bad news on Lehman CDS,” *Financial Times*, October 11, 2008.

pact. To see this, note that the derivative $-A$ has a negative price, and hence an investor is “paid” when taking a long position. Using this derivative as collateral, an investor can only “borrow” a negative amount (i.e., the investor has to put up some of his own cash). In other words, a levered long position in this derivative $-A$ is the same as a short position in derivative A . Therefore, this relabeling has no real impact.

3.6 Equilibrium with Individual Security Margin

We now formally define the investment opportunity sets and the equilibrium given derivative markets. To simplify notations, we will focus on the case with one generic derivative K , which is a claim to a random cash flow \tilde{K} in \mathcal{H} . It is straightforward to generalize to include multiple derivative markets by simply adjusting notations.

If an investor takes a long position in an asset, he can use this asset as collateral to borrow to finance part of the purchase. Suppose he uses one unit of asset C ($C = A$ or K) as collateral to borrow L , with a notional interest rate of $r(L, C)$. Then, at time $t = 1$, the lender receives

$$X(L, C) \equiv \min(L(1 + r(L, C)), \tilde{C}),$$

where \tilde{C} is the value of asset C at $t = 1$. That is, the lender receives $L(1 + r)$ when there is no default, and seizes the collateral asset C upon default (i.e., when $\tilde{C} < L(1 + r)$). Similarly, the payoff to the investor of the levered position is

$$W^+(L, C) \equiv \tilde{C} - X(L, C).$$

Note that this levered position requires $P_C - L$ capital from the investor, where P_C is the price of asset C at $t = 0$.

When taking a short position in an asset, the investor needs to use cash as collateral to back up his promise. We use \tilde{K} to refer to the “credibly promised” cash flow to the investor who holds asset K . For example, if an investor promises a cash flow \tilde{A} but posts only $V_u - \Delta$ cash as collateral, then the “real” promise is $\min(\tilde{A}, V_u - \Delta)$, because when the realized value of \tilde{A} is greater than the collateral, this investor will default on his promise and his counterparty gets only the collateral $V_u - \Delta$. Therefore, the real promise is $\min(\tilde{A}, V_u - \Delta)$ rather than \tilde{A} . In our discussion below, we will use the credible promised cash flow (rather than the artificially promised cash flow) to denote the derivative. This implies that, without loss of generality, we can assume that short sellers post enough collateral and will not default.¹⁰ Hence, if an investor shorts one unit of asset C , he needs to post $\max \tilde{C}$ cash collateral, where $\max \tilde{C}$ is the maximum of the cash flow from asset C at time $t = 1$. This short position requires $\max \tilde{C} - P_C$ capital from the short seller, and its payoff at time $t = 1$ is

$$W^-(C) \equiv \max \tilde{C} - \tilde{C}.$$

We use $\theta_{i,C}^+(L)$, for $i \in \{o, p\}$, $C \in \{A, K\}$ and $\theta_{i,C}^+(L) \geq 0$ for $L \geq 0$, to denote the number of units of asset C that is held by investor i , who borrows L against each unit of asset C . Similarly,

¹⁰For example, the promise of \tilde{A} backed by $V_u - \Delta$ cash collateral is the same as the promise of the cash flow $\min(\tilde{A}, V_u - \Delta)$ with $V_u - \Delta$ cash collateral. In other words, allowing short sellers to under-collateralize is essentially equivalent to introducing another derivative and assuming short sellers do not default.

we use $\theta_{i,C}^*(L)$ to denote the number of units of asset C that is held by investors who borrow from investor i : investor i lends $L\theta_{i,C}^*(L)$ to those investors and takes $\theta_{i,C}^*(L)$ units of asset C as collateral. We then define $M_{i,C}^+(x)$ such that $M_{i,C}^+(0) = 0$ and $dM_{i,C}^+(L) = \theta_{i,C}^+(L)$. That is, $M_{i,C}^+(x)$ is the total number of units of asset C that is held by investor i , who borrows less than or equal to x against each unit of asset C .¹¹ Similarly, we define $M_{i,C}^*(x)$ such that $M_{i,C}^*(0) = 0$ and $dM_{i,C}^*(L) = \theta_{i,C}^*(L)$. Finally, $\theta_{i,C}^-$ denotes the units of asset C that are shorted by investor i , and η_i denotes investor i 's investment in the riskless technology, with $\theta_{i,C}^- \geq 0$ and $\eta_i \geq 0$. With these notations, we can denote investor i 's ($i = o, p$) wealth at time $t = 1$ as

$$W_i = \sum_{C=A,K} \left(\int_0^{P_C} W^+(L, C) dM_{i,C}^+(L) + W^-(C) \theta_{i,C}^- + \int_0^{P_C} X(L, C) dM_{i,C}^*(L) \right) + \eta_i. \quad (4)$$

Note that since one can always give up his collateral and default on his promise, no investor can borrow more than the collateral value. Hence, we don't need to consider the case of $L > P_C$.

Investor i 's objective is to choose his portfolio $(M_{i,C}^+(L), \theta_{i,C}^-, M_{i,C}^*(L), \eta_i)$ for $C = A, K$ and $L \in [0, P_C]$, to maximize his expected wealth at $t = 1$:

$$\max \mathbf{E}_i(W_i) \quad (5)$$

$$s.t. \sum_{C=A,K} \left(\int_0^{P_C} (P_C - L) dM_{i,C}^+(L) + \int_0^{P_C} L dM_{i,C}^*(L) + \theta_{i,C}^- \left(\max \tilde{C} - P_C \right) \right) + \eta_i \leq e + \beta P_A. \quad (6)$$

where the left-hand side of equation (6) is the total capital an investor allocates to long positions, lending, short positions, and the riskless technology; the right-hand side is the investor's initial endowment.

Our focus is to analyze which derivative contract K and loan contracts will be adopted in equilibrium. Before introducing the notion of optimal innovation, however, we first analyze the equilibrium, taking the derivative contract K as given.

Definition 1 *The equilibrium given derivative K is defined as the prices of assets A and K , (P_A, P_K) , investors' holdings, $(M_{i,C}^+(L), \theta_{i,C}^-, M_{i,C}^*(L), \eta_i)$ for $i \in \{o, p\}$, $C \in \{A, K\}$ and $L \in [0, P_C]$, and the notional interest rates, $r(L, C)$, for all adopted loan contracts, such that for all investors, their holdings solve their optimization problem (5), and all markets clear:*

$$\sum_{i=o,p} \left(M_{i,A}^+(P_A) + \theta_{i,A}^- \right) = \beta; \quad (7)$$

$$\sum_{i=o,p} \left(M_{i,K}^+(P_K) + \theta_{i,K}^- \right) = 0; \quad (8)$$

and for $i \in \{o, p\}$, $j \neq i$, $C \in \{A, K\}$, and $L > \min \tilde{C}$:

$$\theta_{j,C}^+(L) = \theta_{i,C}^*(L). \quad (9)$$

¹¹The relation between $\theta_{i,C}^+$ and $M_{i,C}^+$ is similar to that between a Probability Density Function and its corresponding Cumulative Distribution Function. The reason we need to define $M_{i,C}^+(x)$ is that, in the ensuing equilibrium, $\theta_{i,C}^+(L) = 0$ for all but a finite number of values of L . Hence, one cannot calculate the investor i 's total holding in asset C by integrating $\theta_{i,C}^+(L)$ along L .

Equations (7) and (8) state that the aggregate demand is β units for asset A and zero for asset K . Equation (9) implies that borrowing is equal to lending for all loan markets with $L > \min \tilde{C}$. Note that if $L \leq \min \tilde{C}$, this borrowing is riskless and can be arranged through the riskless technology, rather than borrowing from some other investors in the economy.¹²

Due to the disagreement on \tilde{V} , investors would like to speculate on its value at $t = 1$. It is natural to conjecture that, in equilibrium, investors would adopt a derivative contract, asset V , which is a claim to a cash flow \tilde{V} at $t=1$. Before we demonstrate that asset V will indeed be adopted, we first construct the equilibrium, taking the market for asset V as given. Our analysis will focus on the case $\underline{\alpha} \leq \alpha_p \leq \bar{\alpha}$, where

$$\begin{aligned}\underline{\alpha} &\equiv \frac{\gamma - \gamma h_o}{\gamma + \beta h_o}, & \text{with } \gamma &= \frac{h_o [e + \beta (V_d - \Delta)]}{h_o (V_u - V_d) + \Delta}, \\ \bar{\alpha} &\equiv 1 - \frac{h_p e + h_p \beta (V_u - V_d - \Delta)}{e + \beta [h_p (V_u - V_d) + V_d]}.\end{aligned}$$

The equilibrium in other cases is uninteresting and is completely dominated by one group investors. For example, in the case $0 < \alpha_p < \underline{\alpha}$, there are so few pessimists that the equilibrium prices of assets A and V are completely determined by optimists' belief $P_A = \mathbf{E}_o[\tilde{A}]$ and $P_V = \mathbf{E}_o[\tilde{V}]$. Similarly, when $\alpha_p > \bar{\alpha}$, there are so many pessimists that the equilibrium prices are completely determined by pessimists' belief. Hence, our focus will be on the intermediate region $\underline{\alpha} \leq \alpha_p \leq \bar{\alpha}$, where the equilibrium is determined by the interaction between the two groups. To best illustrate our main results, we first analyze the case $\underline{\alpha} \leq \alpha_p < \alpha_1$, where the value of α_1 is given by (22) in the Appendix, and we leave the analysis of the case $\alpha_1 \leq \alpha_p \leq \bar{\alpha}$ to Section 5.

Proposition 1 *In the case $\underline{\alpha} \leq \alpha_p < \alpha_1$, the equilibrium is characterized as follows:*

1. *The prices of assets A and V are given by*

$$P_A = \frac{e\alpha_o + (\gamma + \beta)(V_d - \Delta)}{\gamma + \beta\alpha_p}, \quad (10)$$

$$P_V = \frac{\gamma\alpha_o}{\gamma + \beta\alpha_p}V_u + \frac{(\gamma + \beta)\alpha_p}{\gamma + \beta\alpha_p}V_d. \quad (11)$$

2. *A fraction $\beta/(\gamma + \beta)$ of the optimists holds a levered position in asset A . Each investor holds $(\gamma + \beta)/\alpha_o$ units and borrows $V_d - \Delta$ against each unit of asset A , with an interest rate of 0.*
3. *A fraction $\gamma/(\gamma + \beta)$ of the optimists takes a levered position in asset V . Each of them holds $\frac{e + \beta P_A}{P_V - V_d}$ units and borrows V_d against each unit of asset V , with an interest rate of 0.*
4. *Each pessimist shorts $\frac{e + \beta P_A}{V_u - P_V}$ units of asset V . For each unit of the short position, the investor posts V_u cash as collateral.*

This proposition highlights one of the key points of this paper: An important motivation for financial innovation is to reduce collateral constraints. Note that a unit of asset A gives investors

¹²One interpretation is the following. The cash collateral in the economy is kept at a custodian bank, which can invest the cash only in riskless investments. So, if an investor has sufficient collateral to guarantee no default, he can borrow from this custodian bank at the riskless interest rate.

the same exposure to \tilde{V} as a unit of asset V . However, asset V allows the buyer to take on more leverage: As shown in items 2 and 3 in the proposition, an investor can borrow only $V_d - \Delta$ against each unit of asset A , but can borrow V_d against each unit of derivative V . What is the cause of this difference? Note that the cash flow from the underlying asset has two portions, \tilde{V} and \tilde{U} , but investors are only interested in trading \tilde{V} . To the extent that the “unwanted” portion, \tilde{U} , increases the collateral requirement for the buyer of the underlying asset, it makes the underlying less appealing. That is, to buy the underlying asset, the investor has to “waste” his collateral to cover the risk he is not interested in taking. This is unappealing even if the investor is risk neutral. If assets A and V had the same price, the buyer would strictly prefer the derivative V . Indeed, from equations (10) and (11), one would find that P_A is lower than P_V . This price discount reflects the shadow value of collateral and compensates investors who hold asset A in equilibrium. As shown in items 2 and 3 in the proposition, a fraction $\beta/(\gamma + \beta)$ of optimists holds a levered position in asset A and the rest of the optimists hold a levered position in the derivative V , and they are indifferent about these two positions.¹³

An important feature of our model is that the collateral requirements for both the underlying asset and the derivative are endogenously determined in equilibrium. If an optimist wants to take a levered position in asset A , he can borrow at the riskless interest rate, 0, if he posts enough collateral to guarantee no default. Alternatively, he can reach out to other investors to enter a loan contract, if he and the lenders can agree on the collateral and interest rate. For example, if an optimist borrows more than $V_d - \Delta$ against each share of asset A as collateral, he has to offer a higher interest rate to his lender to compensate the default risk. If a pessimist agrees to lend, this lending choice has to be no worse than his outside option, which is taking a short position in V . In the case $\underline{\alpha} \leq \alpha_p < \alpha_1$, optimists cannot offer an interest rate that is high enough to attract pessimists to lend to them. Therefore, in equilibrium, optimists borrow $V_d - \Delta$ against each share of asset A and the interest rate is 0. For the same reason, optimists borrow V_d against each unit of asset V and the interest rate is 0.

Finally, as shown in item 4 in the proposition, pessimists take a short position in the derivative contract V . No pessimist chooses to short A . This is because that due to the unwanted risk from \tilde{U} , in order to obtain the same exposure to \tilde{V} , shorting A requires more collateral. In addition, as noted earlier, $P_A < P_V$. This makes shorting A even less appealing.

In summary, the above discussion suggests that trading asset A needs more collateral because of the “unwanted” risk from \tilde{U} . The derivative contract is appealing because it “carves” out the unwanted risk. This intuition suggests that the derivative contract that completely carves out the unwanted cash flow should be the “most appealing” financial innovation in the economy, as we formally analyze next.

Definition 2 *A financial innovation X (a claim to a cash flow \tilde{X} in \mathcal{H} at $t = 1$) is optimal if, in the presence of X , one introduced any other derivative contract K (a claim to a cash flow \tilde{K} in \mathcal{H} at $t = 1$), the market for K wouldn't generate any trading, unless \tilde{K} is perfectly correlated with \tilde{X} .*

¹³Investors are risk neutral. Therefore, if they are indifferent about the two strategies, they are also indifferent about any combination of the two strategies. Hence, we can also interpret the result as “a fraction $\beta/(\gamma + \beta)$ of optimists' wealth is invested in the levered position in A and the rest of their wealth in the levered position in V .”

Proposition 2 *The derivative contract V is an optimal financial innovation.*

Due to the disagreement, optimists prefer to transfer their wealth at $t = 1$ to the up state and the pessimists the down state. Derivative V is the most efficient instrument since it allows investors to transfer all their wealth to the states they prefer. For example, as noted in item 3 of Proposition 1, the buyer of asset V can borrow V_d against each unit of asset V . Hence, his wealth is zero at the down state. Alternative assets cannot achieve this goal. Consider asset A , for example. As shown in Proposition 1, an investor can borrow only $V_d - \Delta$ against each unit of asset A . Therefore, this investor cannot completely transfer his wealth to the up state, since his wealth at the down state is always positive unless the realization of \tilde{A} happens to be $V_d - \Delta$. Similarly, a pessimist who shorts asset A cannot transfer all his wealth to the down state. Trading V leads to Pareto improvement because it enables both optimists and pessimists to transfer their wealth to the states they prefer. This explains why no investors short asset A in equilibrium. More generally, the above intuition implies that, in the presence of the market for V , any alternative derivative markets cannot generate any trading.

Our analysis emphasizes the importance of carving out unwanted cash flow to isolate the the component with disagreement. This is different from the standard intuition of completing markets. In fact, it is possible that in a world with N possible states, investors may choose to introduce more than N securities and the markets are still incomplete in the sense that the securities do not span the whole state space. For example, suppose there are M groups and each group has two investors, who disagree on the probabilities of two states but agree that each of the rest of the $N - 2$ states has a probability of $1/N$. For the two states with disagreement, one investor believes their probabilities are $1/N + x$ and $1/N - x$, where $0 < x < 1/N$, and the other investor believes their probabilities are $1/N - x$ and $1/N + x$. In this case, investors want to introduce M securities, each of which facilitates the specific speculation needs for one group. Specifically, each group of investors would introduce one security, which pays \$1 and -\$1 for the two states with disagreement and 0 for all other states. Note that M can be as high as $\binom{N}{2} = N(N - 1)/2$. That is, investors may want to introduce more than N securities. If there are some states for which all investors agree on, they will not introduce securities to span those states. Therefore, it is possible that investors may choose to introduce more than N securities and markets are still incomplete.

3.7 Implementation with Transaction Costs

Proposition 2 states that derivative contract V is an optimal security. It does not, however, pin down the unique contract in the economy. In fact, any linear transformation of asset V serves exactly the same function as asset V . For example, if asset X is a claim to a cash flow $\tilde{X} = a(\tilde{V} + b)$, where a and b are constants, it serves the same economic function as asset V . To see this, we note that a can be normalized to 1 by redefining the size of the derivative contract. So, we only need to consider the case $\tilde{X} = \tilde{V} + b$. The difference between assets X and V is that asset X pays an extra constant cash flow b . Not surprisingly, if we introduce asset X into the economy described earlier, its price would be $P_X = P_V + b$. Using asset X as collateral, an investor can borrow $V_d + b$ and the interest rate is 0. This levered position provides the same payoff as a levered position in asset V . Therefore, asset X is also an optimal financial innovation for any b .

In the next, however, we will illustrate that any infinitesimal transaction cost can pin down the unique optimal contract. Suppose there is a small transaction cost for transferring funds from one account to another. Specifically, the cost for transferring one dollar from an investor to another is k dollars and the sender and receiver each pay $k/2$, where k is a positive constant. For simplicity, we assume that k is infinitesimal. Therefore, at $t = 0$, investors are facing essentially the same problem as analyzed before and the equilibrium prices are the same as those in Proposition 1. Moreover, investors can minimize their transaction costs by “fine-tuning” the derivative contract, as shown in the following proposition.

Proposition 3 *In the presence of the transaction cost described above, the unique optimal financial innovation is asset X with the unique value of b , such that $P_X = 0$.*

This proposition illustrates the appeal of the contract with a zero initial price, a common feature in many derivatives in practice—e.g., futures and swaps. They are the so-called unfunded securities, with the name highlighting the fact that investors can establish their positions without paying at inception and need to post collateral only to cover the daily mark-to-the-market movements. It is easy to see why unfunded securities are appealing. Suppose we had chosen b such that the contract’s initial value is not zero. Then the cash flows from trading this security can be decomposed into two components. The first component is the cash flows from trading a corresponding unfunded security. The second component is the following: Since the initial price of the derivative is not zero, one party gets paid at $t = 0$, but then he may need to pay this amount back at $t = 1$. Note that while the first component serves the economic function by facilitating the speculation among investors, the second one is completely redundant. Making the derivative unfunded avoids this potential “round trip” in fund transfer. To the extent that there is an infinitesimal cost of transferring funds, the unfunded security would be strictly preferred.

4 Violation of the Law of One Price

As noted earlier, buying asset A is inefficient due to its higher collateral requirement. In equilibrium, therefore, to induce an investor to hold it, there has to be a price discount relative to the derivative V . This can potentially lead to the violation of the law of one price—that is, the price of an asset can be different from the price of its replicating portfolio.

Before we proceed with our analysis, it is helpful to first describe the empirical motivation. One example is the so-called corporate bond–CDS basis, the difference between the CDS spread and the corresponding corporate bond yield spread. As noted in Mitchell and Pulvino (2010) and Garleanu and Pedersen (2011), CDS spreads tend to be lower than the corresponding corporate bond yield spreads, although both are measures of the underlying firm’s credit risk and the no-arbitrage relation implies that the difference between the two should be near zero. In other words, a corporate bond can be decomposed into a short position in a CDS contract on the bond issuer plus a Treasury bond. The empirical evidence suggests that the price of the corporate bond is often lower than the price of the portfolio of the CDS and Treasury bond.

Keep this example in mind, let’s now examine the prices in our model. Note that asset A can be decomposed into assets V and U , where asset U is a claim to a cash flow \tilde{U} at $t = 1$. That

is, to analyze the violation of the law of one price, we need to compare P_A with $P_V + P_U$, where P_U is the price of asset U . Note that P_A and P_V have been determined in Proposition 1. How is P_U determined? To see this, it is helpful to map our model to the earlier example. Assets A , V , and U in our model correspond to the corporate bond, CDS, and Treasury bond, respectively. So, how is Treasury bond (asset U) price determined? It is obviously jointly determined by a large number of investors, many of whom are not involved in the corporate bond (asset A) market at all. Therefore, a natural way to think about asset U is to assume that there is another market (the Treasury market in our example), in which a large number of investors trade asset U , and these investors (e.g., sovereign funds, repo trading desks at investment banks) do not trade assets A and V . If the investors in asset U market are all risk neutral and don't have funding constraints (e.g., sovereign funds), we have $P_U = 0$. In contrast, if those investors are risk averse or face funding constraints, P_U is negative. Finally, if investors are attracted by some special features of asset U , its price can be positive. This can happen, for example, during fly-to-quality in crises, or more generally when investors treat Treasury securities as money and value their convenience yield (Krishnamurthy and Vissing-Jorgensen (2010)). Note that if P_U is close to 0, both optimists and pessimists in our model would choose not to trade asset U even if they have access to this market. In this case, the equilibrium prices of A and V are still the same as those in Proposition 1.

4.1 Model Implications

We use B to denote the basis, the price difference between asset A and its replicating portfolio: $B \equiv P_V + P_U - P_A$. Naturally, the basis can be decomposed into two components: $B = S + P_U$, where $S \equiv P_V - P_A$. The first component S reflects the shadow value of collateral. Despite risk neutrality, investors still value A less than V , because V allows investors to bet with less collateral.¹⁴ The second component is derived from the fact that asset U (e.g., Treasury) is probably more liquid and has certain convenience yield, as discussed in Krishnamurthy and Vissing-Jorgensen (2010). It is important to note that the second component reflects how much value investors assign to the convenience yield from Treasury, and so affects the basis for all assets. In contrast, the first component, S , depends on the characteristics of asset A and hence has both cross-sectional and time-series implications on basis, as characterized in the following proposition.

Proposition 4 *The price spread S is positive and has the following properties.*

- (1) S increases when there is less cash in the economy: $\frac{\partial S}{\partial e} < 0$.
- (2) S increases when asset A has more unwanted risk: $\frac{\partial S}{\partial \Delta} > 0$.
- (3) The impact in (1) is stronger when there is more unwanted risk: $\frac{\partial^2 S}{\partial e \partial \Delta} < 0$.
- (4) Suppose that an outside investor has to sell β^* units of asset A to the investors in this

¹⁴If investors are risk averse, the unwanted risk in the underlying asset (i.e., \tilde{U}) also makes asset A unappealing to investors. However, risk aversion alone is not enough to generate a spread between A and V , if investors can hedge the unwanted risk. Nevertheless, if one introduces both risk aversion and cost of hedging to our model, a spread between A and V will arise in equilibrium, reflecting both collateral value and hedging cost. The collateral component of the spread perhaps can be separately empirically identified when, for example, there is an exogenous shock to margin requirement policy.

economy. The spread increases: $\frac{\partial S}{\partial \beta^*}|_{\beta^*=0} > 0$, and this impact is stronger when there is more unwanted risk: $\frac{\partial^2 S}{\partial \beta^* \partial \Delta}|_{\beta^*=0} > 0$.

(5) S increases with the disagreement among investors: $\frac{\partial S}{\partial h_o} > 0$.

(6) S increases when the sizes of the two groups of investors are more balanced: $\frac{\partial S}{\partial \alpha_p} > 0$.

Result (1) shows that this spread increases when investors face tighter funding liquidity. This is because saving collateral becomes more valuable when investors have less cash but need leverage. Similarly, Result (2) says that the spread is larger if \tilde{U} is more volatile (i.e., Δ is larger). The larger the risk in \tilde{U} , the more collateral can be saved by trading V , leading to a larger price spread. Related to these two results, Result (3) shows that when the funding liquidity in the economy tightens (i.e., e decreases), the spread increases more for assets with more volatile unwanted cash flow (i.e., larger Δ). These implications are parallel to those in Garleanu and Pedersen (2011), where the collateral requirements are exogenous. In contrast, our model endogenizes both financial innovation and the collateral requirements, and explains why differential collateral requirements arise in the first place. The model highlights that it is the trading needs among investors that cause both financial innovations and differential collateral requirements across assets.

Our model also has a number of novel predictions. For example, if a large investor (e.g., a failing hedge fund) has to liquidate his positions in asset A at $t = 0$, what is the impact of this supply shock on equilibrium prices? Clearly, the prices of both A and V will drop. Result (4) shows that the price of A drops more—i.e., the supply shock increases the spread. This is because it requires more capital to absorb A than to absorb V . Therefore the price of A is more sensitive to supply shocks. Similarly, the impact of supply shocks is stronger on assets with a larger unwanted risk (larger Δ). Finally, perhaps the most distinctive feature of our model is the speculation among investors. Our model shows that the spread increases with investors' desire to trade. This is because the spread reflects the shadow value of collateral. The stronger the desire to trade, the higher the shadow value of collateral. For example, Result (5) shows that the spread increases with the disagreement among investors. Holding everything else constant, an increase in h_o (i.e., the optimists become even more optimistic) increases the disagreement among investors and the desire to trade. Similarly, Result (6) shows that the spread increases when the sizes of the two groups of investors are more balanced. Intuitively, if the whole population is predominately one type of investors, the belief of those investors would mostly determine asset prices. Since the prices closely reflect their expectation, those investors have a weaker speculation motive. Note that in the case analyzed in Proposition 4, investors are predominately optimists. An increase in pessimists' population size α_p makes it more balanced between the optimists and pessimists, leading to a stronger trading need and so a higher spread.

4.2 Existing Evidence and New Testable Predictions

The above implications shed light on the empirical evidence from a number of studies. The previously mentioned corporate bond–CDS basis arises naturally in our model. Suppose an investor, say a hedge fund, wants to take an exposure on a corporate bond. He can either buy this bond on margin or he can simply short a CDS contract on this corporation. To establish the same exposure

to the default risk of the underlying firm, the corporate bond position requires more collateral because it also has embedded interest rate risk. In other words, if the corporate bond–CDS basis were zero, shorting CDS would be more desirable to the investor. In equilibrium, therefore, the CDS rate is lower, leading to a positive corporate bond–CDS basis.

Moreover, consistent with Result (1), the corporate bond–CDS basis increased substantially during the recent financial crisis, when the funding liquidity was probably tight for most investors. It is possible that fly-to-quality during the crisis disproportionately increased the price of Treasury bonds (P_U in our model) and so contributed to the observed increase in basis. However, this interpretation cannot account for the cross-sectional variation in basis. As documented in Mitchell and Pulvino (2010), during the crisis, the corporate bond–CDS basis tends to be larger for junk bonds than for investment-grade bonds. This evidence is consistent with Result (2), if one takes the interpretation junk bonds have more non-default-related risks (e.g., liquidity risk). Moreover, with the financial crisis unfolding, the basis for junk bonds increases more than that for investment-grade ones, consistent with Result (3).

Another example is the TIPS–inflation swap basis. Fleckenstein, Longstaff, and Lustig (2010) find that the price of TIPS is consistently lower than the price of its replicating portfolio that consists of inflation swaps and nominal Treasury bonds. This is consistent with our model. If an investor decides to hedge inflation, or speculate that inflation will go up, he can buy TIPS, or take a long position in inflation swaps. Note that, relative to the former strategy, the latter needs less collateral to establish the same exposure to inflation. Hence, our model implies that other things equal, investors would prefer the long position in inflation swap, leading to a swap rate that is higher than what is implied by TIPS and nominal Treasury bonds. Moreover, consistent with Result (1), they find that this price discrepancy between inflation swaps and TIPS also increased dramatically during the recent financial crisis.

Counterparty risk may have also contributed to the observed bases. As dealers' default probability increases during the crisis, the CDS contracts they underwrite become less valuable, leading to a lower CDS spread and so a higher corporate bond–CDS basis. Arora, Gandhi, and Longstaff (2010) find that counterparty risk is indeed priced in the CDS market. However, they note that perhaps due to the common practice of full collateralization of swap liabilities, the impact on CDS spread is very small. Moreover, it is not clear whether counterparty risks increase or decrease the TIPS–inflation swap basis. To the extent that the concern of counterparty risks reduces the value of inflation hedge offered by weakened institutions, this would decrease the inflation swap rates and so decreases the basis, opposite to the evidence.

Finally, Proposition 4 offers a number of testable predictions. Result (2) suggests that other things equal, the corporate bond–CDS basis and the TIPS–inflation swap basis should be larger for bonds with higher unwanted risks. This implies, for example, that the basis should be larger for corporate bonds or TIPS with longer maturities, or when the riskless interest rate volatility is higher. Moreover, Result (4) implies that the basis should increase when there is a positive supply shock to the underlying asset (e.g., when a large institution is forced to liquidate its corporate bonds or TIPS). This supply shock impact should be stronger for bonds longer maturities, or when the riskless interest rate volatility is larger. Finally, Results (5) and (6) imply that the basis should be larger when the trading motive is stronger.

5 Other Cases

The analysis so far is focused on the case $\underline{\alpha} \leq \alpha_p \leq \alpha_1$. This section presents the results from other cases. Note that there is no default risk in equilibrium in the case $\underline{\alpha} \leq \alpha_p \leq \alpha_1$. For example, using one share of asset A as collateral, optimists choose to borrow $V_d - \Delta$. Hence, even in the worst case, the value of the collateral is enough to pay back the debt. Optimists have the option to borrow more against the collateral. But, in the case $\underline{\alpha} \leq \alpha_p \leq \alpha_1$, pessimists would charge an interest rate that is too high, so that optimists prefer to borrow only $V_d - \Delta$ to get the riskless interest rate. In the case of $\alpha_1 \leq \alpha_p \leq \bar{\alpha}$, however, pessimists can offer a rate that is also acceptable to optimists. Depending on the relative sizes of the two groups of investors, the equilibrium can be characterized by two subcases. In the first case, $\alpha_1 \leq \alpha_p < \alpha_2$, only a fraction of the asset A -backed debts has default risk; while in the other case, $\alpha_2 \leq \alpha_p \leq \bar{\alpha}$, all asset A -backed debts have default risk, where the expression for α_2 is given by equation (37) in the Appendix.

Proposition 5 *In the case $\alpha_1 \leq \alpha_p < \alpha_2$, the equilibrium is characterized as follows:*

1. *The prices of assets A and V are given by*

$$P_A = \frac{1}{z^* + 1}V_u + \frac{z^*}{z^* + 1}V_d - \left(1 - \frac{1}{h_o} \frac{1}{z^* + 1}\right) \Delta, \quad (12)$$

$$P_V = \frac{1}{z^* + 1}V_u + \frac{z^*}{z^* + 1}V_d, \quad (13)$$

where $z^* \equiv \frac{\alpha_1(\gamma + \beta)}{(1 - \alpha_1)\gamma}$.

2. *Optimists are indifferent about the following three strategies:*

- *A measure x_o^* of them holds a levered position in asset V , where x_o^* is given by equation (32). Each of them holds $\frac{e + \beta P_A}{P_V - V_d}$ units and borrow V_d against each unit of asset V , with an interest rate of 0.*
- *A measure y_o^* of them holds a levered position in asset A , where y_o^* is given by equation (36). Each investor holds $\frac{e + \beta P_A}{P_V - V_d + \Delta}$ units and borrows $V_d - \Delta$ against each unit of asset A , with an interest rate of 0.*
- *The rest of them also hold a levered position in asset A . Each of them holds $\frac{e + \beta P_A}{P_V - L^*}$ units, where L^* is given by (34), and borrows L^* against each unit of asset A , with a positive interest rate that is given by (38).*

3. *Pessimists are indifferent about the following two strategies:*

- *A measure x_p^* of them shorts asset V , where $x_p^* = z^* x_o^*$. Each of them shorts $\frac{x_o^*(e + \beta P_A)}{x_p^*(P_V - V_d)}$ units, and posts V_u cash as collateral for each unit of the short position.*
- *The rest of them lend all their wealth to optimists. For each unit of asset A as collateral, they lend L^* , and the interest rate is given by (38).*

This case is similar to that analyzed in Proposition 1. The main difference is that some optimists can now take more leverage, but need to pay a higher interest rate to compensate their lenders for the credit risk. More precisely, optimists are indifferent about the three strategies: A measure x_o^* of optimists takes a levered position in derivative V . Other optimists take long positions in asset A , but they have two different ways to finance their positions. A measure y_o^* of them chooses to borrow less, so that they face the riskless interest rate. The rest of them, however, chooses to borrow more and faces a higher interest rate. The bigger loan enables them to have a larger position, and this extra expected profit compensates the higher interest they face.

Another new feature is shown in equations (12) and (13): The prices of assets A and V are independent of α_p . In the equilibrium in Proposition 1, both P_A and P_V decrease in α_p . This is intuitive. More pessimists take short positions in V , pushing down P_V . This attracts more optimists from A to V , and hence pushes down P_A as well. In the case of Proposition 5, however, there is another force. With the increase of α_p , more pessimists choose to lend to optimists. This enables optimists to borrow to invest more, pushing up P_A and P_V . These two forces exactly offset each other in the case in Proposition 5, so that P_A and P_V do not depend on α_p .

Proposition 6 *In the case $\alpha_2 \leq \alpha_p \leq \bar{\alpha}$, the equilibrium is characterized as follows:*

1. *The prices of assets A and V are given by*

$$P_A = \left(\frac{1}{x_o^{**} + x_p^{**}} - 1 \right) \frac{e}{\beta}, \quad (14)$$

$$P_V = \frac{x_o^{**}}{x_o^{**} + x_p^{**}} V_u + \frac{x_p^{**}}{x_o^{**} + x_p^{**}} V_d, \quad (15)$$

where x_o^{**} and x_p^{**} are given by (40) and (41).

2. *Optimists are indifferent about the following two strategies:*

- *A measure x_o^{**} of them holds a levered position in asset V . Each of them holds $\frac{e+\beta P_A}{P_V - V_d}$ units and borrows V_d against each unit of asset V , with an interest rate of 0.*
- *The rest of them hold a levered position in asset A . Each of them holds $\frac{e+\beta P_A}{P_V - L^{**}}$ units, where L^{**} is given by (39), and borrows L^{**} against each unit of asset A , with a positive interest rate that is given by (43).*

3. *Pessimists are indifferent about the following two strategies:*

- *A measure x_p^{**} of them shorts asset V . Each of them shorts $\frac{x_o^{**}(e+\beta P_A)}{x_p^{**}(P_V - V_d)}$ units and posts V_u cash as collateral for each unit of the short position.*
- *The rest of them lend all their wealth to optimists. For each unit of asset A as collateral, they lend L^{**} , and the interest rate is given by (43).*

Similar to the previous case, some of loans have default risk. One difference is that in this case, when optimists use A as collateral to borrow, they all prefer to borrow more than $V_d - \Delta$ and pay a positive interest rate. Putting together the three cases in Propositions 1, 5, and 6, we obtain

the plots in Figure 1. The upper panel plots P_A and P_V against α_p . At $\alpha_p = \underline{\alpha}$, both P_A and P_V are completely pinned down by optimists' expectation $P_A = \mathbf{E}_o[\tilde{A}]$ and $P_V = \mathbf{E}_o[\tilde{V}]$. As the population size of pessimists increases, both prices decrease in the case of $\underline{\alpha} \leq \alpha_p < \alpha_1$. In the case of $\alpha_1 \leq \alpha_p < \alpha_2$, however, both prices stay constant while α_p changes, as shown in Proposition 5. Finally, in the region $\alpha_2 \leq \alpha_p < \bar{\alpha}$, when α_p increases, P_V decreases but P_A increases. The reason is that, as noted earlier, two forces arise when pessimists' population size increases. First, their larger short positions in V push down its price and this attracts more optimists to take long positions in V , reducing the number of optimists holding A . Second, more pessimists compete to lend to optimists, and push down the interest rate on asset A -backed debts, giving optimists more purchasing power. This second impact can dominate and so P_A increases with α_p .

Figure 1: Asset Prices and Interest Rates.

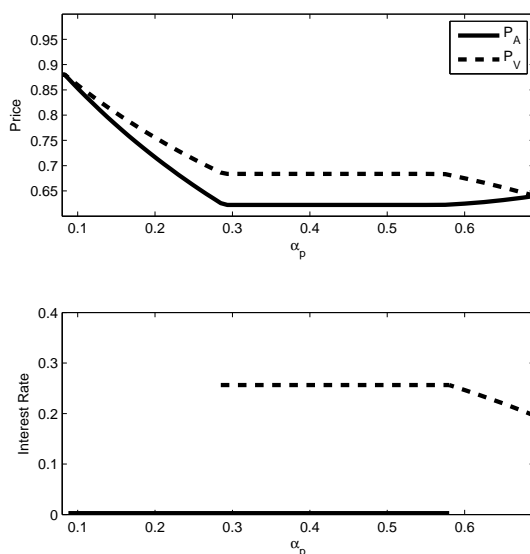


Figure 1: The upper panel plots P_A and P_V against α_p . The lower panel plots the notional interest rates on loans backed by asset A . Parameter values: $h_p = 0.4, h_o = 0.8, V_u = 1, V_d = 0.4, \beta = 1, e = 0.2, \Delta = 0.15$; \tilde{U} is uniformly distributed.

One can see the above intuition more clearly by examining the interest rates. Across all three cases in the region $\underline{\alpha} \leq \alpha_p \leq \bar{\alpha}$, all asset V -backed debts are riskless and have a zero interest rate. In contrast, the credit risk of asset A -backed debts varies across cases. As shown in the lower panel of Figure 1, asset A -backed debts are riskless and have a zero interest rate in the case of $\underline{\alpha} \leq \alpha_p < \alpha_1$. In the case of $\alpha_1 \leq \alpha_p < \alpha_2$, however, some of the asset A -backed debts are riskless and have a zero interest rate, while the rest have default risk and have a positive interest rate, which stays constant throughout the region. Finally, in the case of $\alpha_2 \leq \alpha_p < \bar{\alpha}$, all asset A -backed debts have default risk. Note that the interest rate drops when α_p increases, indicating that when more pessimists compete to lend, they push down the interest rate. This benefits the borrower (i.e., the optimists) and can even lead to the result in the upper panel that P_A increases with α_p .

6 The Impact of Financial Innovation

How does financial innovation affect the economy? To analyze this, we compare the equilibria across two economies. The first is the economy analyzed above. As a comparison, the second economy does not have the market for V and is otherwise identical to the first economy. The analysis of the second economy is similar to that in previous sections, and we leave the details to the Appendix.

Proposition 7 *The introduction of asset V has a mixed impact on the price of asset A .*

The intuition is as follows. The derivative contract V is efficient in facilitating investors' bets. On the one hand, optimists prefer to buy asset V , rather than the underlying asset A . This puts downward pressure on the price of asset A . On the other hand, pessimists are also attracted to shorting V , away from shorting A . This increases the price of A . The overall impact on asset A is mixed, and determined by the tradeoff between these two forces.

Interestingly, with the presence of asset V , the price of asset A can be even lower than the pessimists' expected value $\mathbf{E}_p[\tilde{A}]$. Note that in the economy without asset V , the price of asset A is always between $\mathbf{E}_p[\tilde{A}]$ and $\mathbf{E}_o[\tilde{A}]$, the expected values of the two groups of investors. This is natural. If the price of A were less than $\mathbf{E}_p[\tilde{A}]$, for instance, both investors would want to buy it, which would have pushed up the price. In the presence of the derivative V , however, Figure 1 shows that when α_p is large, the price of asset A is even lower than the pessimist's expected value. Although pessimists find it profitable to buy asset A , they choose not to do so because they find trading asset V even more profitable.

In our previous discussion, we mostly take the heterogeneous belief interpretation. This makes it harder to examine the welfare implications because it is unclear which belief should be used when calculating investors' welfare.¹⁵ As noted in Section 3.2, one can simply adopt the hedging interpretation. Investors' welfare under the hedging interpretation is mathematically identical to their subjective expected utility under the old heterogeneous-belief interpretation, and is reported in the following.

Proposition 8 *The introduction of V has a mixed impact on investors' welfare.*

Intuitively, the derivative contract V helps investors to transfer their wealth to the states they prefer, and so improves their welfare. However, the introduction of V also changes asset prices in the economy and so indirectly affects investors' welfare. For example, if the innovation increases the price of asset A , it decreases the expected utility of optimists since they now have to buy the asset at a higher price. Similarly, the innovation increases the optimists' welfare if it decreases the price of asset A . As a result, financial innovation's mixed impact of the price of asset A , as shown in Proposition 7, translates into the mixed impact on investors' welfare.

¹⁵Brunnermeier, Simsek, and Xiong (2011) propose a solution to welfare analysis with heterogeneous beliefs.

7 Extensions

7.1 A Policy Experiment on Margin Requirement

The Securities and Exchange Act of 1934 empowered the Federal Reserve Board (Fed) to set the margin requirements for trading stocks and bonds. One of the main goals of this policy is to contain the volatility caused by speculation. The Fed was initially active in adjusting these margin requirements. For example, since the margin requirement for stock trading was established on October 1, 1934, the Fed adjusted it 23 times in the next 40 years. After this margin requirement was set to 50% on January 3, 1974, however, the Fed hasn't made any more changes. A large number of empirical studies have attempted to examine the impact of margin requirements on the stock market volatility, leading to opposite conclusions and debates on the adopted methodologies. Most of the recent studies, however, generally conclude that the reduction in margin requirements does not lead to a noticeable increase in stock market volatility.¹⁶

How does the margin policy affect the economy in our model? How does financial innovation affect the effectiveness of the policy? To analyze these issues, we conduct a “policy experiment” — i.e., introduce a margin policy into our model. Specifically, the policy requires that when investors buy asset A on margin, they can at most borrow \hat{L} against each share of asset A as collateral. This margin policy, however, does not apply to the derivative market. As seen in our previous analysis, investors can design the derivative to be unfunded and are fully collateralized in equilibrium. It seems difficult for regulators to enforce restrictions on the collateral levels for over-the-counter derivatives.

Since our main interest is to assess the policy's impact on speculation, we compare the volatility of the wealth-share fluctuations in the following four economies. The first economy is the benchmark economy without financial innovation, which was analyzed in Section 6. The second economy has the margin policy imposed on the market for asset A , and is otherwise identical to the first economy. The third economy is the one with financial innovation, which was analyzed in Sections 3–5. Finally, the fourth economy has the margin policy imposed on the market for asset A , and is otherwise identical to the third economy.

Comparing the first two economies, we can assess the impact of the margin policy in an environment without financial innovation. In particular, we measure the policy impact by $Q^B \equiv \sigma^{BR}/\sigma^B$, where σ^B and σ^{BR} denote the standard deviation of the change in the wealth share of the optimists in economies 1 and 2, respectively. Note that the standard deviation of the change of wealth share of the optimists is the same as that of the pessimists, since the wealth shares of the two groups always add up to 1. Moreover, the measure Q^B does not depend on the belief we choose in calculating the standard deviation, as long as the belief “agrees” with the investors on the distribution of \tilde{U} and is non-degenerate on \tilde{V} (i.e., it assigns positive probabilities to V_u and V_d). Similarly, we can measure the policy impact in an environment with financial innovation by $Q \equiv \sigma^R/\sigma$, where σ and σ^R denote the standard deviation of the change in the wealth share of the optimists in the third and fourth economies, respectively.

We relegate the details of the equilibrium construction for economies 2 and 4 to the Appendix.

¹⁶See Alexandera et al. (2004) for a recent review.

In the following, for simplicity, we focus our discussion on the parameter region $\alpha_p \in [\underline{\alpha}_p, \bar{\alpha}_p]$, where $\underline{\alpha}_p$ and $\bar{\alpha}_p$ are constants and given in the Appendix. We focus on this region because the equilibria for all 4 economies are explicitly available. Numerical evaluations for other regions lead to qualitatively the same results.

Proposition 9 *Suppose $\alpha_p \in [\underline{\alpha}_p, \bar{\alpha}_p]$. If the margin policy is loose, $\hat{L} \geq V_d - \Delta$, it has no impact on the wealth fluctuation in the economy: $Q^B = Q = 1$. If the margin policy is tight, $\hat{L} < V_d - \Delta$, then $Q^B < Q < 1$. Moreover, Q^B , Q , and Q^B/Q all increase in \hat{L} .*

The proposition shows that when the margin policy is not binding ($\hat{L} \geq V_d - \Delta$), it has no impact on the equilibrium. Recall that Proposition 1 shows that investors prefer to borrow $V_d - \Delta$ against one unit of asset A . Since the margin policy allows investors to borrow more, it has no impact on investors' choices. When the margin policy is binding, however, it forces optimists to reduce leverage and so reduces the wealth fluctuation in the economy. Hence, both Q^B and Q are less than 1. When the margin policy gets tighter (i.e., \hat{L} decreases), it reduces the volatility in the economy (i.e., both Q^B and Q reduce).

More interestingly, the proposition shows that, in the presence of the derivative market, the margin policy is less effective in reducing wealth fluctuations—i.e., $Q^B < Q$. The intuition is as follows. When the regulator imposes a tight margin requirement, investors can no longer take large levered positions in asset A , leading to less volatility in the economy. In an economy with innovations, however, the impact on investors' overall speculative position is limited because investors can move to the derivative markets to set up their speculative positions. Similarly, when the margin policy loosens (\hat{L} increases), the economy without innovations responds more—i.e., Q^B increases more than Q . Therefore, Q^B/Q increases in \hat{L} .

It is interesting to note the coincidence that the Fed stopped actively adjusting its margin regulations in 1974 and the derivative markets started taking off at around that time. One possibility is that, as our model suggests, the existence of active derivative markets makes the margin policy less effective. Indeed, conversations with practitioners appear to suggest that active investors such as hedge funds barely view this margin policy as a binding constraint nowadays.

More generally, this analysis has implications on policies that focus on the spot markets only. In his Janeway Lectures in 1972, James Tobin proposed the idea of taxing transactions in foreign currency markets to contain speculations. Our model illustrates the limitation of the impact from this Tobin tax on transactions in the spot market. Transaction tax clearly makes it less appealing to speculate in the spot market. However, the impact of this tax on speculation is likely to be limited because investors can move their positions to derivative markets, where the transaction tax is less effective. To see this, note that investors can design the derivative to be unfunded, so that the tax is zero if it is levied as a fraction of the transaction price.

7.2 Financial Intermediation

Our prior analysis abstracts away from the role of financial intermediation—investors can simply introduce the securities they prefer. In this section, we explicitly model financial intermediation. In particular, we assume that only the financial intermediary (e.g., exchanges or brokers) can design

new securities. To model the objective of the financial intermediary, we follow Duffie and Jackson (1989) and assume that it can charge a fee for each contract investors trade, and their objective is to maximize the total fee by designing the contracts and setting the fee level. Taking the derivatives designed by the financial intermediary as given, investors maximize their expected wealth at $t = 1$ by choosing their portfolios at $t = 0$, and face the individual security margin constraint described in Section 3.5. The equilibrium in this economy is obtained when all investors and the financial intermediary optimize and all markets clear.

Since investors now have to pay a fee (e.g., bid-ask spread) when they trade the derivatives designed by the financial intermediary, the equilibrium prices in this economy are generally different from those in Sections 3–5. How does this financial intermediation affect the design of the derivative contracts? We first consider the case in which the intermediary has monopoly power.

Proposition 10 *In the case of $\beta = 0$, asset V is an optimal contract for the monopolistic financial intermediary.*

This proposition shows that in equilibrium, the profit-optimizing financial intermediary chooses the same derivative as in the model where investors can design the contract themselves. As shown earlier, given the collateral friction in Section 3.5, investors find asset V most appealing. Hence, the financial intermediary can maximize the fees from investors if it introduces the market for asset V . If it introduces an alternative derivative contract, the financial intermediary won't be able to charge as much in fees since investors find alternative contracts less useful. In other words, this monopolistic financial intermediary does not affect the design of the derivative, which is still the one investors want most. However, the intermediary can extract all the surplus from this innovation.

This result is in contrast to that in Duffie and Jackson (1989), where the intermediary may choose a contract that is inefficient from investors' perspective but maximizes the intermediary's revenue. The key difference is that in our model, the fee charged by the intermediary is endogenous: if the contract is more appealing to investors, the intermediary can charge a higher fee. In Duffie and Jackson (1989), the fee is set exogenously, so that the intermediary does not have the incentive to design the security that is most appealing to investors, but focuses on maximizing the trading volume instead. Proposition 10 focuses on the case of $\beta = 0$. This is because that, in the case of $\beta > 0$, the financial intermediation affects investors' initial endowment through its impact on the price of asset A . We set $\beta = 0$ to shut down this endowment effect.¹⁷

Finally, in the above discussion, the financial intermediary faces no competition. It is easy to see that competition would drive down the fee that the financial intermediary can charge. In the extreme case in which the financial intermediary faces no cost in designing securities, competition will drive the fee to zero and the equilibrium will converge to the one analyzed in Sections 3–5. In other words, competition among intermediaries does not affect the design of the derivative either. However, competition reallocates the surplus from the innovation from the intermediaries to investors.

¹⁷For many cases (e.g. the design of swap contracts), it seems far fetched to believe this endowment effect is an important consideration of financial intermediaries. However, this endowment effect may be important if the intermediary has a significant position in asset A . For example, if the intermediary is involved in both packaging mortgage-backed securities (MBS) and designing mortgage-related derivatives, he naturally would take into account the impact of his design of mortgage-related derivatives on the MBS he sells. We leave this to a separate study.

8 General Discussions

8.1 Common Theme Behind Various Innovations

This collateral view highlights the common theme behind a variety of financial innovations, despite their strikingly different appearances. For example, many successful derivative contracts are unfunded and allow investors to take large positions with very little collateral. In addition to the previously mentioned swaps, futures are another example. A futures contract allows investors to take an exposure to the fluctuations of the price of a certain asset without the physical process of buying or selling the asset. This is especially important for commodity futures, where transactions are costly and time consuming. In other words, similar to the case for swaps, futures contracts save collateral by isolating the variables that investors want to bet on (i.e., the underlying asset prices).

This collateral view of innovation is not restricted to the invention of new securities. It also sheds light on the evolution of a legal practice, the *de facto* superseniority of derivatives and repos. Although derivatives and repos are not supersenior in a strict statutory sense, it has been a common practice in the U.S.: When a company goes bankrupt, its repo and derivative counterparties can simply seize the collateral posted in the transactions up to the amount the company owes them, instead of going through the lengthy and costly bankruptcy procedure. This exceptional treatment accorded derivatives and repos in bankruptcy is recent and has been evolving over time. In the context of our analysis, this practice can be viewed as carving unwanted cash flows out of derivative and repo transactions: Suppose an investor enters an interest rate swap and his goal is to hedge or speculate on interest rate risk, rather than taking an exposure to his counterparty's credit risk. Without this superseniority, even if his counterparty posts a large amount of collateral, the investor would still not feel safe since he would have to go through automatic stay when his counterparty defaults. With superseniority, however, the investor can immediately seize the collateral upon default, and so can be better protected even by a smaller amount of collateral. In other words, superseniority separates the counterparty's assets into two parts, the collateral posted to the swap transaction and other assets. Given the investor's purpose, the second part is unwanted. The collateral efficiency is achieved when those unwanted cash flows are carved out of the swap transaction by superseniority.¹⁸

Financial innovation may also take the form of legal entities. For example, special purpose vehicles (SPVs) have become prevalent in recent decades with the rise of securitization. In the context of our analysis, we can view creating an SPV as, again, carving out unwanted cash flows (i.e., the firm's assets other than those allocated to the SPV). This interpretation is similar to the theory proposed in Gorton and Souleles (2006), which emphasizes the benefit of making SPVs bankruptcy remote to avoid bankruptcy cost.

The collateral friction in our model arises from the imperfection in cross-netting. In practice, it is becoming increasingly possible to have limited cross-netting. For example, on December 12, 2006, the SEC approved a rule change, which made limited cross-netting available to some investors in the exchange-traded options market. Another example is the efforts from brokers and hedge funds

¹⁸Our focus here is the collateral efficiency from superseniority for derivative traders. We do not attempt to evaluate the overall impact of this practice. See Bolton and Oehmke (2012) for an analysis of the impact of superseniority on corporate policies. Duffie and Skeel (2012) offer more broad discussions on the pros and cons of this practice.

that attempt to get around the regulation-induced margin requirements (see, e.g., Brunnermeier and Pedersen (2009)). One can view the continuing efforts by regulators and market participants in modifying the margin procedure as one form of collateral-motivated financial innovation. Their goal is simply to satisfy the demand from market participants to alleviate their collateral constraints.

8.2 Open Issues

In practice, collateral is likely to play many more roles than what is captured in the above model. For example, our model assumes that all investors can handle the procedure of posting collateral. In practice, some derivative users may find it more costly to deal with posting collateral and daily marking-to-the-market. For example, corporations may find it onerous to deal with these issues and the induced cash flow uncertainty. As noted in Ross (2011), many over-the-counter (OTC) derivatives are very similar to those traded on exchanges, and the volume in the OTC markets is much larger. One conjecture is that the collateral flexibility for OTC contracts might be the key. According to the data from the Federal Reserve Board, for the credit exposure in their OTC derivative positions with corporations, a bank or broker on average demanded only 0.1% collateral at the end of 2010. In contrast, it demanded 72% collateral for its exposures to hedge funds and 45% for exposures to other banks or brokers. This tremendous variation simply highlights many unanswered questions. What is the role of collateral in this financial system with heterogeneous institutions? What is the optimal innovation in this more complex world? Analysis of these issues is likely to shed light on the role of OTC markets for the macroeconomy. It also helps to understand the impact of the policy that moves all OTC derivatives to centralized markets.

Another open issue is the impact of financial innovation on market liquidity. For example, Dang, Gorton, and Homstrom (2011) show that information-insensitive securities discourage information production, which avoids adverse selection and hence is beneficial for liquidity provision. That is, “ignorance is a bliss for liquidity.” Our analysis, however, shows that financial innovation helps investors to take larger positions. This naturally encourages investors to produce more information, and so may jeopardize some of the liquidity benefits from information-insensitive securities. We leave these issues to a separate study.

9 Conclusion

This paper proposes a collateral view of financial innovation: Many successful innovations, despite their strikingly different appearances, share the common motive of reducing collateral requirements to facilitate trading. We illustrate this insight in an equilibrium model in which both the financial innovation and collateral requirements are endogenous. In the presence of a collateral friction in cross-netting, the optimal security is the one that isolates the variable with disagreement. If one introduced any other derivatives, those markets would not generate any trading. The model not only has a number of asset-pricing implications that are broadly consistent with existing empirical evidence, but also generates some new testable predictions. Our model also shows that financial innovations can limit the impact of policies that target spot markets only (e.g., Regulation T, Tobin tax) because investors can move to derivative markets to set up similar positions.

Appendix

Proof of Proposition 1

We first conjecture that the equilibrium is as follows: $x_o \in [0, \alpha_o)$ optimists invest all their wealth in a levered long position in asset V and the remaining, $\alpha_o - x_o$, invest all their wealth in a levered long position in asset A , and all pessimists short asset V . Moreover, using each unit of asset A as collateral, an optimist borrows $V_d - \Delta$, and the interest rate is 0. Using each contract V as collateral, an optimist borrows V_d and the interest rate is 0. To short each contract, a pessimist needs to post V_u cash as collateral. We then derive the market-clearing prices under this conjecture before verifying that the above conjecture is indeed sustained in equilibrium.

To hold one unit of asset V , an investor needs $P_V - V_d$ capital since he can use the asset as collateral to borrow V_d . So the aggregate demand from x_o optimists is $x_o \frac{e + \beta P_A}{P_V - V_d}$. Similarly, pessimists' aggregate short position in asset V is $\alpha_p \frac{e + \beta P_A}{V_u - P_V}$. So the market-clearing condition in the market for asset V is:

$$x_o \frac{e + \beta P_A}{P_V - V_d} = \alpha_p \frac{e + \beta P_A}{V_u - P_V}. \quad (16)$$

Similarly, the market-clearing condition in the market for asset A is:

$$(\alpha_o - x_o) \frac{e + \beta P_A}{P_A - (V_d - \Delta)} = \beta. \quad (17)$$

Moreover, the expected utility for an optimist to borrow V_d to hold one share of asset V is $\mathbf{E}_o[\tilde{V}] - V_d$. So the expected utility from investing one dollar in this levered position in asset V is $\frac{\mathbf{E}_o[\tilde{V}] - V_d}{P_V - V_d}$. Similarly, the expected utility from investing one dollar in the levered position in asset A is $\frac{\mathbf{E}_o[\tilde{A}] - (V_d - \Delta)}{P_A - (V_d - \Delta)}$. Optimists are indifferent about these two strategies:

$$\frac{\mathbf{E}_o[\tilde{V}] - V_d}{P_V - V_d} = \frac{\mathbf{E}_o[\tilde{A}] - (V_d - \Delta)}{P_A - (V_d - \Delta)}. \quad (18)$$

Similarly, pessimists' expected utility from a one-dollar investment in shoring asset V is

$$\frac{V_u - \mathbf{E}_p[\tilde{V}]}{V_u - P_V}. \quad (19)$$

From (16)–(18), we obtain (10), (11), and $x_o = \beta / (\beta + \gamma)$.

To verify that the conjecture is an equilibrium, we need to show the following:

- (a) No investor prefers to invest in the riskless technology.
- (b) Pessimists prefer to short V rather than shorting A .
- (c) Optimists prefer to borrow $V_d - \Delta$ against each unit of asset A as collateral.
- (d) Optimists prefer to borrow V_d against each unit of asset V as collateral.

Equation (11) implies that $\mathbf{E}_o[\tilde{V}] < P_V < \mathbf{E}_o[\tilde{A}]$. Therefore, trading V strictly dominates the investment in the riskless technology, implying (a). It is also straightforward to verify (b) by directly calculating the expected utility from shorting V and shorting A .

Using each unit of asset A as collateral, optimists would not borrow less than $Y_d - \Delta$. This is because they can borrow more at 0 interest rate and their expected return for the investment in asset A is positive. Hence, to prove (c), we just need to verify that optimists would not choose to borrow more than $Y_d - \Delta$. Note that if they borrow more than $Y_d - \Delta$, they have to compensate the lender by offering a higher interest rate. Equations (18) and (19) imply that if optimists and pessimists agree on the loan, the following two inequalities have to hold

$$\frac{\mathbf{E}_p \left[\min \left\{ \tilde{A}, L(1+r) \right\} \right]}{L} \geq \frac{V_u - \mathbf{E}_p[\tilde{V}]}{V_u - P_V}, \quad (20)$$

$$\frac{\mathbf{E}_o \left[\max \left\{ \tilde{A} - L(1+r), 0 \right\} \right]}{P_A - L} \geq \frac{\mathbf{E}_o[\tilde{V}] - V_d}{P_V - V_d}, \quad (21)$$

where r is the notional interest rate in the loan contract. The left-hand side of (20) is pessimists' expected return from lending, and the right-hand side is their expected return from shorting asset V . Similarly, the left-hand side of (21) is optimists' expected return from the levered position in A , and the right-hand side is their expected return from the levered position in asset V .

By changing the inequalities in (20) and (21) into equalities, we obtain an equation system of L and r . We show in the online Appendix that there exists a unique value α^* , $0 < \alpha^* < 1$, such that if $\alpha_p = \alpha^*$ there is a unique solution for this equation system. We define

$$\alpha_1 \equiv \alpha^*. \quad (22)$$

Moreover, if $\underline{\alpha} \leq \alpha_p < \alpha_1$, inequalities (20) and (21) cannot hold simultaneously for any values of L and r —i.e., optimists and pessimists cannot agree on any loan contract. This verifies (c). The proof for (d) is similar.

Proof of Proposition 2

We offer here an intuitive proof, and leave the algebra to the online Appendix. Asset V allows investors to transfer all their $t = 1$ wealth to the states they prefer. So it weakly dominates all other instruments. Moreover, if the derivative K is not perfectly correlated with V , traders in this market cannot all transfer wealth to the states they prefer. Suppose, in equilibrium, asset K generates some trading. Then, it is possible to Pareto-improve the welfare of those traders if they simply move to the market for asset V , a contradiction.

Proof of Proposition 3

Suppose X is an unfunded security. The flows of funds for those investors who trade X are the following: There is no need to transfer funds across investors at $t = 0$ since $P_X = 0$. At $t = 1$, the dealer just needs to transfer all the wealth of short sellers to the long side if V_u is realized, or all the wealth from the long side to short sellers if V_d is realized. Now, suppose $P_X \neq 0$. Then the fund flows induced by trading X are those in the above case with an unfunded security, plus a “round trip” for P_X , i.e., transferring P_X from the long side to the short side at $t = 0$ and then transferring it back at $t = 1$. Hence, the total transfer in the case with a funded security is always higher than or equal to that in the case with an unfunded security. Equality occurs when the long

side happens to lose all his wealth to the short sellers at $t = 1$. Moreover, investors may have to borrow to trade a funded security and so induce even more fund flows. The flows induced by trading other securities are not affected by contract X . Therefore, the total transfer induced by a funded security is always higher than or equal to that induced by an unfunded one.

Proof of Proposition 4

Differentiating S leads to all results except those in item 4. To prove item 4, we derive the equilibrium prices when the total supply of asset A is $\beta + \beta^*$. Results in item 4 can then be obtained by differentiating S in this new equilibrium.

Proof of Propositions 5 and 6.

The proof is similar to that of Proposition 1. We first calculate the equilibrium prices based on the portfolio holdings described in items 2 and 3. Market-clearing conditions for assets A and V imply:

$$x_o^* \frac{e + \beta P_A}{P_V - V_d} = x_p^* \frac{e + \beta P_A}{V_u - P_V}. \quad (23)$$

$$y_o^* \frac{e + \beta P_A}{P_A - (V_d - \Delta)} + (\alpha_o - x_o^* - y_o^*) \frac{e + \beta P_A}{P_A - L^*} = \beta. \quad (24)$$

Suppose the loan contract is such that optimists borrows L^* against each unit of asset A and promise to pay back Y^* . The market-clearing condition for the loan market is

$$(\alpha_o - x_o^* - y_o^*) \frac{e + \beta P_A}{P_V - L^*} L^* = (\alpha_p - x_p^*) (e + \beta P_A). \quad (25)$$

Optimists being indifferent about the three strategies in item 2 of Proposition 5 implies

$$\frac{\mathbf{E}_o \tilde{V} - V_d}{P_V - V_d} = \frac{\mathbf{E}_o \tilde{A} - (V_d - \Delta)}{P_A - (V_d - \Delta)}, \quad (26)$$

$$\frac{\mathbf{E}_o \tilde{V} - V_d}{P_V - V_d} = \frac{\mathbf{E}_o [\max(\tilde{A} - Y^*, 0)]}{P_A - L^*}. \quad (27)$$

Pessimists being indifferent about the two strategies in item 3 of Proposition 5 implies

$$\frac{V_u - \mathbf{E}_p \tilde{V}}{V_u - P_V} = \frac{\mathbf{E}_p [\min(Y^*, \tilde{A})]}{L^*}. \quad (28)$$

From equation (27), we obtain Y^* as a function of L^* . Denote it as $Y^* = f_1(L^*)$. An optimist prefers to borrow if

$$Y^* \leq f_1(L^*). \quad (29)$$

Similarly, from equation (28), we obtain Y^* as a function of L^* . Denote it as $Y^* = f_2(L^*)$. A pessimist prefers to lend if

$$Y^* \geq f_2(L^*). \quad (30)$$

One necessary condition for L^* and Y^* to satisfy both (29) and (30) is

$$f_1'(L^*) = f_2'(L^*). \quad (31)$$

Rearranging equations (24)–(28) and equation (31), we obtain a seven-equation system: equations (12)–(13) and the following five

$$x_o^* = \frac{\frac{e+\beta(V_d-\Delta)}{e+\beta P_A} - \alpha_p \frac{V_d-\Delta}{L^*}}{1 + z^* \left(1 - \frac{V_d-\Delta}{L^*}\right)}, \quad (32)$$

$$x_p^* = z^* x_o^*, \quad (33)$$

$$L^* = \frac{1}{1 - h_p} \frac{z^*}{z^* + 1} \mathbf{E}_p \left[\min \left\{ \tilde{A}, Y^* \right\} \right], \quad (34)$$

$$Y^* = V_d + F_D^{-1} \left(\frac{1}{1 - h_p} \frac{1 - h_p - h_o z^*}{1 - h_o - h_o z^*} \right), \quad (35)$$

$$y_o^* = \frac{1}{z^* + 1} \frac{V_u - V_d + \frac{\Delta}{h_o}}{L^* - (V_d - \Delta)} \left[\frac{(1 + z^*) \left(1 - \frac{V_d-\Delta}{L^*}\right)}{1 + z^* \left(1 - \frac{V_d-\Delta}{L^*}\right)} \alpha_o - \frac{\beta(P_A - L^*)}{e + \beta P_A} - \frac{\frac{e+\beta(V_d-\Delta)}{e+\beta P_A} - \frac{V_d-\Delta}{L^*}}{1 + z^* \left(1 - \frac{V_d-\Delta}{L^*}\right)} \right]. \quad (36)$$

Define α_2 as

$$\alpha_2 \equiv 1 - \frac{\frac{\beta(P_A - L^*)}{e + \beta P_A} \left[1 + z^* \left(1 - \frac{V_d-\Delta}{L^*}\right)\right] + \frac{e+\beta(V_d-\Delta)}{e+\beta P_A} - \frac{V_d-\Delta}{L^*}}{(1 + z^*) \left(1 - \frac{V_d-\Delta}{L^*}\right)}. \quad (37)$$

In the online Appendix, we show that in the case of $\alpha_1 \leq \alpha_p < \alpha_2$, the equation system has a unique solution. The notional interest rate in equilibrium is then

$$r(L^*, A) = \frac{Y^*}{L^*} - 1. \quad (38)$$

The proof of Proposition 6 is similar. Now the equation system has one less equation since optimists have two, rather than three, strategies. Following the same logic, we obtain

$$L^{**} = \frac{1}{1 - h_p} \frac{x_p^{**}}{x_o^{**} + x_p^{**}} \mathbf{E}_p \left[\min \left(\tilde{A}, Y^{**} \right) \right], \quad (39)$$

$$x_o^{**} = \frac{\alpha_o h_o e}{h_o e + \beta \mathbf{E}_o \left[\max \left(\tilde{A} - Y^{**}, 0 \right) \right]}, \quad (40)$$

$$x_p^{**} = \frac{\alpha_p (1 - h_p) e}{(1 - h_p) e + \beta \mathbf{E}_p \left[\min \left(\tilde{A}, Y^{**} \right) \right]}, \quad (41)$$

and Y^{**} is the unique positive solution to

$$\frac{x_p^{**}}{x_o^{**}} = \frac{1 - h_p}{h_o} \frac{1 - (1 - h_o) F(Y^{**} - V_d)}{1 - (1 - h_p) F(Y^{**} - V_d)}. \quad (42)$$

Hence, the notional interest rate in equilibrium is

$$r(L^{**}, A) = \frac{Y^{**}}{L^{**}} - 1. \quad (43)$$

Proof of Propositions 7 and 8

The calculation of the equilibrium in the benchmark economy without the derivative contract V is similar to that on Propositions 1, 5, and 6 and is reported in the online Appendix. To prove

Proposition 7, it is sufficient to see that in a subset of $\alpha_p \in [\underline{\alpha}^B, \alpha_1^B]$, the price of asset A in this benchmark economy, P_A^B , is given by

$$P_A^B = \frac{e\alpha_o + (\gamma^B + \beta)(V_d - \Delta)}{\gamma^B + \beta\alpha_p}, \quad \text{where } \gamma^B = \frac{e + \beta(V_d - \Delta)}{V_u - V_d + 2\Delta},$$

and the expected utility of an optimistic investor, J_o^B , is given by

$$J_o^B = (e + \beta P_A^B) \frac{\mathbf{E}_o[\tilde{A}] - (V_d - \Delta)}{P_A^B - (V_d - \Delta)} = \frac{h_o}{\alpha_o} [e + \beta(V_d - \Delta)] \frac{\gamma^B + \beta}{\gamma^B} \frac{V_u - V_d + \frac{\Delta}{h_o}}{V_u - V_d + 2\Delta}.$$

In the economy with the derivative V the price of asset A , P_A is given by (10), and the expected utility for an optimistic investor, J_o , is given by

$$J_o = (e + \beta P_A) \frac{\mathbf{E}_o[\tilde{V}] - V_d}{P_V - V_d} = \frac{h_o}{\alpha_o} [e + \beta(V_d - \Delta)] \frac{\gamma + \beta}{\gamma}.$$

It is straightforward to see that in the case of $\alpha_p \in [\underline{\alpha}^B, \alpha_1^B] \cap [\underline{\alpha}, \alpha_1]$

$$P_A \gtrless P_A^B \text{ iff } h_o \lesseqgtr \frac{1}{2},$$

$$J_o \gtrless J_o^B \text{ iff } h_o \gtrless \frac{1}{2}.$$

Similarly, we can calculate the pessimist's welfare in the economy with and without the derivative V , J_p , and J_p^B , and obtain

$$J_p \gtrless J_p^B \text{ iff } h_p \lesseqgtr \frac{1}{2}.$$

That is, the introduction of asset V has a mixed impact price and welfare.

Proof of Proposition 9

Let ω_o and ω_p be the $t = 1$ wealth share for optimists and pessimists, respectively. Since $1 = \omega_o + \omega_p$, we have $\omega_o - \mathbf{E}\omega_o = \mathbf{E}\omega_p - \omega_p$, where the expectations are taken with respect to a certain probability measure that assigns the same distribution to \tilde{U} as \mathcal{P}_o and \mathcal{P}_p , but assigns $\tilde{V} = V_u$ a probability h and $\tilde{V} = V_d$ a probability $1 - h$, with $0 < h < 1$. It thus follows that the variance of the wealth share for the two groups are identical—i.e., $\mathbf{Var}(\omega_o) = \mathbf{Var}(\omega_p)$.

Similar to the analysis for economy 1 (analyzed in Section 6) and 3 (analyzed in Section 3–5), we construct the equilibria for economies 2 and 4 in the online Appendix. For economies 1 through 4, we obtain closed-form expressions for the equilibria when α_p is in the interval $[\underline{\alpha}^B, \alpha_1^B]$, $[\underline{\alpha}^{BR}, \bar{\alpha}^{BR}]$, $[\underline{\alpha}, \alpha_1]$, and $[\alpha_1^R, \alpha_2^R]$, respectively. Denote the intersection of the four intervals by $[\underline{\alpha}_p, \bar{\alpha}_p]$.

If $\hat{L} \geq V_d - \Delta$, the margin policy never binds and hence had no impact on the equilibria. Therefore, we have $Q^B = Q = 1$. If $\hat{L} < V_d - \Delta$, the margin policy always binds. Using ω_p^B , ω_p^{BR} , ω_p , and ω_p^R to refer to the wealth share of pessimists in economies 1 through 4, respectively, we obtain

$$\omega_p^B = \vartheta^B \omega_p, \text{ where } \vartheta^B = \frac{(1 - h)(V_u - V_d) + \Delta}{V_u - V_d + 2\Delta} < 1.$$

$$\omega_p^{BR} = \vartheta^{BR} \omega_p, \text{ where } \vartheta^{BR} = \frac{(1-h)(V_u - V_d) + \Delta}{V_u + \Delta - C_A} \vartheta^R < \vartheta^R.$$

$$\omega_p = \begin{cases} 0 \\ \frac{e + \beta(V_d - \Delta)}{e + \beta(V_d + \varepsilon)} \end{cases}, \text{ with probability } \begin{matrix} h \\ 1-h \end{matrix}.$$

$$\omega_p^R = \vartheta^R \omega_p, \text{ where } \vartheta^R = \frac{e + \beta C_A}{e + \beta(V_d - \Delta)} < 1.$$

We then have $Q^B = \frac{\sigma^{BR}}{\sigma^B} = \frac{V_u - V_d + 2\Delta}{V_u + \Delta - \hat{L}} \frac{\sigma^R}{\sigma} < \frac{\sigma^R}{\sigma} = Q$. Note that the conclusion that $Q^B < Q$ does not depend on the value of h .

Proof of Proposition 10

Suppose the intermediary issues a derivative

$$\tilde{Y} = \begin{cases} Y_u + \tilde{\eta}_u & \text{if } \tilde{V} = V_u, \\ Y_d + \tilde{\eta}_d & \text{if } \tilde{V} = V_d, \end{cases}$$

where $\mathbf{E}[\tilde{\eta}_u | \tilde{V} = V_u] = \mathbf{E}[\tilde{\eta}_d | \tilde{V} = V_d] = 0$. Let the support of $\tilde{\eta}_u$ and $\tilde{\eta}_d$ be $[-\underline{\eta}_u, \bar{\eta}_u]$ and $[-\underline{\eta}_d, \bar{\eta}_d]$, respectively. The financial intermediary offers a bid price P_Y and an ask price $P_Y + \varepsilon$. Hence, the intermediary earns the bid-ask spread ε for each pair of long-short positions in the derivative.

Suppose that in equilibrium a fraction x_o of optimists and a fraction x_p of pessimists would like to trade this derivative. We obtain the aggregate trading volume M :

$$M = x_o \frac{e}{P_Y + \varepsilon - (Y_d - \underline{\eta}_d)} = x_p \frac{e}{Y_u + \bar{\eta}_u - P_Y},$$

From this, we obtain

$$P_Y = \frac{x_o}{x_o + x_p} (Y_u + \bar{\eta}_u) + \frac{x_p}{x_o + x_p} (Y_d - \underline{\eta}_d - \varepsilon). \quad (44)$$

Hence, the trading volume and the intermediary's revenue are

$$M = \frac{(x_o + x_p) e}{Y_u - Y_d + \bar{\eta}_u + \underline{\eta}_d + \varepsilon},$$

$$\varepsilon M = \frac{\varepsilon}{Y_u - Y_d + \bar{\eta}_u + \underline{\eta}_d + \varepsilon} (x_o + x_p) e.$$

The expected return for the two groups of investors are:

$$J_o(\tilde{Y}) = \frac{\mathbf{E}_o[\tilde{Y}] - (Y_d - \underline{\eta}_d)}{P_Y + \varepsilon - (Y_d - \underline{\eta}_d)}, \quad J_p(\tilde{Y}) = \frac{(Y_u + \bar{\eta}_u) - \mathbf{E}_p[\tilde{Y}]}{(Y_u + \bar{\eta}_u) - P_Y}.$$

The monopolistic intermediary's problem is then to maximize his total revenue subject to investor's participation constraint and market-clearing conditions:

$$\max_{\varepsilon} \varepsilon M, \text{ s.t. } J_o(\tilde{Y}) \geq 1, J_p(\tilde{Y}) \geq 1 \text{ and } P_Y \text{ is determined by (44).}$$

The solution to this optimization program is $\varepsilon = (h_o - h_p)(Y_u - Y_d)$. The equilibrium is as follows.

In case 1: $\alpha_p \leq \frac{(1-h_p)(Y_u-Y_d)+\bar{\eta}_u}{(1-h_p+h_o)(Y_u-Y_d)+\bar{\eta}_u+\underline{\eta}_d}$,

$$\begin{aligned} x_p &= \alpha_p, \\ x_o &= \frac{h_o(Y_u - Y_d) + \underline{\eta}_d}{(1 - h_p)(Y_u - Y_d) + \bar{\eta}_u} \alpha_p, \\ \varepsilon M &= \frac{(h_o - h_p)(Y_u - Y_d)}{(1 - h_p)(Y_u - Y_d) + \bar{\eta}_u} \alpha_p e. \end{aligned}$$

In case 2: $\alpha_p > \frac{(1-h_p)(Y_u-Y_d)+\bar{\eta}_u}{(1-h_p+h_o)(Y_u-Y_d)+\bar{\eta}_u+\underline{\eta}_d}$,

$$\begin{aligned} x_o &= \alpha_o, \\ x_p &= \frac{(1 - h_p)(Y_u - Y_d) + \bar{\eta}_u}{h_o(Y_u - Y_d) + \underline{\eta}_d} (1 - \alpha_p), \\ \varepsilon M &= \frac{(h_o - h_p)(Y_u - Y_d)}{h_o(Y_u - Y_d) + \underline{\eta}_d} (1 - \alpha_p) e. \end{aligned}$$

In both cases, we have $J_o(\tilde{Y}) = J_p(\tilde{Y}) = 1$. It is straight forward to see that setting $\bar{\eta}_u = \underline{\eta}_u = \bar{\eta}_d = \underline{\eta}_d = 0$ (i.e., turning asset Y into asset V) maximizes the intermediary's revenue.

Careful readers may have noticed that although asset V is optimal for the intermediary, it is not unique. In case 1, for example, setting $\bar{\eta}_u = \underline{\eta}_u = 0$ would be enough. This is because there are “too many” optimists. Even if the security is not appealing to them (i.e., $\bar{\eta}_d$ and $\underline{\eta}_d$ are not 0), plenty of them would still trade it. Similarly, there are “too many” pessimists in case 2, the monopolistic intermediary does not have to “serve them well” (i.e., $\bar{\eta}_u$ and $\underline{\eta}_u$ don't have to be 0). The optimal security will be uniquely determined (up to a constant) if the economic environment is not this extreme (In the model, the intermediary can extract all the surplus from trading, so that all investors' expected returns are 0). For example, if investors have some outside option and so their participation constraint is that their expected return is positive, rather than strictly 0. In this case, the intermediary has to adopt V to appeal all investors.

References

- Alexandera, Gordon, Evren Ors, Mark Peterson, and Paul Seguin, 2004, Margin regulation and market quality: a microstructure analysis, *Journal of Corporate Finance*, 10, 549–574.
- Allen, Franklin and Douglas Gale, 1988, Optimal Security Design, *Review of Financial Studies* 1, 229–263.
- Allen, Franklin and Douglas Gale, 1994, Financial Innovation and Risk Sharing, MIT Press
- Alvarez, F., and U. Jermann, 2000, Efficiency, Equilibrium, and Asset Pricing with Risk of Default, *Econometrica*, 68, 775–798.
- Arora, Navneet, Priyank Gandhi, and Francis Longstaff, 2010, Counterparty Credit Risk and the Credit Default Swap Market, working paper.
- Banerjee, Snehal and Jeremy Graveline, 2011, Trading in Derivatives When the Underlying is Scarce, working paper.
- Basak, Suleyman and Benjamin Croitoru, 2000, Equilibrium mispricing in a capital market with portfolio constraints, *Review of Financial Studies* 13, 715–748.
- Benston George and Clifford Smith, Jr., 1976, A Transactions Cost Approach to the Theory of Financial Intermediation, *Journal of Finance*, 31 215–231.
- Bhamra, Harjoat and Raman Uppal, 2009, The Effect of Introducing a Non-Redundant Derivative on the Volatility of Stock-Market Returns When Agents Differ in Risk Aversion, *Review of Financial Studies* 22, 2303–2330.
- Biais, Bruno, Jean-Charles Rochet and Paul Woolley, 2010, Innovations, Rents and Risk, working paper.
- Bolton, Patrick and Martin Oehmke, 2012, Should Derivatives Be Privileged in Bankruptcy?, working paper.
- Brunnermeier, Markus and Lasse Pedersen, 2009, Market Liquidity and Funding Liquidity, *Review of Financial Studies*, 22, 2201–2238.
- Brunnermeier, Markus, Alp Simsek, and Wei Xiong, 2011, A Welfare Criterion for Models with Heterogeneous Beliefs, working paper.
- Chien, YiLi and Hanno Lustig, 2010, The Market Price of Aggregate Risk and the Wealth Distribution, *Review of Financial Studies* 23, 1596–1650.
- Culp, Christopher and Merton Miller, 1995, Metallgesellschaft and the Economics of Synthetic Storage. *Journal of Applied Corporate Finance* 7, 62–76.
- Dang, Tri Vi, Gary Gorton, Bengt Homstrom, Opacity and the Optimality of Debt for Liquidity Provision, working paper.
- DeMarzo, Peter and Darrell Duffie, 1999, A Liquidity-Based Model of Security Design, *Econometrica* 67, 65–99.
- Duffie, Darrell and Matthew Jackson, 1989, Optimal Innovation of Futures Contracts, *Review of Financial Studies* 2, 275–296.

- Duffie, Darrell and Rohit Rahi, 1995, Financial Market Innovation and Security Design: An Introduction, *Journal of Economic Theory* 65, 1-42
- Duffie, Darrell and David Skeel, 2012, A Dialogue on the Costs and Benefits of Automatic Stays for Derivatives and Repurchase Agreements, working paper.
- Dubey, P., J. Geanakoplos, and M. Shubik, 2005 Default and Punishment in General Equilibrium, *Econometrica*, 73(1): 1-37.
- Fleckenstein, Matthias, Francis A. Longstaff, and Hanno Lustig, 2010, Why does the Treasury issue TIPS? The TIPS-Treasury bond puzzle, working paper.
- Fortune, Peter, 2000, Margin Requirements, Margin Loans, and Margin Rates: Practice and Principles, *New England Economic Review*, Sep-Oct, 20–44.
- Fostel A., Geanakoplos J. 2008. Leverage Cycles and the Anxious Economy, *American Economic Review* 98, 1211–1244.
- Frazzini, Andrea and Lasse Pedersen, 2011, Embedded Leverage, working paper.
- Garleanu, Nicolae and Lasse Pedersen, 2011, Margin-Based Asset Pricing and Deviations from the Law of One Price, working Paper.
- Geanakoplos, John, 1997, Promises, Promises, in *The Economy as an Evolving Complex System II*, ed. W. Brian Arthur, Steven Durlauf, and David Lane, 285–320, Reading, MA: Addison-Wesley.
- Geanakoplos, John. 2003. Liquidity, Default, and Crashes: Endogenous Contracts in General Equilibrium. in *Advances in Economics and Econometrics: Theory and Applications*, Eighth World Conference, Vol. 2, 170–205, Econometric Society Monographs.
- Geanakoplos, John, 2009, The Leverage Cycle, NBER Macroeconomics Annual.
- Geczy, Christopher, David Musto, and Adam Reed, 2002, Stocks are Special Too: An Analysis of the Equity Lending Market, *Journal of Financial Economics* 66, 241–269.
- Gennaioli, Nicola, Andrei Shleifer, Robert Vishny, 2011, Neglected risks, financial innovation, and financial fragility, working paper.
- Gorton, Gary and George Pennacchi, 1990, Financial Intermediaries and Liquidity Creation, *Journal of Finance* 45, 49–72.
- Gorton, Gary and Nicholas Souleles, 2006, “Special Purpose Vehicles and Securitization” in *The Risks of Financial Institutions*, ed. Rene Stulz and Mark Carey, University of Chicago Press.
- Gromb, Denis and Dimitri Vayanos, 2002, Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs, *Journal of Financial Economics*, 66, 361–407.
- He, Zhiguo, and Wei Xiong, 2009, Multi-market Delegated Asset Management, working paper.
- He, Zhiguo, and Wei Xiong, 2010, Equilibrium Debt Financing, working paper.
- Jiang, Wenxi and Hongjun Yan, 2012, Levered ETFs, working paper.
- Kehoe, T., and D. K. Levine, 1993, Debt Constrained Asset Markets, *Review of Economic Studies*, 60, 856–888.

- Krishnamurthy, Arvind, and Annette Vissing-Jorgensen, 2010, The Aggregate Demand for Treasury Debt, working paper.
- Kiyotaki, Nobuhiro and John Moore, 1997, Credit Cycles, *Journal of Political Economy* 105, 211–248.
- Liu, Jun, and Francis Longstaff, 2004, Losing Money on Arbitrages: Optimal Dynamic Portfolio Choice in Markets with Arbitrage Opportunities, *Review of Financial Studies* 17, 611–641.
- Merton, Robert, 1989, On the Application of the Continuous-Time Theory of Finance to Financial Intermediation and Insurance, *Geneva Papers on Risk and Insurance*, 225–262
- Miller, Merton, 1986, Financial Innovation: The Last Twenty Years and the Next, *Journal of Financial and Quantitative Analysis* 21, 459–471.
- Mitchell, Mark and Todd Pulvino, 2011, Arbitrage crashes and the speed of capital, *Journal of Financial Economics*, forthcoming.
- Rampini, Adriano and S. Viswanathan, 2010, Collateral, risk management, and the distribution of debt capacity, *Journal of Finance* 65, 2293–2322.
- Rahi, Rohit and Jean-Pierre Zigrand, 2009, Strategic Financial Innovation in Segmented Markets, *Review of Financial Studies* 22, 2941–2971.
- Ross, Stephen, 1989, Institutional Markets, Financial Marketing, and Financial Innovation, *Journal of Finance*, 44, 541–556.
- Ross, Stephen, 2011, The Recovery Theorem, working paper.
- Santos, Tano and Jose Scheinkman, 2001, Competition Among Exchanges, *Quarterly Journal of Economics*, 1027–1061.
- Simsek, Alp, 2011, Belief Disagreements and Collateral Constraints, working paper.
- Simsek, Alp, 2012, Speculation and Risk Sharing with New Financial Assets, working paper.
- Tufano, Peter, 2003, Financial Innovation, in Handbook of the Economics of Finance (Volume 1a: Corporate Finance), ed. George Constantinides, Milton Harris and Rene Stulz, Elsevier, 307–336.
- White, Lawrence, 2000, Technological Change, Financial Innovation, and Financial Regulation in the US: The Challenges for Public Policy, Performance of Financial Institutions: Efficiency: Innovation, Regulation. Patrick Harker and Stavros Zenios, eds, 388–415.
- Zapatero, Fernando, 1998, Effects of Financial Innovations on Market Volatility When Beliefs Are Heterogeneous, *Journal of Economic Dynamics and Control*, 597–626.