Risk-Taking, Capital Allocation and Monetary Policy*

Joel M. David† David Zeke‡

FRB Chicago USC

February 22, 2022

Abstract

We study the implications of firm heterogeneity for business cycle dynamics and monetary policy. Firms differ in their exposure to aggregate risk, which leads to dispersion in costs of capital that influence micro-level resource allocations. The heterogeneous firm economy can be recast as a representative firm New Keynesian model, but where total factor productivity (TFP) endogenously depends on the micro-allocation. The monetary policy regime determines the nature of aggregate risk and hence shapes the allocation and long-run level/dynamics of TFP. Welfare losses from policies ignoring heterogeneity can be substantial, which stem largely from a less productive allocation of resources.

JEL Classifications: D24, E12, E22, E44, E52, E63

Keywords: monetary policy, heterogeneous firms, aggregate TFP, reallocation, risk premia

*We thank Jonas Fisher, Gadi Barlevy, Francois Gourio, Florin Bilbiie, Rohan Kekre (discussant), Moritz Lenel, Tom Winberry (discussant), Nicolas Caramp (discussant), Elisa Rubbo, Pablo Kurlat, Andy Atkeson, Narayana Kocherlakota, Mark Bils, Tarek Hassan, Stephen Terry and various seminar and conference participants for helpful comments and suggestions. The views expressed here are those of the authors and not necessarily those of the Federal Reserve Bank of Chicago or the Federal Reserve System.

†joel.david@chi.frb.org.

‡zeke@usc.edu.

1 Introduction

Firm-level micro-data reveal large differences in cyclicality and hence, exposure to aggregate risk. These patterns can have direct consequences for the micro-level allocation of capital: firms that are more sensitive to aggregate shocks are riskier and face a higher cost of capital. All else equal, these firms choose lower levels of capital and have higher marginal revenue products of capital (MRPK).[1] Aggregate total factor productivity (TFP) depends on the cross-sectional distribution of capital and is thus a function of the micro-allocation. In this paper, we develop these insights and explore their implications for (i) equilibrium business cycle dynamics and (ii) the effects and optimal conduct of monetary stabilization policy.

Our point of departure is a workhorse New Keynesian business cycle model featuring shocks to aggregate technology. We augment this framework with two key elements: (i) a cross-section of heterogeneous firms that differ ex-ante in their cyclicality, i.e., exposure to these shocks, and (ii) cyclical distortions in firm-level investment decisions that lead to inefficiencies in the allocation of capital. We adopt a flexible specification for these distortions designed to capture various types of capital market frictions, which show up as 'wedges' to the stochastic discount factor (SDF) in firm-level investment Euler equations.[2] Although for most our analysis we remain agnostic as to the precise source of these frictions, we show that such a wedge arises naturally in salient recent models of financial frictions in quantitative business cycle environments.[3] Throughout the paper, we assume that the aggregate capital stock is an exogenous fixed endowment, which is a standard approach in New Keynesian models and allows us to hone in on the new allocational considerations in our framework. The heterogenous firm economy can be recast in a representative firm formulation, but where aggregate TFP, usually taken as an exogenous driving force, becomes in part endogenous and determined by the resource allocation. Under our assumptions, we can derive a tractable log-linear representation of the equilibrium system, enabling us to formally prove our main results.

At the heart of our model is the risk-return tradeoff governing firm-level capital investment and hence the allocation of capital across firms. Firms equate the expected return to capital, i.e., the MRPK, to a firm-specific cost of capital, which depends on its exposure to aggregate risk and the nature of that risk. The latter is a function of movements in the discount factor (and other state prices), which depend on the properties of capital market distortions and the

[1] We illustrate each of these facts in Figure 1 in Section 4.1 below.

[2] Our approach mirrors the large literature on the misallocation of resources across firms, which has persuasively documented the presence of micro-level mis-allocative distortions and models them in a similar vein, e.g., Hsieh and Klenow (2009) and Restuccia and Rogerson (2008), and in a business cycle context, Hall (2011).

[3] For example, we show that models of frictional financial intermediation as in Gertler and Karadi (2011) or limited asset market participation as in Debortoli and Galí (2018) both lead to wedges in the relevant SDF used to price assets and hence show up in investment Euler equations exactly as in our framework.

path of aggregate output/consumption. In turn, output can be decomposed into (i) the natural rate, or potential output, that would hold in a flexible price economy and depends on TFP alone and (ii) the output gap stemming from nominal rigidities. Thus, the two inefficiencies – fluctuations in the output gap and the capital wedge – distort the risk-return tradeoff and the micro-level allocation. In the aggregate, TFP depends on the micro-allocation. We derive a sharp expression expressing TFP as a function of two micro-level moments: the dispersion in MRPK across firms and the covariance of firm-level capital with cyclical exposures. The first pins down the long-run (average) level of TFP and the second its cyclical volatility. The micro-allocation and aggregate TFP are both endogenous objects jointly determined in equilibrium.

The risk-return relationship at the micro-level implies a tradeoff between the long-run level and volatility of TFP. In the face of aggregate risk, agents shift resources from risky, procyclical firms to safer, less cyclical ones, reducing the covariance between firm-level capital and risk exposures and hence the economy's responsiveness to exogenous shocks. On the other hand, shifting capital in response to aggregate risk introduces a wedge between firm-level capital investment and productivity, which leads to dispersion in MRPK and reduces the level of achieved TFP. The result can be understood as a form of self-insurance – although there are no aggregate savings in the economy, insurance against business cycle risk is attained by shifting capital to less cyclical firms, which endogenously reduces aggregate cyclicality. The cost of this insurance is the foregone output caused by lower TFP due to the mis-alignment of firm-level capital and productivity relative to the allocation that maximizes aggregate production. Note that private insurance contracts do not undo these dynamics: the mechanism is consistent with complete markets and relies on *systematic* risk, which cannot be diversified away.

The nature of aggregate risk and hence the amount of risk-taking depend on the dynamics of TFP, capital distortions and the output gap. The latter is determined by the conduct of monetary policy. Thus, monetary policy affects the macroeconomic risk that agents face and plays a key role in shaping the allocation and behavior of TFP. Aggressive countercyclical policy (e.g., output gap) reduces the volatility of output/consumption and the degree of aggregate risk. The private sector responds by taking on more risk, i.e., shifting capital to more procyclical firms, which improves the productivity of the allocation by reducing MRPK dispersion but increases the volatility of TFP. Less countercyclical policy has the opposite effects.[4] A striking implication is that monetary policy is not neutral in the long-run – in particular, the degree of monetary stabilization influences the distribution of capital across firms and hence shifts the economy closer to/further from its long-run production possibilities frontier, captured by the level of TFP. A second implication is that firm heterogeneity and reallocation opportunities

[4]The result extends the classic notion of the 'Fed put' in financial markets, e.g., Cieslak and Vissing-Jorgensen (2021), Poole (2008) and Miller et al. (2002), to real resource allocations and macroeconomic outcomes.

dampen the effects of stabilization policies. Consider again the case of countercyclical policy. While the direct impact of such a policy is to reduce the cyclicality of the output gap and hence output/consumption, agents endogenously take on more risk by shifting capital to more procyclical firms, increasing the cyclicality of TFP and partially offsetting the direct effect. These allocational effects of monetary policy rely only on the presence of nominal rigidities and firm heterogeneity, irrespective of whether there are other (e.g., capital market) distortions at play in the economy or not.

Capital market imperfections distort the allocation and so the behavior of TFP. The cyclical nature of these distortions calls for a cyclical policy in response. We derive a representation of the welfare criterion that can be decomposed into four terms – first, volatility in inflation and the output gap, as in the standard representative firm setup. Second, the level and volatility of TFP enter directly. Intuitively, these latter two forces determine the dynamics of the natural rate of output. In the representative firm environment, monetary policy affects the output gap but the natural rate is beyond its influence. Here, in contrast, monetary policy affects both the output gap and the natural rate via the dynamics of TFP and so both terms are relevant for understanding the welfare implications of policy. Optimal policy is chosen to best balance these four objectives.

In an undistorted economy – i.e., without nominal rigidities or capital market frictions – the capital allocation and behavior of TFP are efficient. Thus, the economy attains the socially optimal tradeoff between volatility and the long-run level of production. The result has an important implication: when nominal rigidities are the only inefficiency, a version of the 'divine coincidence' holds – the monetary authority can attain the first best by completely stabilizing inflation and restoring the flexible price outcome. In contrast, when additional forces are present that distort the allocation, e.g., financial frictions, complete price stabilization is no longer optimal and the divine coincidence fails. For example, consider the case when these distortions are countercyclical, which leads to an allocation that is overly conservative. Optimal policy entails a countercyclical output gap that raises output in downturns/reduces it in expansions. The usual arguments for stabilization are present, but there is a rationale for even more aggressive countercyclical policy – such a policy helps alleviate the distortions to the allocation by incentivizing risk-taking and more closely aligning firm-level capital and productivity. The resulting redistribution of resources leads to a higher long-run level of TFP and output. The larger the heterogeneity across firms, the more opportunities for reallocation and the more costly are allocative distortions, further strengthening the motives for countercyclical policy.

The effects of heterogeneity and risk do not hinge on monetary policy as the tool of stabilization or the presence of capital distortions per se. For example, similar results hold when fiscal policy is the tool of stabilization (e.g., through cyclical labor income taxes) and with labor

market rather than capital market distortions. Further, we show that heterogeneity affects the standard tradeoff between output gap and inflation volatility in response to cost-push shocks, even when nominal rigidities are the only distortion in the economy. Specifically, optimal policy is more accommodative of these shocks, since acting to neutralize them is more costly – doing so not only generates inefficient output gap fluctuations in the usual way, but also distorts the resource allocation and dynamics of TFP.

In the second part of the paper, we provide a numerical evaluation of our findings. The key new parameters govern the extent of firm heterogeneity and the properties of the capital distortion. We calibrate these parameters to match two important aspects of the micro-data, namely, the observed dispersion in cyclicality across firms and the slope of the relationship between firm-level MRPK and cyclicality. This latter moment is a direct measure of the risk-return tradeoff in firm-level capital choices. The data point to a strong positive (and statistically significant) relationship that is steeper than that implied by preferences and observed aggregate dynamics alone. This gap leads us to a significantly countercyclical estimate of the wedge. This finding is further supported by direct quantification of two micro-founded models of financial frictions using different calibration approaches and data (i.e., household financial market participation and financial sector leverage), which also imply a quantitatively significant countercyclical wedge.

Our results show, first, that heterogeneity and risk can have significant effects on TFP dynamics and welfare. For example, when monetary policy follows a standard Taylor rule, the long-run level of TFP is lower by almost 1.4% and its unconditional volatility by almost 30% (relative to the case with no heterogeneity and/or risk adjustments in the allocation). In contrast, in the first-best, these values are only 0.003% and 9%, respectively. These findings imply that the equilibrium allocation is inefficiently conservative – there is excess shifting of capital to less cyclical firms, which reduces TFP volatility but increases marginal product dispersion, with detrimental effects on long-run TFP and output. Strikingly, of the total welfare losses including from fluctuations in inflation and the output gap, the lion's share – roughly 85% – stems from the reduction in long-run TFP due to the distorted allocation. Thus, abstracting from allocational considerations may miss important welfare effects of business cycle fluctuations. In other words, the long-run allocational effects of stabilization policies seem to be first-order.

Second, we find an important role for policy to improve on equilibrium outcomes. For example, relative to a Taylor rule, optimal policy increases welfare by about 0.65%. The gain in long-run TFP is about 0.44%, which accounts for about two-thirds of the total welfare improvement. In contrast, if the central bank were to set policy to the optimal one ignoring heterogeneity – which, in our simple environment entails complete stabilization of inflation and the output gap – the long-run TFP loss is close to (indeed, slightly larger than) the equilibrium under the Taylor rule. The total welfare gain from such a policy is about 0.4% – thus, accounting

for heterogenity and allocational effects adds a significant contribution to the potential gains from policy, about 0.25% of lifetime steady state consumption.

Related literature. Our paper relates, first, to a burgeoning literature exploring the implications of micro-level heterogeneity for business cycle dynamics and the transmission mechanism of monetary policy, important examples of which include Bilbiie (2008), Kaplan et al. (2018) and Auclert (2019) among others, and for the implementation of optimal policy, e.g., Bilbiie (2018), Challe (2020), Acharya, Challe, and Dogra (2020), Bilbiie and Ragot (2020) and Bhandari et al. (2021).[5] Recent papers studying the transmission of monetary policy through risk premia and asset revaluations include Kekre and Lenel (2021) and Caramp and Silva (2021). The focus of this work has in large part been on the role of household heterogeneity, whereas our paper focuses on the reallocative effects of policy on the production side of the economy. Our results echo a number of important lessons from this literature as forcefully summarized in Kaplan (2021), namely: (i) reallocation opportunities can reduce the effectiveness of policy; (ii) the distributional consequences of policy can be just as – if not more – important than the direct effects; and (iii) under reasonable calibrations, optimal policy entails a redistribution towards agents (in our case, firms) that are more exposed to the business cycle.

Our theory linking monetary policy to capital (mis)allocation and TFP relates to Baqaee et al. (2021) and González et al. (2022), who draw a similar connection in heterogeneous firm models with markup dispersion and financial frictions, respectively, and Kurtzman and Zeke (2020), who highlight the mis-allocative effects of central bank large-scale asset purchases of corporate securities. Relatedly, David et al. (2021) link marginal product dispersion to aggregate risk exposures in a partial equilibrium setting and calculate the implications for measures of misallocation and TFP. More broadly, our paper connects to recent work studying monetary policy in distorted economies with production-side heterogeneity, recent examples of which include Ottonello and Winberry (2020), who study the investment channel of policy with heterogeneous firms and financial frictions, and Angeletos and La'O (2020) and La'O and Tahbaz-Salehi (2020) and Rubbo (2020), who study optimal policy in environments with dispersed private information held by firms and input-output linkages, respectively. We contribute to this line of work by exploring the implications of a different dimension of heterogeneity, namely, the risk-return tradeoff highlighted in the asset pricing literature, which is a robust feature of the micro-data. Our theoretical framework incorporates heterogeneous risk premia into a workhorse business cycle environment. Our findings point to an important role for this form of heterogeneity in determining the equilibrium effects and optimal conduct of monetary policy via its impact on

[5] A large body of work studies optimal monetary policy in a rich variety of representative agent business cycle settings. For a textbook treatment see Galí (2015) and for a recent review of the literature see Woodford (2010). Barlevy (2005) surveys the literature on the benefits of stabilization.

the resource allocation and behavior of aggregate TFP.

A large literature has studied the link between volatility and the level and/or growth rate of economic activity.[6] We build on this body of work by studying the role of micro-level allocations in influencing both aggregate volatility and the level of aggregate activity. Our framework shows that when volatility and long-run outcomes are jointly determined in equilibrium, understanding the underlying primitives is crucial – for example, insufficient policy stabilization leads to more volatile output and lower TFP, whereas cyclical distortions dampen both. Relatedly, a number of recent papers have linked monetary policy and TFP growth in models with endogenous innovation.[7] In contrast, we tease out this relationship in an environment where volatility and long-run outcomes are jointly determined by (re)distributions across heterogeneous agents and explicitly assess the role of monetary stabilization policies in shaping the resulting equilibrium.

2 The Model

Preferences and technology. A continuum of households indexed by $j \in [0,1]$ seek to maximize expected lifetime utility from consumption and leisure, given by

$$\mathbb{U} = (1-\rho)\,\mathbb{E}_{-1}\left[\sum_{t=0}^{\infty}\rho^t\left(\frac{C_{jt}^{1-\gamma}}{1-\gamma}-\chi\frac{L_{jt}^{1+\varphi}}{1+\varphi}\right)\right], \tag{1}$$

where ρ denotes the time discount factor and all other notation is standard. Because we are studying the effects of risk, we assume throughout that households are sufficiently risk averse, specifically, $\gamma > 1$.

We assume nominal rigidities in the form of sticky wages. The setup is standard and we provide only a brief overview. Households monopolistically supply differentiated labor services, which are then bundled into the final labor input using a CES aggregator with elasticity of substitution ν_w. Wage changes are subject to quadratic adjustment costs $\frac{\theta_w}{2}\left(\Pi_t^w-1\right)^2 Y_t$, where Π_t^w denotes gross nominal wage inflation. Following Auclert et al. (2018), we assume these costs enter as an extra additive disutility term in (1).[8] Households receive labor income, income from capital that they rent to firms, the return on holdings of nominally risk-free bonds and any distributed profits from firms.

[6] Examples include Kormendi and Meguire (1985), Ramey and Ramey (1991, 1995), Obstfeld (1994), Acemoglu and Zilibotti (1997), Barlevy (2004), Jones et al. (2005), Koren and Tenreyro (2007), Aghion et al. (2010) and Koren and Tenreyro (2013).

[7] For example, Moran and Queralto (2018), Garga and Singh (2021) and Jordà et al. (2020).

[8] Although not necessary for any of our results, this assumption ensures that aggregate output exactly equals total consumption (else the equality only holds up to a first-order approximation). The framework also accommodates Calvo pricing frictions with a modified slope of the Phillips Curve.

The household optimality conditions lead to a standard New Keynesian wage Phillips curve (discussed in more detail below) and to aggregate labor supply that satisfies

$$\mathcal{M}_t W_t = \chi L_t^{\varphi} C_t^{\gamma} \ , \tag{2}$$

where W_t is the real wage (relative to the price of the final consumption good), L_t and C_t aggregate labor supply and consumption, respectively, and $\mathcal{M}_t > 0$ denotes a labor 'wedge' or (inverse) markup between the real wage and the marginal rate of substitution between consumption and labor that stems from the household wage-setting problem. The household's intertemporal marginal rate of substitution, or stochastic discount factor (SDF), is given by

$$\Lambda_t = \rho \left(\frac{C_t}{C_{t-1}} \right)^{-\gamma} \ .$$

To ease notation later, define the log SDF as $\lambda_t \equiv \log \Lambda_t$.

The final consumption good is produced by a competitive representative firm, which bundles a continuum of intermediate goods, indexed by $i \in [0, 1]$, using a constant elasticity of substitution (CES) aggregator:

$$Y_t = \left(\int Y_{it}^{\nu} di \right)^{\frac{1}{\nu}} \ , \tag{3}$$

where $\nu \in (0, 1]$ and $\frac{1}{1-\nu}$ is the elasticity of substitution between intermediate goods.

Intermediate goods are produced using capital and labor according to

$$Y_{it} = A_{it} K_{it}^{\alpha_1} L_{it}^{\alpha_2}, \qquad \alpha_1 + \alpha_2 \leq 1 \ .$$

Importantly, intermediate good firms are heterogeneous in productivity, where A_{it} denotes the productivity of firm i in period t.

Throughout the paper, we abstract from accumulation considerations and assume the total capital stock is an exogenous and fixed endowment, i.e., $K_t = \overline{K} \ \forall \ t$.[9] This is a common assumption in the New Keynesian literature and allows us to hone in on the new allocational effects in our framework. With this assumption, the economy here exactly nests the standard New Keynesian model without capital.

The resource constraints in the economy are then given by

$$C_t = Y_t, \qquad \int K_{it} di = K_t = \overline{K}, \qquad \int L_{it} di = L_t \ .$$

[9]The value of $\overline{K}$ plays no role in the analysis.

Demand and revenue. Profit maximization by the final good producer yields a demand function for intermediate good i:

$$P_{it} = \left(\frac{Y_{it}}{Y_t}\right)^{\nu-1} ,$$

where P_{it} denotes the relative price of good i in terms of the final good. Revenues for intermediate firm i at time t are

$$P_{it}Y_{it} = Y_t^{1-\nu}Y_{it}^{\nu} = Y_t^{1-\nu}A_{it}^{\nu}K_{it}^{\alpha_1\nu}L_{it}^{\alpha_2\nu} .$$

Input choices. Intermediate firms hire labor period-by-period to maximize current profits. The optimal choice of labor satisfies

$$\alpha_2\nu\frac{Y_t^{1-\nu}Y_{it}^{\nu}}{L_{it}} = W_t , \tag{4}$$

which shows that firms equalize the marginal revenue product of labor. Operating profits (revenues less labor expenses) are proportional to revenues and are equal to

$$\Pi_{it} = P_{it}Y_{it} - W_tL_{it} = (1-\alpha_2\nu)\,P_{it}Y_{it} = GY_t^{\frac{1-\nu}{1-\alpha_2\nu}}A_{it}^{\frac{\nu}{1-\alpha_2\nu}}W_t^{-\frac{\alpha_2\nu}{1-\alpha_2\nu}}K_{it}^{\alpha} , \tag{5}$$

where $\alpha \equiv \frac{\alpha_1\nu}{1-\alpha_2\nu}$ is the effective curvature of operating profits with respect to capital and $G \equiv (1-\alpha_2\nu)\,(\alpha_2\nu)^{\frac{\alpha_2\nu}{1-\alpha_2\nu}}$.

At the end of period $t-1$, firms rent capital for use in period t at rate R_{t-1}^K. The firm chooses capital to maximize expected discounted profits, i.e., to solve

$$\max_{K_{it}} \ \mathbb{E}_{t-1}\left[\Lambda_tT_{\Lambda t}\Pi_{it}\right] - R_{t-1}^KK_{it} .$$

Firms discount the payoffs from capital using the household's discount factor, Λ_t. The term $T_{\Lambda t}$ denotes a capital or financial 'wedge' that distorts firms' investment choices and hence the allocation of capital. The wedge is meant to capture a broad set of cyclical distortions in capital markets that influence firms' investment decisions. Rather than take a stand on the exact source of these distortions, we adopt this flexible – albeit reduced-form – specification designed to encompass various types of capital market frictions. We show below that such a wedge arises naturally across a range of detailed models of financial or capital market frictions.[10]

[10] Hall (2011) takes a similar approach to modeling financial frictions in a business cycle context and contains a detailed discussion of the mapping of such a financial wedge to theories of financial frictions.

The optimal choice of capital satisfies

$$\mathbb{E}_{t-1}\left[\Lambda_t T_{\Lambda t} MRPK_{it}\right] = R^K_{t-1}, \qquad MRPK_{it} \equiv \alpha \frac{\Pi_{it}}{K_{it}}\,, \tag{6}$$

which shows that firms equalize the expected discounted marginal revenue product of capital.

Capital market frictions – examples. Our formulation of the wedge in firms' capital decisions is designed to capture many potential inefficiencies in the functioning of capital markets. Here, we give two concrete examples based on salient recent models of financial frictions in quantitative business cycle environments. We show that these more detailed models lead to exactly such a wedge, which in these cases manifests itself as a distortion to the stochastic discount factor. We give a brief description of each setup and provide further details in Appendix A. In Section 4.3 we quantitatively evaluate the contribution of these channels to the capital wedge we infer from the firm-level micro-data.

Example 1: frictional financial intermediation – Consider a model of frictional financial intermediation along the lines of Gertler and Karadi (2011) (GK). Firms issue equity liabilities to financial intermediaries in order to finance their capital. Intermediaries borrow from households at the risk-free rate in order to provide this financing. Intermediary assets include the total market value of firm liabilities. Intermediary liabilities consist of deposits and equity, or net worth. Intermediaries exit at exogenous rate σ and are immediately replaced by new entrants. Intermediaries act to maximize the expected discounted stream of dividends payed out to households. Due to a moral hazard/costly enforcement problem, intermediaries face collateral constraints that limit their ability to obtain deposits. Specifically, a fraction θ of assets can be diverted, which implies an incentive constraint limiting intermediary collateral.

Assuming that the collateral constraint binds, the intermediary's optimality conditions yield an expression analogous to (6), where

$$T_{\Lambda t} = 1 - \sigma + \sigma \frac{\partial V_t}{\partial N_t}\,. \tag{7}$$

Here, V_t and N_t denote the market value and net worth of the intermediary, respectively. The expression shows that models of frictional financial intermediation lead to a capital wedge – specifically, a distortion to the relevant SDF pricing assets – that reflects the shadow marginal value of net worth to the intermediary, $\frac{\partial V_t}{\partial N_t}$.

Example 2: limited asset market participation – Next, consider a model of limited asset market participation on the part of households as in the two agent New Keynesian (TANK) model of Debortoli and Galí (2018). A constant fraction θ of households are financially constrained,

i.e., they do not participate in financial markets and simply consume their labor income in each period. The relevant SDF for pricing assets is then $\Lambda_t^U = \rho \left(\frac{C_t^U}{C_{t-1}^U} \right)^{-\gamma}$, where C_t^U denotes the consumption of an unconstrained household. Firms face (common) price markup shocks denoted T_t^p, which leads to variation in the share of output paid to labor vs. capital and profits.[11] Using a first-order approximation to the resource constraint, unconstrained consumption (in logs, henceforth denoted with lowercase) is equal to

$$c_t^U = c_t + \frac{\alpha_2}{1-\alpha_2}\theta\tau_t^p, \qquad \tau_t^p \equiv \log T_t^p \ ,$$

where c_t is (log) total consumption. Using the resulting SDF to price assets yields (6), where

$$\log T_{\Lambda t} = -\frac{\alpha_2}{1-\alpha_2}\gamma\theta\Delta\tau_t^p, \qquad \Delta\tau_t^p \equiv \tau_t^p - \tau_{t-1}^p \ . \tag{8}$$

Thus, a simple TANK model featuring limited asset market participation and price markup shocks leads to a capital wedge, which shows up as a distortion to the SDF that reflects the proportion of constrained households and time-variation in their share of aggregate income.[12]

In this simple setup, the capital wedge is purely exogenous. However, this may not be the case in more complicated versions (e.g., with sticky output prices in addition to sticky wages) and for other possible frictions (e.g., the GK model), where the wedge may be endogenous to policy. For simplicity, and given the wide range of distortions likely at play in the data, we abstract from these issues here and consider the problem of a monetary authority which takes the wedge as given.

Additional examples – Although financial frictions are a natural candidate (and, as just shown, connect closely to recent work studying financial frictions in New Keynesian environments), the capital wedge can capture a wide range of potential distortions in the capital choice. For instance, it clearly picks up any cyclical tax on operating profits. Appendix A provides two additional examples based on theories of externalities in preferences and expectational biases.

Aggregation and equilibrium. The following result shows that from a macroeconomic perspective, the equilibrium of the heterogeneous firm economy is observationally equivalent to a representative firm economy with endogenous TFP (proof in Appendix B.1):

Proposition 1. *The aggregate variables of the heterogenous firm economy behave identically to a representative firm economy with TFP Ψ_t, where (i) the log of aggregate output obeys the*

[11] Although we micro-found these shocks as stemming from markups, this interpretation is not strictly necessary, only that there is time-variation in the distribution of output between labor and capital/profits.

[12] The result is reminiscent of Longstaff and Piazzesi (2004) who show that fluctuations in the corporate earnings/consumption ratio can explain a significant portion of the equity premium.

following stochastic process:

$$y_t = \phi_\psi \psi_t + \mu_t, \qquad \textit{where} \qquad \psi_t \equiv \log \Psi_t, \quad \mu_t \equiv \phi_l \log \mathcal{M}_t \tag{9}$$

and (ii) aggregate TFP is defined by:

$$Y_t = \Psi_t K_t^{\alpha_1} L_t^{\alpha_2}, \quad \Psi_t = \left(\int A_{it}^{\frac{\nu}{1-\alpha_2 \nu}} \left(\frac{K_{it}}{K_t} \right)^{\alpha} di \right)^{\frac{1-\alpha_2 \nu}{\nu}}, \quad \frac{K_{it}}{K_t} = \frac{\mathbb{E}_{t-1} \left[\Lambda_t T_{\Lambda t} \Pi_{it} \right]}{\int \mathbb{E}_{t-1} \left[\Lambda_t T_{\Lambda t} \Pi_{it} \right] di}, \tag{10}$$

where $\phi_\psi > 0$ and $\phi_l > 0$ are constant composites of production and preference parameters.

Expression (9) decomposes fluctuations in output into (i) the 'natural rate' or potential output, $y_t^n \equiv \phi_\psi \psi_t$, which is a function of TFP only and (ii) the output gap, denoted μ_t, which is a function of the markup coming from the nominal wage rigidities. Because it is a pre-determined state variable when shocks are realized, our definition of the output gap holds the capital allocation fixed across actual and potential output. An implication of this definition is that the natural rate of output is not independent of monetary policy since, as we describe below, the conduct of policy affects the allocation and hence TFP. Although other notions of the output gap exist in the presence of endogenous state variables, this definition is most appropriate in our context since it explicitly accounts for the current productive capacity of the economy (which depends on the allocation) and will turn out to be exactly the gap that determines the inflation/real activity tradeoff facing the monetary authority.[13]

The left-hand expression in (10) defines aggregate TFP through a reduced-form aggregate production function, which inherits the form of the micro-level production function and is identical to a representative firm economy with TFP Ψ_t. The middle expression shows the key result: though the economy can be written as if there was a representative firm, TFP of that firm is endogenous and depends on the efficiency of the capital allocation across firms. In particular, TFP is equal to an average of firm-level productivities, weighted by their shares of the aggregate capital stock. The right-hand expression shows that these shares are determined by individual firm expected discounted profits relative to average expected discounted profits.

The remaining equilibrium conditions are the Phillips curve and consumption Euler equation (IS curve). Because the household side of the model is standard, these take the usual form and we provide the expressions in Appendix B.1.

[13] Our definition follows that advocated by Woodford (2003) in the presence of pre-determined state variables (Chapter 5, Section 3.4) and is equivalent to one where the natural rate of output is defined as the level of output that would be obtain if prices in period t onward were flexible, given the actual allocation of capital entering t. An alternative would be to define the natural rate as the level of output that would hold if prices had always been flexible, which would place deviations in TFP from the efficient level into the output gap, but would imply a disconnect between the natural rate and the actual current productive capacity of the economy.

Proposition 1 reveals the first main insight of our paper: macroeconomic dynamics depend on the micro-level allocation of capital via the behavior of TFP, Ψ_t. In reverse, the allocation itself depends on aggregate dynamics – the SDF (which is a function of aggregate consumption/output) directly enters firms' capital decisions, as do other aggregate variables (i.e., the wage and aggregate output) through movements in expected profitability, Π_{it}. Thus, the allocation and dynamics of aggregate TFP are both endogenous objects that are jointly determined in equilibrium. We show below that under our assumed processes for firm and aggregate shocks, aggregation in our model is exact and we can obtain an exact analytical expression for TFP.

Stochastic processes. Firm productivity (in logs) is given by

$$a_{it} = \hat{\beta}_i a_t + \mathcal{O}_{it}, \quad a_t = \delta a_{t-1} + \varepsilon_t \quad \text{where} \quad \varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right), \quad \hat{\beta}_i \sim \mathcal{N}\left(1, \sigma_{\hat{\beta}}^2\right) . \tag{11}$$

Here, a_t is an aggregate shock to technology that follows an AR(1) process with persistence δ and variance of the innovations σ_ε^2. Crucially, firms are heterogeneous in their sensitivity, or exposure, to movements in a_t and hence their degree of cyclicality, captured by $\hat{\beta}_i$.[14] The mean beta is unity. The cross-sectional variance in beta, $\sigma_{\hat{\beta}}^2$, captures the extent of this heterogeneity.

The term $\mathcal{O}_{it}$ denotes a number of adjustments to offset Jensen's inequality terms when taking expectations both over time and across firms and maintain log-linearity of the economy. These ensure that when agents are risk-neutral with respect to all state prices and/or firms are homogeneous, TFP equals the exogenous aggregate shock. Since these terms are independent of risk and policy, they play no other role in the analysis and can safely be ignored for purposes of exposition. We detail the adjustment terms in our derivations in Appendix B.1.

The capital wedge is a constant elasticity function of the aggregate shock, i.e.,[15]

$$\tau_{\Lambda t} \equiv \log T_{\Lambda t} = -\tau_{\Lambda a} a_t \, . \tag{12}$$

We refer to the wedge as countercyclical when $\tau_{\Lambda a} > 0$, in which case it acts like a countercyclical tax on firm profits. For intuition, consider the case where the wedge distorts the SDF (as in the two examples laid out above). A countercyclical wedge implies a discount factor that is inefficiently countercyclical, i.e., it falls (rises) excessively in expansions (downturns) relative to what preferences and the dynamics of consumption would dictate. In this case agents act

[14] For simplicity, we abstract from firm-level idiosyncratic shocks. Although important for matching micro-level investment moments, with complete markets, agents can perfectly diversify these shocks, implying that they bear no risk premium. Additionally, they are independent of policy. Thus, they would play no role in the analysis (other than adding constant terms to a number of the equilibrium equations).

[15] We assume optimal time-invariant subsidies are in place to offset goods and labor market monopoly distortions. This ensures an efficient steady-state and that only the time-varying distortions play a role.

excessively averse to bearing aggregate risk. The opposite holds when the wedge is procyclical.

For simplicity, the remainder of the analytical analysis considers the special case when the aggregate shock is i.i.d., i.e., $\delta = 0$. Under this special case, we obtain particularly sharp expressions for our main theoretical results. We relax this assumption in our quantitative exercises in Section 4, but show that the same intuition from this simpler case carries through.

The conduct of monetary policy. We assume the central bank conducts policy to directly control the output gap. In particular, the monetary authority chooses a value μ_a that determines the cyclicality of the output gap such that[16]

$$\mu_t = \mu_a a_t \ . \tag{13}$$

More aggressive countercyclical policy entails a more negative μ_a, i.e., a more countercyclical output gap that falls in expansions and rises in downturns. In Appendix B.2 we show that common interest rate rules imply an output gap of exactly this form, e.g., choosing the cyclicality of the nominal interest rate (the response of the interest rate to the aggregate shock) or the reaction coefficients in standard Taylor rules. More aggressive countercyclical policy entails a more procyclical nominal interest rate, i.e., a nominal rate that rises (falls) more in response to positive (negative) realizations of the aggregate shock. Given the behavior of the output gap in (13), the path of inflation is determined by the Phillips curve.

Log-linear solution. Under an output gap of the form in (13), the model allows for exact aggregation and the aggregate quantity variables (e.g., TFP and output) have an *exact* solution that is log-linear in the shock, a_t. The nominal variables are determined by the non-linear Phillips curve and Euler equation. However, conditional on the behavior of the output gap, these expressions do not affect any of the real variables, but only the behavior of inflation and the nominal interest rate. We exploit this feature to derive exact expressions for all quantity variables. We then perform a standard log-linearization of the Phillips curve and Euler equation.[17] This approach yields analytic tractability of the equilibrium system and ensures that our framework and solution fully nest the standard representative firm New Keynesian model.

Proposition 2. *Under (i) the assumed stochastic processes on the exogenous shocks and capital*

[16] Our results do not depend on this precise specification – for example, Section 4 shows that optimal policy under full commitment and more complicated rules lead to similar qualitative effects of policy.

[17] As is standard in this class of model, a first-order approximation to the the Phillips curve is sufficient to obtain a second-order approximation to the utility function for our welfare analysis in Section 3.

wedge and (ii) the conduct of monetary policy, (log) aggregate TFP is equal to

$$\psi_t = \overline{\psi} + \psi_a a_t \ , \tag{14}$$

where $\overline{\psi}$ *and* ψ_a *are endogenous constants that depend on the capital allocation. The log-linear equilibrium is characterized by (9), (14) and the log-linearized Phillips curve and Euler equation:*

$$\pi_t^w = \rho \mathbb{E}_t \left[\pi_{t+1}^w\right] + \lambda_w \mu_t \tag{15}$$

$$y_t = \mathbb{E}_t \left[y_{t+1}\right] - \frac{1}{\gamma}\left(i_t - \mathbb{E}_t \left[\pi_{t+1}^p\right]\right) \ . \tag{16}$$

The Phillips curve relates wage inflation, π_t^w, to expected wage inflation and the output gap, μ_t. The slope is determined by the composite parameter $\lambda_w \equiv \frac{\alpha_2 \nu_w}{\theta_w \phi_l}$. The Euler equation relates expected output/consumption growth to the nominal interest rate, i_t, and expected price inflation, π_{t+1}^p. The proposition shows that the log-linear equilibrium is completely characterized by the path of output, the definition of TFP, the Phillips curve and the Euler equation. Notice that this system is identical to the textbook representative firm New Keynesian model (e.g., Galí (2015)), the only difference being the endogeneity of TFP. As discussed above, conditional on the assumed form of monetary policy, expressions (9) and (14) are exact. Expressions (15) and (16) are approximated and pin down the path of the nominal variables (the nominal interest rate and inflation). For much of our analysis, we can focus on the former set of exact equations characterizing the real variables. The latter set come into play when characterizing the behavior of the nominal rate that supports a given policy and the path of inflation for our welfare results.

Proposition 2 introduces two endogenous constants – $\overline{\psi}$ and ψ_a – that determine (i) the long-run average level of TFP and (ii) the loading, or responsiveness, of TFP to exogenous shocks. In particular, the long-run level of TFP is equal to $\overline{\psi}$ and its standard deviation to $\sigma\left(\psi_t\right) = \psi_a \sigma_\varepsilon$. We provide exact solutions for these constants in expressions (62) and (63) in Appendix B.1 and show that they both depend on the micro-level capital allocation. Note that they are not in general equal to zero and one, respectively, implying that endogenous TFP is not equal to exogenous technology. Characterizing TFP entails characterizing these two terms. The next subsection lays out in detail how these terms are determined and in particular, how they depend on aggregate risk and the micro-allocation.

2.1 Micro Allocations and Aggregate TFP

As described above, Proposition 2 presents an exact solution for aggregate TFP with explicit characterizations for $\overline{\psi}$ and ψ_a in Appendix B.1. It turns out, however, that we can derive particularly sharp expressions using a series of first and second-order approximations, and

hence for this section only, we use such an approximation simply to provide as transparent intuition as possible. We provide detailed derivations of this approach in Appendix B.3, which ensures that we retain the effects of aggregate risk while at the same maintaining analytical clarity by ensuring linearity of firm-level variables in beta.[18] We return to the exact solution for our welfare results and quantitative work in Sections 3 and 4.

First, we can approximate aggregate TFP in (10) to obtain

$$\psi_t \equiv \log \Psi_t = \underbrace{-\frac{1}{2}\omega \text{var}\left(mrpk_i\right)}_{\overline{\psi}} + \underbrace{\left(1 + \alpha_1 \text{cov}\left(k_i, \beta_i\right)\right)}_{\psi_a} a_t \ , \tag{17}$$

where $\omega \equiv \frac{\alpha_1(1-\alpha_2\nu)}{1-(\alpha_1+\alpha_2)\nu}$ and $\beta_i \equiv \frac{\nu}{1-\alpha_2\nu}\hat{\beta}_i$. The expression shows that the endogenous terms determining TFP depend on two moments of the micro-allocation evaluated at the ergodic mean (denoted without time subscripts), namely (i) the cross-sectional variance of $mrpk_i$ and (ii) the covariance of k_i with β_i. The level term, $\overline{\psi}$, reflects a familiar result from the misallocation literature, e.g., Hsieh and Klenow (2009) and David and Venkateswaran (2019): in log-linear environments, the variance of $mrpk$ is a sufficient statistic for the TFP losses from mis-allocated capital. The loading of TFP on the shock, ψ_a, is novel to our setting but also intuitive: from Proposition 1, aggregate TFP is an average of firm-level productivities weighted by their shares of the aggregate capital stock. When highly cyclical firms are larger, TFP is more cyclical. In the approximate expression, the covariance between firm-level capital and cyclicality (beta) is a sufficient statistic for this effect. Clearly, the behavior of TFP – its long-run mean and cyclical volatility – depends crucially on the micro-allocation, specifically, moments (i) and (ii).

Risk-taking and capital allocation. The capital allocation is shaped by the risk-return in firm-level capital choices. To gain intuition, approximate the optimality condition (6) to obtain

$$\mathbb{E}_{t-1}\left[mrpk_{it}\right] + \frac{1}{2}\text{var}_{t-1}\left(mrpk_{it}\right) = -\text{cov}_{t-1}\left(mrpk_{it}, \tilde{\lambda}_t\right) \ , \tag{18}$$

where $\tilde{\lambda}_t = \lambda_t + \tau_{\Lambda t}$ is the log of the (distorted) discount factor. The expression shows that the capital allocation is determined by a standard asset pricing equation relating the expected $mrpk$ (i.e., the return on capital) to a firm-specific risk premium in the cost of capital. The risk premium equals the negative of the conditional covariance of the return with the (log of the) distorted SDF, $\tilde{\lambda}_t$. More procyclical firms have $mrpk$ that is high in good times, i.e., when marginal utility and the SDF are low. Thus, these firms are riskier and must offer a higher

[18] Specifically, we use second-order approximations to evaluate expectations that involve aggregate risk and first-order approximations to linearize any remaining non-linear terms.

expected return as compensation. Less procyclical firms have higher $mrpk$ in downturns when marginal utility is high and thus are safer and offer a lower expected return.

The capital wedge leads to inefficiencies in the pricing of risk and hence distorts the risk-return tradeoff and capital allocation. If $\tau_{\Lambda a} > 0$, i.e., the distortion is countercyclical, it augments movements in the SDF and the risk-return slope is excessively steep – agents act inefficiently averse to bearing risk, which implies the cost of capital is too high for risky firms and too low for safe ones and an allocation that is overly conservative. If the wedge is procyclical, agents are taking on excessive risk and the allocation is too aggressive.

The following proposition characterizes the allocation as a function of a single object – an equilibrium 'risk adjustment,' κ – that jointly determines the risk-return tradeoff and the allocation of capital and hence the dynamics of TFP:

Proposition 3. *(i) The optimal choice of capital satisfies*

$$mrpk_{it} = \underbrace{\log \tilde{A}_{it} - \log \mathbb{E}_{t-1}\left[\tilde{A}_{it}\right]}_{uncertainty} + \underbrace{\beta_i \kappa \sigma_\varepsilon^2}_{risk}, \qquad k_{it} = \frac{1}{1-\alpha}\left(\underbrace{\log \mathbb{E}_{t-1}\left[\tilde{A}_{it}\right]}_{productivity} - \underbrace{\beta_i \kappa \sigma_\varepsilon^2}_{risk}\right). \tag{19}$$

(ii) The endogenous terms in TFP are given by

$$\psi_a = 1 - \omega \kappa \sigma_\varepsilon^2 \sigma_\beta^2, \qquad \overline{\psi} = -\frac{1}{2}\omega \left(\kappa \sigma_\varepsilon^2\right)^2 \sigma_\beta^2, \tag{20}$$

where (iii) κ is an equilibrium risk adjustment equal to

$$\kappa = \underbrace{\kappa_\psi \psi_a}_{TFP} + \underbrace{\tau_{\Lambda a}}_{capital\ wedge} + \underbrace{\kappa_l \mu_a}_{output\ gap} \tag{21}$$

and $\tilde{A}_{it} = A_{it}^{\frac{\nu}{1-\alpha_2 \nu}}$ and κ_ψ and κ_l are positive constants.

The term κ is a risk adjustment in the capital allocation and indeed, is a sufficient statistic that captures the effects of all sources of aggregate risk in the cross-sectional allocation. Specifically, Appendix B.4 shows that κ measures the (negative of the) elasticity of the discounted profitability of capital to aggregate shocks operating through discount factor effects and via equilibrium effects on the wage, W_t, and aggregate demand, Y_t. Put another way, κ captures the aggregate risk facing the firm through the joint movement of all relevant state prices. The price of risk is positive when $\kappa > 0$.

Expression (21) makes clear that aggregate risk stems from movements in (i) TFP (e.g., the natural rate of output), (ii) the capital distortion and (iii) the output gap. In Appendix B.4 we explicitly characterize and provide further intuition for the constants κ_ψ and κ_l. The risk

adjustment κ plays the same role in the exact solution in Appendix B.1 and takes precisely the form in (21), i.e., that expression is exact. Although the remaining expressions are somewhat more complicated, the intuition from the approximate expressions carry through. In particular, κ is a sufficient statistic for the role of aggregate risk in determining the allocation and TFP.

Risk-taking and dispersion – Expression (19) illustrates the two key effects of aggregate risk on micro-level resource allocations. First, the right-hand equation shows how firm-level risk affects the capital choice, which depends on expected productivity and the firm-specific risk premium, which is linear and increasing in the firm's exposure to aggregate risk, β_i: conditional on expected productivity, more procyclical firms are riskier, face a higher cost of capital and hence, choose a lower level of capital. Thus, the risk premium drives a wedge between firm-level capital and expected productivity. The strength of this effect is determined by the risk adjustment, κ, and the degree of aggregate volatility, σ_ε^2. The extent of aggregate risk determines the amount of risk-taking at the micro-level, with more aggregate risk, i.e., larger κ, leading capital to shift towards less cyclical low beta firms, reducing the risk profile of the capital allocation.

Second, there is a dispersion effect – the left-hand equation in (19) shows that $mrpk$ depends on the realization of unexpected shocks ('uncertainty') and the risk premium. Riskier firms must offer a higher rate of return on capital as compensation for bearing that risk. The expression shows that $\kappa\sigma_\varepsilon^2$ is the 'price' of aggregate risk exposure in capital decisions: for each unit increase in β_i, the return on capital must increase by this amount. By introducing a wedge between capital choices and expected productivity, exposure to aggregate risk induces differences in $mrpk$. The effects of aggregate risk on firm-level $mrpk$ and capital are two sides of the same coin: greater exposure to aggregate risk leads to a higher expected $mrpk$; due to diminishing marginal returns to capital in production, higher $mrpk$ entails a smaller capital stock.

The level and volatility of TFP – In turn, expression (20) reveals the two key effects of aggregate risk on macroeconomic dynamics via the micro-allocation. First, there is a smoothing effect: in the absence of heterogeneity and/or risk adjustments in the allocation, the term ψ_a, which captures the elasticity of TFP to the exogenous shock, is equal to one – exogenous technology and endogenous TFP are the same. With heterogeneity and risk, ψ_a is strictly less than one and is decreasing in κ, i.e., $\frac{\partial \psi_a}{\partial \kappa} < 0$. In the face of exogenous shocks, the endogenous reallocation of capital from more to less procyclical firms lowers cov (k_i, β_i) in (17), reducing the responsiveness of TFP to those shocks. The larger the risk adjustment, the less is risk-taking at the micro-level (i.e., the lower/more negative this covariance) and the lower the sensitivity of TFP. Second, there is a level effect: by inducing dispersion in $mrpk$, the risk adjustment lowers the long-run level of TFP, $\overline{\psi}$. In the absence of heterogeneity and/or risk, $\overline{\psi}$ is equal to zero. With these elements, it is negative and decreasing in κ, i.e., $\frac{\partial \overline{\psi}}{\partial \kappa} < 0$ – when the risk adjustment is larger, the

capital allocation depends more on firm-level risk than expected productivity, which increases var $(mrpk_i)$ in (17) and depresses long-run productivity.

Heterogeneity in risk generates a tradeoff between the long-run level and cyclical volatility of TFP. A smaller risk adjustment leads to a more productive allocation of resources, but more volatile TFP. A larger risk adjustment the opposite. The result can also be understood as a form of self-insurance. There are no savings in the economy (the aggregate capital stock is fixed and bonds are in zero net supply). Yet agents can partially self-insure against cyclical fluctuations by shifting capital to less cyclical firms, which endogenously reduces the extent of aggregate risk. The cost of doing so is higher marginal product dispersion, which reduces the productiveness of the resource allocation and hence the long-run level of output/consumption.

The risk adjustment, κ, determines and is in turn determined by both the micro-level allocation and the nature of macroeconomic dynamics. Thus, these are both endogenous objects that are jointly determined in equilibrium, as is κ itself: from expressions (19) and (20), κ determines the allocation and hence ψ_a, the loading of TFP on exogenous shocks; from expression (21), κ is determined by ψ_a. In Appendix B.3 we provide an explicit solution for κ under our approximation in terms of model primitives (and the exact analog in Appendix B.1).

Simple example. We can use an even simpler example to sharply illustrate these links between risk, macroeconomic dynamics and resource allocations. Consider the special case where $\alpha_2 = 0$ so that capital is the only factor of production; $\nu = 1$ so that intermediate goods are perfect substitutes; all prices are flexible so that $\mathcal{M}_t = 1$; and capital markets are efficient, i.e., $T_{\Lambda t} = 1$. Under these assumptions, there are no aggregate movements in factors of production at all, yet the economy features rich dynamics arising from the resource allocation alone.

Proposition 1 implies that the equilibrium is fully determined by the following system:

$$Y_t = \Psi_t K_t^{\alpha_1}, \quad \Psi_t = \int A_{it} \left(\frac{K_{it}}{K_t}\right)^{\alpha_1} di, \quad \frac{K_{it}}{K_t} = \frac{\left(\mathbb{E}_{t-1}\left[\Lambda_t A_{it}\right]\right)^{\frac{1}{1-\alpha_1}}}{\int \left(\mathbb{E}_{t-1}\left[\Lambda_t A_{it}\right]\right)^{\frac{1}{1-\alpha_1}} di}, \tag{22}$$

and using the form of TFP from Proposition 2, the discount factor satisfies

$$\log \Lambda_t - \mathbb{E}_{t-1} \log \Lambda_t = -\gamma y_t = -\gamma \psi_a a_t = -\kappa a_t \,, \tag{23}$$

where $\kappa \equiv \gamma \psi_a$ is the (negative) elasticity of the discount factor, Λ_t, to the aggregate shock, A_t. In this simple environment where goods are perfect substitutes and capital is the only factor of production, the only aggregate state price that affects capital decisions is the SDF.

The capital allocation satisfies a simplified version of (19) with modifications to the relevant curvature, specifically, $\beta_i = \hat{\beta}_i$, $\tilde{A}_{it} = A_{it}$ and $\alpha = \alpha_1$. Similarly, TFP is determined by (17) with

$\omega = \frac{\alpha_1}{1-\alpha_1}$. Evaluated at the ergodic mean, the variance of $mrpk$ is approximately $\text{var}(mrpk_i) = (\kappa\sigma_\varepsilon^2)^2 \sigma_\beta^2$ and the covariance of firm-level capital and beta is $\text{cov}(k_i, \beta_i) = \frac{1}{1-\alpha_1}\kappa\sigma_\varepsilon^2\sigma_\beta^2$.

We can use this simplified example to clearly highlight the equilibrium nature of the risk adjustment, κ. First, from its definition, $\kappa = \gamma\psi_a$, i.e., κ depends on the nature of macro dynamics, specifically, the endogenous response of TFP to exogenous shocks. On the other hand, applying (17) yields $\psi_a = 1 - \frac{\alpha_1}{1-\alpha_1}\kappa\sigma_\varepsilon^2\sigma_\beta^2$, i.e., this responsiveness is itself a function of κ. Combining yields κ in terms of model primitives:

$$\kappa = \frac{\gamma}{1 + \gamma\frac{\alpha_1}{1-\alpha_1}\sigma_\varepsilon^2\sigma_{\hat{\beta}}^2} \ . \tag{24}$$

The interpretation of the expression is straightforward. The numerator captures the direct (partial equilibrium) effect of risk, which is simply the coefficient of relative risk aversion. The denominator captures the general equilibrium effects that work through the resource allocation, which smooth TFP and hence lower the extent of aggregate risk. The larger the heterogeneity across firms, the greater the opportunities for reallocation and hence the larger these effects.

2.2 Aggregate Risk, Monetary Policy and TFP

At the heart of our model is the risk-return tradeoff in firm-level capital choices, which depends on the nature of aggregate risk in the economy as captured by κ. Expression (21) makes explicit the key role of monetary policy in determining the slope of this tradeoff: monetary policy pins down the behavior (i.e., the cyclicality) of the output gap, μ_a, and thus directly affects κ. Intuitively, the cyclicality of the economy, e.g., aggregate output – the source of aggregate risk – depends on both the natural rate of output and the output gap. A procyclical output gap increases the cyclicality of realized output and thus the amount of aggregate risk; a countercyclical output gap has the opposite effects. Formally, $\frac{\partial\kappa}{\partial\mu_a} > 0$. Thus, the allocational effects of policy boil down to its impact on a single object, κ.

Expressions (19) and (20) show that the effects of policy feed through to the resource allocation and via this channel, to the behavior of TFP, i.e., ψ_a and $\overline{\psi}$. Through its choice of the output gap, the monetary authority determines (i) the effective degree of aggregate risk facing firms and the slope of the risk-return tradeoff, (ii) the amount of risk-taking and the micro-level capital allocation and (iii) the dynamics of TFP. As an example, consider the case of more aggressive countercyclical policy, i.e., a lower value of μ_a and hence κ. From (19), such a policy reduces the slope of firm-level $mrpk$ on beta, which causes a reallocation of capital towards more cyclical high beta firms, increasing the covariance of firm-level capital and beta and lowering mean $mrpk$ dispersion. From (17) and (20), this reallocation leads to (i) a higher

value of ψ_a – TFP becomes more volatile as more cyclical firms become larger – and (ii) a higher value of $\overline{\psi}$ – the long-run level of TFP increases as $mrpk$ dispersion falls. Intuitively, more aggressive stabilization by the central bank mitigates the extent of aggregate risk by reducing the cyclicality of aggregate output. This incentivizes further risk-taking on the part of the private sector, which leads to more volatile TFP, but also a more productive allocation.

A striking implication is that monetary policy has permanent effects – the long-run level of TFP (and hence output/consumption) is in part determined by the extent of stabilization, which, by influencing the capital allocation and dispersion in $mrpk$, moves the economy closer to/further from its production possibilities frontier. In the heterogeneous firm environment, aggregate production possibilities depend not only on the state of aggregate technology, but also on how resources are allocated at the micro-level. Thus, monetary policy is not neutral in the long-run but instead affects the economy's long-run level of production via the resource allocation. More formally, we have $\frac{\partial \overline{\psi}}{\partial \mu_a} < 0$, which shows that long-run TFP depends on the monetary policy regime. The result holds despite our standard formulation of nominal rigidities.

A second implication is that through the endogenous behavior of TFP, firm heterogeneity and reallocation opportunities dampen the effectiveness of stabilization policies. In particular, the standard deviation of output is equal to

$$\sigma\left(y_t\right) = \left(\phi_\psi \psi_a + \mu_a\right) \sigma_\varepsilon \ ,$$

which shows that output fluctuations stem from movements in TFP and the output gap. In a representative firm environment, ψ_a is simply equal to one and outside the influence of policy – the central bank can act to smooth output through its choice of μ_a leaving the dynamics of TFP (e.g., the natural rate of output) unchanged. Here, in contrast, TFP depends on the actions of the central bank. As the central bank acts to smooth fluctuations through a more countercyclical output gap (lower/more negative μ_a), it reduces aggregate risk (κ), which causes capital to shift towards more procyclical firms, increasing ψ_a and so the volatility of TFP. More formally, we have $\frac{\partial \psi_a}{\partial \mu_a} < 0$, i.e., increased risk-taking by the private sector in response to the attempted stabilization mitigates the effects of the policy. Thus, if the central bank is targeting the overall volatility of output, a more aggressive countercyclical policy is required. This is a form of the Lucas Critique at work – in the representative firm setup, the dynamics of TFP are assumed to be exogenous and invariant to policy; in the heterogenous firm economy, this is no longer the case and the central bank must take into account the effects of its actions on TFP and the natural rate of output.[19]

[19]It is straightforward to show that $\frac{\partial \sigma(y_t)}{\partial \mu_a} > 0$, i.e., the direct stabilizing effect of countercyclical policy on output always dominates the indirect effect through the response of TFP. The result simply implies that such policies are less effective when the private sector is able to reallocate resources in response to the policy rule.

The effects of monetary policy are independent of the capital wedge and hold even in economies that are otherwise undistorted. As evidenced by (21) distortions in capital markets are also a source of aggregate risk and impact the risk-return tradeoff – e.g., a countercyclical distortion that worsens in bad times ($\tau_{\Lambda a} > 0$) increases the effective amount of risk that firms face with all the accompanying ramifications for the allocation and TFP. The result also bears relevance for the large literature studying the relationship between volatility and growth – comparative statics exercises show that $\frac{\partial \sigma(y_t)}{\partial \mu_a} > 0$ and $\frac{\partial \overline{\psi}}{\partial \mu_a} < 0$, i.e., less countercyclical policy leads to higher output volatility and lower long-run TFP/output, while $\frac{\partial \sigma(y_t)}{\partial \tau_{\Lambda a}} < 0$ and $\frac{\partial \overline{\psi}}{\partial \tau_{\Lambda a}} < 0$, i.e., larger countercyclical capital distortions dampen both. Thus, the two distortions – fluctuations in the output gap and the capital wedge – can point to either a positive or negative sign on this relationship, which depends crucially on the nature of the distortions at work in the economy.

3 Optimal Policy

In this section, we study the optimal conduct of monetary policy in the presence of heterogeneity and distortions to the capital allocation. Appendix C derives the following second order approximation to the welfare loss function, expressed in terms of the equivalent consumption decline measured as a fraction of consumption in the non-stochastic steady state:

$$\mathbb{W} = (1-\rho)\,\mathbb{E}_{-1}\sum_{t=0}^{\infty}\rho^t\left[-\overline{\psi} + \frac{1}{2}(\gamma-1)\,\phi_\psi\left(\psi_t - \overline{\psi}\right)^2 + \frac{1}{2}\frac{1}{\phi_l}\mu_t^2 + \frac{1}{2}\frac{\alpha_2\nu_w}{\lambda_w\phi_l}\left(\pi_t^w\right)^2\right], \tag{25}$$

which expresses the loss as a function of (i) the long-run level of TFP, (ii) the volatility of TFP, (iii) the volatility of the output gap, and (iv) the volatility of wage inflation, all with appropriate weights. It is straightforward to verify that the last two terms correspond exactly to the welfare function in the textbook representative firm New Keynesian model and capture the standard losses from output gap and inflation fluctuations, respectively, conditional on the dynamics of TFP.[20] The effects of heterogeneity enter the loss function through the first two terms, which are new to our setting – welfare is increasing in the level of TFP and decreasing in its volatility.[21] With heterogeneity, these two terms are endogenous, dependent on the conduct of monetary policy and are not independent – indeed, they are both pinned down by the risk adjustment, κ. Optimal policy thus trades off four considerations: the volatility of inflation

Further, the result highlights that monetary policy has different effects on TFP and output: the former is only affected through the allocation, while the latter is also affected directly by the policy.

[20] For example, compare these terms to the welfare function in Galí (2015), equation (26) of Chapter 6, and note that $\frac{1}{\phi_l} = \gamma + \frac{\varphi+1-\alpha_2}{\alpha_2}$.

[21] We can also interpret the first two terms as capturing the level and volatility of the natural rate of output, y_t^n: multiply and divide to obtain $-\frac{1}{\phi_\psi}\overline{y^n} + \frac{\text{var}(y_t^n)}{2}\frac{\gamma-1}{\phi_\psi}$, where $\overline{y^n}$ is the long-run mean of the natural rate.

and the output gap as in a representative firm model and the level and volatility of TFP.

The policy maker chooses the behavior of the output gap to minimize the welfare loss subject to the responses of the private sector to its choice, i.e., the definitions of $\overline{\psi}$ and ψ_a and the Phillips curve. To obtain intuitive expressions, for our analytic results in this section we study optimal policies among the class that satisfy the simple rule in (13). This is the set of policies that are implementable by standard Taylor rules.[22] In our numerical work in Section 4 we solve for the unrestricted full commitment optimal policy and show that it is approximated quite well by the optimal one among the restricted class studied here.

Optimal risk adjustment. Before turning to our analysis of monetary policy, we study as a benchmark the allocation chosen by a social planner in two cases: first, when the planner has access to enough instruments to correct both distortions in the economy, i.e., sticky prices and the capital wedge, and second, a constrained planner who faces, but takes as given, the sticky price distortion, μ_t. Because the allocation is summarized by κ, the planner's solution can be characterized by the choice of this single object. The solution is the same in both cases: the planner sets $\kappa^* = \kappa_\psi \psi_a^*$, where ψ_a^* is the value of ψ_a in an economy with no distortions.

The result has two implications. First, in the absence of distortions, the economy is at first-best, yet the value of κ is not zero. The economy faces a tradeoff between the long-run level of TFP and its volatility, but this tradeoff is efficient. Reducing the risk adjustment further is possible and may be effective in raising long-run TFP/output, but would cause a shifting of capital towards more procyclical firms and inefficient volatility of TFP. Increasing the risk adjustment would have opposite effects, smoothing TFP but reducing its long-run level. Thus, even at first-best, the economy features both TFP volatility and marginal product dispersion. Both are symptoms of the economy's efficient response to fundamental shocks.

Second, even when aggregate dynamics are distorted, i.e., due to sticky prices, the constrained optimal risk adjustment is independent of those distortions. In other words, even though the economy may exhibit inefficiencies on other other margins (e.g., labor supply), the optimal κ does not further distort the allocation of capital across firms.

Optimal monetary policy. As a single policy tool, optimal monetary policy does not attain the first-best value for κ. Intuitively, monetary policy alone cannot fully correct the inefficiencies in the allocation without further distorting the economy on other margins and must balance these considerations. Appendix C.1 derives the optimal policy, which is fully characterized by

[22] We derive the mapping between (13) and interest rate rules in Appendix B.2. More generally, this is the class in which policy is a function only of current state variables, but cannot condition on lagged states. In contrast, the optimal policy under full commitment that we solve for in Section 4 will require committing to a response to the infinite sequence of realized shocks.

the elasticity of the output gap to the realization of the exogenous shock:

Proposition 4. *Optimal monetary policy satisfies*

$$\mu_t = \mu_a^* a_t \qquad \textit{where} \qquad \mu_a^* = -\tau_{\Lambda a}\sigma_\beta^2 \Phi_\Lambda \tag{26}$$

and $\Phi_\Lambda > 0$ *is a constant composite of model parameters.*

To build intuition, consider first the representative firm model where $\sigma_\beta^2 = 0$. Clearly, we have $\mu_a^* = 0$: TFP is exogenous and thus outside the control of policy, optimal policy completely stabilizes both inflation and the output gap, and the economy is at first-best. Thus, in the usual way, the 'divine coincidence' holds and the central bank faces no tradeoff in setting policy. The result also holds in the presence of heterogeneity, but when the capital distortion is absent (i.e., $\tau_{\Lambda a} = 0$), but for a different reason – in this case, TFP is endogenous but other than the output gap distortion it is efficient, and again the central bank is able to achieve the first-best by eliminating the output gap. In other words, when the flexible price allocation is efficient, output gap fluctuations are unambiguously distortionary and an extended version of the divine coincidence holds – complete stabilization of the output gap eliminates distortionary fluctuations in labor supply and inflation in the usual way, but now also ensures the micro-allocation and aggregate TFP attain the first-best.

This logic breaks down when firms are heterogeneous ($\sigma_\beta^2 > 0$) and the allocation is distorted ($\tau_{\Lambda a} \neq 0$). If the distortion is countercyclical, i.e., $\tau_{\Lambda a} > 0$, it strengthens the incentives for countercyclical policy. From expression (21), the wedge generates an inefficiently high risk adjustment, which leads to an overly conservative allocation, an inefficiently low cyclicality of TFP and excessively high marginal product dispersion, depressing the level of TFP. Thus, the optimal policy response entails more aggressive countercyclical policy – specifically, a countercyclical output gap (negative value for μ_a^*) that leans against cyclical fluctuations to partially offset these effects. The opposite holds if the capital distortion is procyclical.

Because $\mu_a^* \neq 0$ in the distorted heterogeneous firm economy, optimal policy allows for fluctuations in inflation and the output gap. By doing so, it helps correct the inherent inefficiencies in the allocation. However, policy does not deviate so far from complete stabilization as to fully correct the allocational distortions – doing so would lead to an overly cyclical output gap that would lead to costly volatility in that object and inflation.[23] Thus, the divine coincidence does not hold – optimal policy takes on an intermediate value that balances allocational considerations against inflation/output gap volatility.[24]

[23] More formally, we can show that $|\mu_a^*| < |-\frac{1}{\kappa_l}\tau_{\Lambda a}|$, which from (21) is the value of μ_a that would completely correct the capital wedge.

[24] In this sense, capital market distortions in conjunction with heterogeneity generate an endogenous cost-

Finally, note that the (absolute value of the) cyclicality of the output gap under the optimal policy is strictly increasing in the extent of heterogeneity, i.e., $\frac{\partial \mu_a^*}{\partial \sigma_\beta^2} < (>)\ 0$ if $\tau_{\Lambda a} > (<)\ 0$, which shows that the policy response to the real friction is more aggressive when there is more heterogeneity – with larger differences across firms, the distortionary effects of the wedge on the allocation and TFP are more costly relative to output gap and inflation volatility and optimal policy calls for larger deviations from complete stability in response. In contrast, in the case of no heterogeneity, the capital wedge has no effects at all and optimal policy can safely ignore it.

Optimal nominal interest rate. Appendix C.1 derives the nominal interest rate that implements the optimal output gap as a function of the exogenous shock:

$$i_t^* = \left(\Phi^i + i_a^*\right) a_t \qquad \text{where} \qquad i_a^* = \tau_{\Lambda a} \sigma_\beta^2 \Phi_\Lambda^i, \quad \Phi_\Lambda^i > 0\ .$$

Further, we can show $\frac{\partial i_a^*}{\partial \sigma_\beta^2} > (<)$ if $\tau_{\Lambda a} > (<)\, 0$. The result aligns closely with Proposition 4: (i) the cyclicality of the optimal nominal rate is a linear and increasing (in absolute value) function of the capital wedge, so that a countercyclical wedge ($\tau_{\Lambda a} > 0$) leads to a more aggressive rise in the nominal rate in response to expansionary shocks and (ii) the strength of this effect is increasing in heterogeneity, i.e., for a given value of the wedge, more heterogeneity leads the optimal nominal rate to be more procyclical. Indeed, i_a^* is zero in the absence of heterogeneity, in which case $i_t^* = \Phi^i a_t$, which sets the nominal rate (which is also the real rate since $\mu_t = 0$ and thus there is no inflation) equal to the natural interest rate. The same result clearly holds in an economy with nominal rigidities but no capital market distortion. The cyclicality of the nominal rate thus provides a natural metric to gauge the cyclicality of policy.

Costs of policy mistakes. As we have seen, the combination of firm heterogeneity and distortions to the resource allocation changes the prescription for optimal policy. A corollary to the result is that deviations from the optimal policy are more costly in the presence of heterogeneity (with or without capital distortions). Formally, assume that policy is set according to (13) where $\mu_a \neq \mu_a^*$, i.e., policy is not set to the optimal one.[25] Appendix C.1 proves that the welfare losses from such a deviation are increasing in the extent of heterogeneity, i.e., $\frac{\partial(\mathbb{W}(\mu_a) - \mathbb{W}(\mu_a^*))}{\partial \sigma_\beta^2} > 0$.

push shock. Indeed, defining x_t as the difference between actual and efficient output, i.e., $x_t = y_t - y_t^e$, we can write the Phillips curve as $\pi_t^w = \rho \mathbb{E}_t\left[\pi_{t+1}^w\right] + \lambda_w x_t + \lambda_w \phi_\psi \left(\psi_t^e - \psi_t\right)$ where the last term resembles a standard cost-push shock. Different than an exogenous cost-push shock, however, (e.g., of the type studied in Section 3.1), (i) this 'shock' is endogenous and depends on the conduct of policy and (ii) because $\psi_t^e - \psi_t$ enters the welfare function in addition to its effects on inflation, x_t is not the main welfare-relevant object; rather μ_t is.

[25] We can write such a policy as $\mu_t = \left(\mu_a^* + e_a\right) a_t$, where the first term in the parentheses is the optimal policy and the second term is an error, which when positive implies that the output gap is not countercyclical enough relative to the optimum and when negative the opposite. Note that the error captures a systematic bias in policy, i.e., it is known and anticipated by the private sector.

The intuition is clear: in a representative firm environment, the costs of policy mistakes stem from their effects on inflation and the output gap. With heterogeneity, these mistakes lead to a sub-optimal amount of aggregate risk that distorts the resource allocation. As an example, consider the case where $\mu_a > \mu_a^*$, i.e., the output gap is less countercyclical than the optimal policy prescribes. In addition to the costs of higher output gap/inflation volatility (which are independent of heterogeneity), aggregate output/consumption are also overly volatile and hence there is excessively high aggregate risk. The private sector responds by taking on a more conservative capital allocation, which reduces TFP and welfare.

3.1 Other Distortions/Policies

Thus far, we have focused on cyclical capital market imperfections that directly distort the capital allocation and the role of monetary policy as a stabilization tool in an economy with nominal rigidities. The effects of aggregate risk and firm heterogeneity are not limited to these considerations. Here, we show that heterogeneity affects the optimal policy response to a broader class of distortions, i.e., labor market distortions and traditional cost-push shocks that generate an output/inflation tradeoff even in the absence of heterogeneity and/or additional distortions. We also show that our main results hold when fiscal policy (modeled as cyclical labor income taxes) is the tool of stabilization, both in flexible and sticky price economies.

Additional distortions. We add two additional distortions: (i) a cost-push shock to the Phillips curve that generates a trade-off for the central bank even in the absence of heterogeneity and (ii) a labor market distortion ('labor wedge') that generates inefficient movements in labor supply.[26] The Phillips curve and aggregate labor supply conditions then take the form:

$$\begin{aligned} \pi_t^w &= \rho \mathbb{E}_t \left[\pi_{t+1}^w\right] + \lambda_w \mu_t + \eta_t \\ w_t + \frac{1}{\phi_l}\mu_t + \tau_{lt} &= \gamma y_t + \varphi l_t \ , \end{aligned}$$

which are augmented versions of (15) and (2). We specify the cost-push shock as a constant elasticity function of a_t, i.e., $\eta_t = \eta_a a_t$, where $\eta_a > 0$ captures a procyclical shock to (wage) inflation and $\eta_a < 0$ the opposite. We specify the labor wedge as $\tau_{lt} = \tau_{la} a_t$, where $\tau_{la} > 0$ captures a countercyclical distortion to labor supply (i.e., labor supply is inefficiently procyclical).

[26] Together, the capital and labor wedges and cost-push shock span the set of possible aggregate distortions to the three choice variables, aggregate labor, wage setting and capital allocation, i.e., although additional distortions could be added, they would be redundant.

With these additional distortions, optimal policy takes the form (proof in Appendix C.1):

$$\mu_a^* = -\tau_{\Lambda a}\sigma_\beta^2\Phi_\Lambda - \eta_a\Phi_\eta - \tau_{la}\Phi_l, \qquad \text{where} \qquad \Phi_\eta > 0, \quad \frac{\partial\Phi_\eta}{\partial\sigma_\beta^2} < 0, \quad \Phi_l > 0, \quad \frac{\partial\Phi_l}{\partial\sigma_\beta^2} > 0\ , \quad (27)$$

which is an extended version of (26) that incorporates the optimal policy response to the capital wedge (which is the same as before), but additionally to the labor wedge and cost-push shock.

Because of the cost-push shock, in the standard way, the central bank faces a tradeoff and cannot simultaneously stabilize inflation and the output gap. Stabilizing the output gap requires setting $\mu_a^* = 0$, but this implies complete accommodation of the cost-push shock. In reverse, stabilizing inflation requires setting $\mu_a^* = -\frac{\eta_a}{\lambda_w} \neq 0$, so clearly allows for fluctuations in the output gap. The slope of the optimal tradeoff between inflation and the output gap is given by Φ_η. Expression (27) shows that Φ_η is strictly decreasing in σ_β^2 – with more heterogeneity, the central bank responds less aggressively to the inflationary pressure from the cost-push shock. The typical cost of offsetting this shock is output gap volatility; here, there is an additional cost – allowing μ_a to depart from zero to lean against the shock distorts the capital allocation. For example, consider the case where $\tau_{\Lambda a} = 0$. The allocation and dynamics of TFP are efficient without the effects of policy and the impact of the policy response to the cost-push shock on these margins is unambiguously distortionary. Thus, heterogeneity reduces the magnitude of the optimal response to the cost-push shock and affects the tradeoff between inflation and output gap volatility (even with no additional distortions). The central bank must account for the fact that responding to this shock induces an inefficient reallocation of capital across firms.

The effect of the labor wedge is similar to the capital wedge: if the wedge is countercyclical ($\tau_{la} > 0$), it further strengthens the incentives for countercyclical policy.[27] For example, consider the case with no other distortions and flexible prices, i.e., $\tau_{\Lambda a} = \eta_a = \lambda_w = 0$. We can show that optimal policy entails completely neutralizing the labor distortion, i.e., $\mu_a^* = -\tau_{la}$.[28] With nominal rigidities, the optimal response is less than one-for-one, i.e., $\Phi_l < 1$: countercyclical policy aimed to stabilize the labor wedge generates costly inflation volatility. The positive derivative of Φ_l with respect to σ_β^2 shows that more heterogeneity strengthens the response of policy to the labor wedge. Intuitively, the cost of the wedge is increasing in heterogeneity: it not only has a direct effect on the cyclicality of labor supply, but also leads to inefficient aggregate risk, which distorts the allocation and dynamics of TFP. This latter effect is larger when there are more opportunities for reallocation, i.e., σ_β^2 is large. Thus, optimal policy responds more aggressively than in the case of a representative firm.

[27] A large body of work dating back at least to Chari et al. (2007) documents a countercyclical labor wedge.
[28] Of course, with flexible prices, monetary policy cannot achieve this outcome. However, fiscal policy can.

Fiscal policy. The main insights do not hinge on nominal rigidities, or monetary policy as the instrument of stabilization – here we show that similar results go through with cyclical fiscal policy, both in sticky and flexible price economies.

The fiscal authority chooses a cyclical labor income tax/subsidy, which in some abuse of notation, we denote τ_{lt}. Optimal fiscal policy sets the cyclicality of the tax, i.e., a value τ_{la} such that $\tau_{lt} = -\tau_{la} a_t$, to maximize household welfare. The sign convention implies that $\tau_{la} < 0$ captures a procyclical tax, i.e., a tax rate that increases in expansions and falls in downturns. $\tau_{la} > 0$ implies the opposite. It is straightforward to verify (see Appendix C.1) that with flexible prices, the equilibrium and policy objective function are special cases of the environment laid out above where (i) the nominal side of the economy is completely disentangled from real quantities and the macro dynamics are fully characterized by (9) and (14) with the output gap driven by the labor tax so that $\phi_l \tau_{la}$ replaces μ_a and (ii) the welfare function takes the form in (25) with no costs of inflation. From here, we can show that the optimal policy takes an analogous form to (26), i.e., $\tau^*_{la} = -\tau_{\Lambda a} \sigma^2_\beta \Phi^f_\Lambda$ where $\Phi^f_\Lambda > 0$ and $\frac{\partial \tau^*_{la}}{\partial \sigma^2_\beta} < 0$ if $\tau_{\Lambda a} > 0$.

The intuition is the same as with monetary policy. In the absence of distortions, the optimal policy is a laissez-faire one – the policy-maker sets the tax to zero. With countercyclical distortions, i.e., $\tau_{\Lambda a} > 0$, the optimal policy is countercyclical (a procyclical tax rate) – the distortion implies an inefficiently high amount of aggregate risk and the fiscal authority reduces the cyclicality of labor supply in order to (partially) correct this inefficiency, even though such a policy distorts labor supply. The strength of the policy response is increasing in the degree of heterogeneity. Even with flexible prices, however, the policy-maker cannot replicate the first-best allocation – although within the set of attainable allocations, doing so would be overly costly due to the distortionary effect on labor supply margin.[29]

Fiscal-monetary coordination. A last case we consider is when fiscal and monetary policy are set optimally in tandem, i.e., there is coordination between the fiscal and monetary authorities. In this case, we can prove (i) optimal monetary policy sets $\mu_a = 0$, i.e., completely stabilizes inflation, and (ii) optimal fiscal policy is the same as in the flexible price economy. Thus, when both fiscal and monetary policy are jointly put to work, a natural ordering emerges: first, monetary policy is set to replicate the flexible price outcome. Then, fiscal policy is set as it would be if prices were indeed truly flexible.

[29] Replicating the first-best capital allocation would entail setting τ_{la} such that $\tau_{\Lambda a} + \kappa_l \tau_{la} = 0$ or $\tau_{la} = -\frac{1}{\kappa_l} \tau_{\Lambda a}$. We can show, however, that optimal fiscal policy under flexible prices is more aggressively countercyclical than the corresponding monetary policy in the sticky price economy. This is because there are no costs of inflation in the flexible price economy and hence this countervailing force is absent in the policy-maker's objective.

4 Quantitative Exercise

In this section, we provide a numerical evaluation of the policy effects studied analytically in the last section.

First, we return to the more general case of persistent aggregate shocks, i.e., we allow δ in (11) to be non-zero. With this modification, the equilibrium conditions in Proposition 2 remain unchanged with the exception of aggregate TFP, which now takes the form

$$\psi_t = \overline{\psi} + \delta a_{t-1} + \psi_a \varepsilon_t \ , \tag{28}$$

which is easily shown to be an ARMA(1,1). In other words, although the exogenous shock follows an AR(1), endogenous TFP follows an altered dynamic process. The other macroeconomic variables, such as output and labor, follow a similar process.

We assume that monetary policy in the baseline equilibrium follows a Taylor rule in expected price inflation and the output gap (see, e.g., Clarida et al. (2000)), given by

$$i_t = \phi_y \mu_t + \phi_\pi \mathbb{E}_t \left[\pi^p_{t+1} \right] \ . \tag{29}$$

Appendix D shows that such a rule implies an output gap that satisfies

$$\mu_t = \mu_{a_{-1}} a_{t-1} + \mu_a \varepsilon_t \ , \tag{30}$$

i.e., the coefficients in the rule map to a pair of values $\mu_{a_{-1}}$ and μ_a that govern the response of the output gap to the two natural state variables of the model, a_{t-1} and ε_t.[30]

Last, rather than restricting the form of optimal policy to a rule like (13), we solve for the unrestricted optimal policy under full commitment (derivation in Appendix C.2). We discuss the important role of commitment in more detail below and additionally show that the appropriate choice of coefficients in a simple rule of the type in (13) studied above can approximate quite well the full unrestricted policy.

4.1 Calibration

We begin by assigning values to the more standard preference and production function parameters of our model. Given our focus on capital investment decisions, we assume a period length of one year and set the annual discount factor, ρ, to 0.96. We set the inverse Frisch elasticity of labor supply, φ, to 1 and risk aversion, γ, to 10. Although this is somewhat high for the macro literature (and at the upper bound of what is typically deemed the 'reasonable' range,

[30] We obtain similar results using a rule in the output gap and realized wage inflation.

e.g., Mehra and Prescott (1985)), it is a standard value (indeed, at the lower end) in the finance literature studying issues of risk premia (e.g., Bansal and Yaron (2004)) as do we.[31] We assume constant returns to scale in production and set α_1 and α_2 to values of one-third and two-thirds, respectively. We set the substitution parameter across intermediate goods to $\nu = 0.8$, which is a common value in the literature, e.g., Atkeson and Burstein (2010). Following, e.g., Broer et al. (2020) and Galí (2015), we set the elasticity of substitution across labor types, ν_w, to 6. The coefficients in the Taylor rule are set to $\phi_y = 0.5$ and $\phi_\pi = 1.5$.

The remaining parameters are calibrated jointly to match five salient moments of the data. The parameters governing the aggregate shock process, δ and σ_ε^2, are chosen to match the persistence of observed aggregate TFP and the standard deviations of TFP growth rates, respectively, which yields $\delta = 0.7$ and $\sigma_\varepsilon = 0.05$.[32] The wage adjustment cost parameter, θ_w, is chosen to match the slope of the wage Philips curve. Specifically, we regress wage inflation on measures of the output gap both in the model and data, and set θ_w so that the estimated coefficients are identical.[33] Note that the coefficient from this regression does not directly map into a structural parameter, since from equation (15), inflation expectations, which are correlated with the output gap, are in the error term. Rather, the identification is indirect and follows from matching a salient moment from the model and data. This procedure yields a value of $\theta_w = 866$ to match the flat slope of the empirical wage Phillips curve (see, e.g., Hazell et al. (2020) for related evidence on the flat slope of the price Phillips curve).

The key new parameters of the model are (i) the degree of heterogeneity across firms, $\sigma_{\hat{\beta}}$, and (ii) the cyclicality of the capital market distortion, $\tau_{\Lambda a}$. We set $\sigma_{\hat{\beta}}$ to match the observed dispersion in cyclicality among Compustat firms. Specifically, we estimate firm-by-firm time-series regressions of firm-level productivity growth on aggregate TFP growth. The coefficients from these regressions yield an observable measure of firm cyclicality. We denote this 'observed beta' β_i^{obs} (relative to the 'true beta', β_i). Because TFP is not equal to the exogenous shock, $\beta_i^{obs} \neq \beta_i$. However, Appendix D shows how we can use the equilibrium process on TFP in (28) to derive an adjustment factor mapping the two, allowing us to recover the true betas (up to

[31] Our results are not sensitive to using even much higher values for γ. The reason is that due to partial equilibrium and general equilibrium effects, γ appears both in the numerator and denominator of κ (see (24) for a simple example and (65) for the exact expression) and κ asymptotes to a constant when γ becomes large. Thus, further increases have little effect. We have also experimented with Epstein-Zin preferences: increasing γ under these preferences while holding the elasticity of intertemporal substitution fixed also has little effect.

[32] We use annual HP-filtered aggregate TFP calculated from data on real GDP and aggregate capital and labor from the Bureau of Economic Analysis, which has a standard deviation of about 0.03.

[33] We use the difference between potential real GDP as computed by the BEA and realized real GDP as a measure of the output gap. Using various (annual, HP-filtered) measures of wages from the Bureau of Labor Statistics yields slope coefficients ranging from about 0.1 (average hourly earnings of production and nonsupervisory employees) to 0.3 (business sector compensation per hour). We target a slope of 0.2, approximately the midpoint of this range.

an additive constant). An important concern is that these estimates may be affected by issues of sampling and measurement error. Appendix D outlines an approach to adjust the estimated distribution for these types of errors. Implementing this procedure yields a value of $\sigma_{\hat{\beta}} = 3.2$ (the value without correcting for measurement/sampling error would have been over 7).[34]

Last, to pin down the capital distortion, $\tau_{\Lambda a}$, we continue to use the micro-data and directly target the covariance of firm-level *mrpk* and beta. This is the slope of the micro-level risk-return tradeoff that is at the heart of the allocational mechanism in our theory. For intuition, consider the following regression of *mrpk* on beta:

$$mrpk_{it} = \text{const.} + \lambda_\beta \beta_i + \xi_{it} \ . \tag{31}$$

From (19), we have that the time-averaged coefficient from such a regression is approximately $\lambda_\beta \approx \kappa \sigma_\varepsilon^2$.[35] Given values for the other parameters, we can invert the expression for κ (e.g., (21)) and solve for $\tau_{\Lambda a}$, which is the only remaining unknown. The estimated $\tau_{\Lambda a}$ is increasing in the estimated regression coefficient, i.e., holding other parameters fixed, a steeper slope of $mrpk_{it}$ on β_i implies a larger risk adjustment driven by a more countercyclical capital distortion.

Properly implementing the regression in (31) requires a panel of betas. To obtain such a panel, we estimate shorter horizon (10 year) backwards-looking rolling window regressions of firm-level TFP growth on aggregate TFP growth. However, just as with our estimates of the cross-sectional dispersion in betas, estimation of (31) is complicated by issues of measurement and sampling error in these measures of beta, which would lead to attenuation bias in the resulting coefficient.[36] To address this challenge, we develop a two-stage instrumental variable approach in which we instrument for the observed TFP betas (β_i^{obs}) we estimate from production-side data with stock market betas calculated from financial data on firm and aggregate stock market returns. First, we estimate firm-level stock market betas, β_i^s, using a standard market model of the form $r_{it} = \text{const} + \beta_{it}^s r_{mt} + \zeta_{it}$, which relates firm-level stock returns, r_{it}, to the aggregate market return, r_{mt} (with exposure β_{it}^s) and an idiosyncratic component, ζ_{it}. Again, we estimate this model using backwards-looking rolling window regressions for each individual firm (we use quarterly data and 10 year horizons), which yields a panel of stock market betas, β_{it}^s. We then use these values as instruments for the observed TFP betas, β_{it}^{obs}.

[34] Firm-level data are obtained from Compustat over the period 1964-2019. We include firms with at least 40 observations (the data are at the annual frequency). The estimated dispersion is on par with that in David et al. (2021) who find a value of 3.4, which they estimate using stock market returns in conjunction with a structural model (there are other differences in the estimation approaches as well, e.g., assumptions on curvature, the sample of firms studied and the frequency of the data).

[35] Specifically, the coefficient from a single cross-section is approximately $\kappa\sigma_\varepsilon^2 + \varepsilon_t \neq \kappa\sigma_\varepsilon^2$. However, pooling observations over time eliminates the mean zero term ε_t. To derive the approximation, we use $\mathbb{E}_{t-1}\left[\tilde{A}_{it}\right] \approx 1 + \mathbb{E}_{t-1}\left[\tilde{a}_{it}\right]$. We provide the exact expression in Appendix D.

[36] Error in measured *mrpk* is less of a concern since it is the left-hand side variable.

Table 1: Slope of MRPK on Observed Betas

	$mrpk_{it+1}$	
	OLS	**IV**
	(1)	(2)
β_{it}^{obs}	0.008***	0.116***
	[4.09]	[3.07]
Observations	71546	57165
First stage F-statistic		16.86

Notes: This table reports regressions of firm *mrpk* on observed beta, β^{obs}. Column (1) displays OLS estimates. Column (2) displays IV estimates with stock market betas as an instrument. Standard errors are clustered two ways by firm-year. t-statistics in brackets. Significance levels denoted by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Specifically, the first-stage regression is one of the observed TFP betas, β_{it}^{obs}, on stock market betas, β_{it}^{s}. The second stage regression is of $mrpk_{it}$ on the predicted values from the first stage (to avoid issues of simultaneity, we use one period lagged betas in the second stage regression). Scaling by the adjustment factor mapping the observed β_{i}^{obs} to the true β_i yields the slope of *mrpk* on beta. This two-stage approach is designed to address measurement concerns in the direct estimates of TFP betas – as long as the errors are uncorrelated across the production and financial-side data, using the covariance between the two measures will eliminate them.[37]

Table 1 displays the results of these regressions. Column (1) presents the OLS estimate from a regression of $mrpk_{it+1}$ on the directly observed β_{it}^{obs}. Column (2) presents the IV estimates (we also include a full set of industry-by-year fixed effects). Although the OLS estimate is positive and statistically significant at standard levels, the IV estimate is substantially larger, in line with the presence of measurement errors that artificially amplify beta dispersion and attenuate the OLS coefficient.[38] The IV strategy addresses this concern. Note that this approach does not constitute a test of a particular asset pricing theory, but rests only on the fact that the exposure of firm-level variables to aggregate shocks should be reflected in the sensitivity of stock market returns. This connection is evidenced by the large first-stage F-statistic reported in Table 1, which verifies that stock market betas are a strong instrument for the production-side betas. The coefficient estimate is also significant in economic magnitude: a one-unit increase in the observed cyclicality measure, β_{it}^{obs}, is associated with a 12% increase in MRPK. The adjustment factor mapping β_{it}^{obs} to the true β_{it} is roughly 0.44 and so a one unit increase in β_{it} is associated with a 5% increase in MRPK. Because this slope is steeper than can be explained by preferences and aggregate dynamics alone, the data point to inefficiently low risk-taking at the micro-level

[37] In the case of positively correlated errors, the regression estimate would be biased downward, leading our results to be conservative.

[38] Indeed, our direct correction for sampling error in calculating the cross-sectional dispersion accounts for roughly one-half of the difference in the OLS and IV estimates.

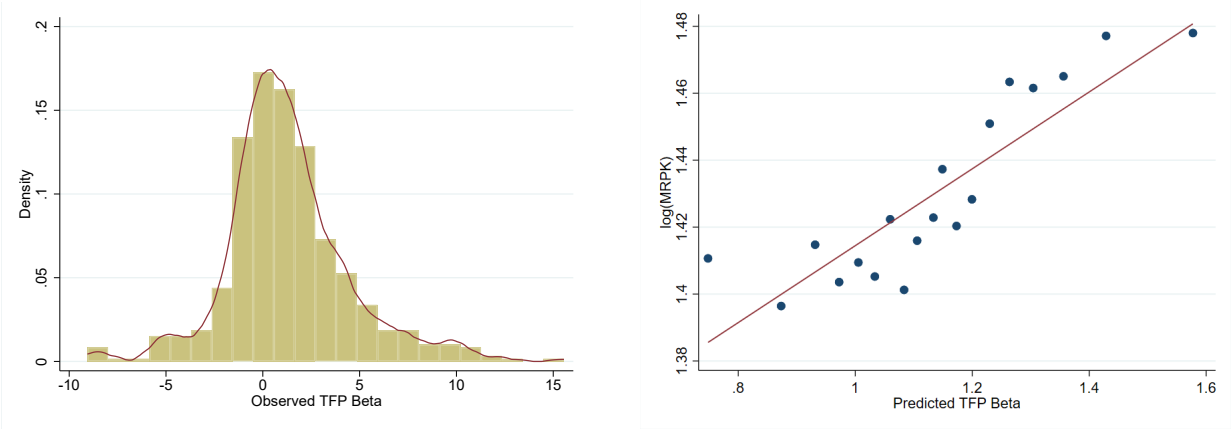

Figure 1: Firm Cyclicality and MRPK

Notes: Panel A displays the cross-sectional distribution of observed firm cyclicality (β_i^{obs}) obtained from regressions of firm-level productivity growth on aggregate TFP growth as described in the text. The vertical bars show the histogram of the distribution and the solid lines the result of kernel smoothing with a bandwidth of 0.25. Panel B displays a bin-scatter of $mrpk_{it+1}$ on the predicted values of β_{it}^{obs} obtained from a regression of β_{it}^{obs} on stock market betas, β_{is}^{s}.

and a strongly countercyclical capital market distortion.

Figure 1 illustrates the two main features of the micro-data that are informative for σ_β and $\tau_{\Lambda a}$. The left-hand panel plots the cross-sectional distribution of the raw estimates of β_i^{obs} (i.e., before adjusting for sampling error) and shows the wide dispersion in observable measures of firm cyclicality. The right-hand panel of the figure displays a bin-scatter of $mrpk_{it+1}$ against the predicted values of β_{it}^{obs} from the first stage of the IV (because the predicted values are linear functions of the instrument, the plot is essentially showing the reduced form of the IV) and illustrates the strong positive relationship between MRPK and beta. Table 2 summarizes the full set of parameter values and Appendix D provides further details of the calibration strategy.

4.2 Policy Evaluation

Table 3 presents the equilibrium and counterfactual policy exercises. Each column displays welfare losses (top panel) and a number of equilibrium statistics (bottom panel) under alternative policy regimes. We report the total welfare loss, as well as a decomposition of the loss into its four components: the level of TFP and the volatilities of TFP, the output gap and (wage) inflation. We report four salient statistics of the equilibrium under each policy: the reduction in TFP volatility relative to the case with no risk adjustment in the capital allocation, i.e., where $\kappa = 0$ (denoted $\Delta\sigma(\psi_t)$), the volatilities of the output gap and inflation, and lastly, a measure of the cyclicality of monetary policy, namely, the elasticity of the nominal interest rate

Table 2: Calibration – Summary

Parameter	Description	Value
Preferences		
ρ	Discount factor	0.96
φ	Inverse Frisch elasticity	1
γ	Risk aversion	10
Production		
α_1	Capital share	1/3
α_2	Labor share	2/3
ν	Intermediate good substitutability	0.8
δ	Persistence of agg. shock	0.7
σ_ε	Std. dev. of agg. shock	0.05
$\sigma_{\hat{\beta}}$	Std. dev. of risk exposures	3.2
Wage-setting & distortions		
ν_w	Labor elasticity of substitution	6
θ_w	Wage adjustment cost	866
$\tau_{\Lambda a}$	Capital wedge	14.7

to the realization of TFP (denoted ϵ_{i_t,ψ_t}).[39] Column (1) corresponds to the equilibrium under the Taylor rule; column (2) reports results from the first-best allocation; column (3) reports results under optimal policy with full commitment; and column (4) when monetary policy is set to the optimal one in the representative firm case, i.e., when the central bank acts as if there was a representative firm, ignoring micro-level heterogeneity and allocational concerns.

The results suggest (i) firm-level heterogeneity and risk can have sizable effects on TFP dynamics and welfare, and (ii) accounting for this heterogeneity can have important implications for the conduct of monetary policy through its effects on the resource allocation.

First, column (1) shows that under the Taylor rule, long-run TFP is lower by almost 1.4% (relative to the case with no risk adjustment in the allocation). At the same time, the volatility of TFP is also lower, by about 29%. The welfare costs of depressed TFP are directly equal to the TFP loss itself. The welfare costs of TFP volatility turn out to be relatively small. Column (2), the first-best allocation, provides a natural benchmark for these values. In this case, TFP losses are extremely small, only about 0.003%. TFP is more volatile, only about 9% less so than in the case with no risk adjustment. These findings imply that due to capital market distortions, the equilibrium allocation is inefficiently conservative – there is an excessive shifting of capital towards less cyclical firms, which reduces TFP volatility relative to the first-best, but generates excessively high marginal product dispersion, which reduces the level of long-run TFP. In contrast, the first-best allocation features a more productive allocation of

[39]The case with $\kappa = 0$ is equivalent to one with a representative firm facing the same exogenous shocks. We can also show that this is the allocation that maximizes expected TFP, taken to an appropriate power.

Table 3: Heterogeneity and Optimal Monetary Policy

	Taylor Rule (1)	First-Best (2)	Optimal Policy (3)	Ignoring Hetero. (4)
Welfare loss (%)				
Total	1.784	0.193	1.146	1.393
TFP level	1.357	0.003	0.925	1.385
TFP volatility	0.009	0.190	0.026	0.008
Output gap volatility	0.100	0.000	0.171	0.000
Inflation volatility	0.318	0.000	0.025	0.000
Equilibrium statistics				
$\Delta\sigma(\psi_t)$ (%)	-28.88	-8.59	-26.72	-28.97
$\sigma(\mu_t)$	1.34	0.00	1.69	0.00
$\sigma(\pi_t^w)$	0.28	0.00	0.08	0.00
ϵ_{i_t,ψ_t}	0.25	-	0.78	-0.25

capital with higher long-run TFP, but also higher TFP volatility. In total, welfare under the Taylor rule is about 1.6% (1.8% − 0.19%) lower than in the first-best. Strikingly, of the total loss, the vast majority – roughly 85% (calculated as $\frac{1.36\%-0.003\%}{1.6\%}$) – is due to the losses in long-run TFP. Since the first-best allocation features higher TFP volatility than the Taylor rule, equilibrium welfare losses from this source are smaller than in the first-best. The remaining welfare difference is due to changes in inflation and output gap volatility – these are modest but non-negligible under the Taylor rule and zero in the first-best, which is completely undistorted.

Second, the results suggest an important role for policy to improve on equilibrium outcomes and further, highlight the importance of accounting for heterogeneity when setting optimal policy. Column (3) shows that relative to the Taylor rule in column (1), the optimal monetary policy increases long-run TFP by 0.44%. The central bank achieves these gains through a more countercyclical policy, which reduces the extent of aggregate risk and induces the private sector to take on an allocation that more closely aligns firm-level capital and productivity – for example, the elasticity of the nominal interest rate to the realization of TFP is roughly three times larger under optimal policy than under the Taylor rule, i.e., the optimal nominal rate is significantly more procyclical. At the same time, there is an increase in TFP volatility, but this offsetting effect turns out to be small (the standard deviation of TFP is about 2% higher than under the Taylor rule), as are the resulting welfare losses. The total welfare gain from implementing the optimal policy relative to the Taylor rule is about 0.65%, of which about two-thirds is due to the effects on long-run TFP. Thus, the optimal policy eliminates about 40% of the gap between equilibrium and first-best welfare ($\frac{1.80\%-1.15\%}{1.80\%-0.19\%}$). Monetary policy cannot achieve the first-best since using cyclical policy to influence the resource allocation also affects

the output gap and inflation (under this calibration, optimal policy reduces inflation volatility but increases output gap volatility relative to the Taylor rule).

Finally, how does the presence of heterogeneity change the conduct of optimal policy? Comparing columns (3) and (4) shows that if the central bank were to set policy to the optimal one ignoring heterogeneity – which, in this simple environment entails complete stabilization of inflation and the output gap – the welfare gain relative to the Taylor rule is about 0.4%, almost entirely due to the elimination of fluctuations in these variables. The properties of TFP change only minimally so that the central bank is almost entirely missing out on the gains from improving the resource allocation (indeed, the level of TFP under this policy is slightly lower than the Taylor rule). Thus, accounting for micro-level heterogeneity makes a significant contribution to the potential gains from policy, about 0.25% ($1.40\% - 1.15\%$) of lifetime consumption in the steady state. These gains come wholly from an improved allocation and higher long-run TFP – TFP under optimal policy is almost 0.5% higher than under the policy ignoring heterogeneity. The cost of this gain is somewhat higher volatility on all dimensions, with inflation volatility being the costliest form. Again, the central bank achieves this gain through a more aggressively countercyclical policy – the interest rate elasticity to TFP is not procyclical enough when not accounting for heterogeneity (indeed, the optimal nominal rate ignoring heterogeneity tracks the natural interest rate, which in this setup is countercyclical).

The role of commitment and simple rules. Monetary policy affects the capital allocation primarily through anticipation effects – in particular, it is not unexpected policy shocks that shape the allocation, but rather the private sector's expectations of the systematic stabilization component of policy. Thus, the ability to commit to future policy is crucial to the policy maker's ability to affect the allocation. To see this most clearly, consider the case with no commitment, which is the polar opposite of the full commitment case in Table 3. In the absence of commitment, the policy maker cannot influence private sector expectations and simply acts to maximize welfare on a period-by-period basis, but must take as given the resource allocation and dynamics of TFP. Thus, without the ability to commit, the policy maker's problem is exactly the same as the case of discretion in the representative firm environment and the best it can do is set $\mu_t = 0$, i.e., completely stabilize the output gap and inflation. Because this is precisely the same outcome as the case of ignoring heterogeneity in column (4) of Table 3, we can equivalently interpret the welfare losses in that column relative to the optimal policy in column (3) as the welfare costs of lack of commitment.

Although commitment is crucial, it turns out that full commitment to responses to the entire sequence of realized shocks is not. To see this, consider a 'simple' rule of the form in (28), which sets the output gap as a function of the lagged and current realizations of the shock only, the

two state variables in the economy. This form includes the set of policies implementable with standard Taylor rules. In Appendix C.2 we solve for the optimal policy within this class and find that welfare losses are only negligibly larger than under the unconstrained policy in Table 3. The result implies that the lion's share of the allocational gains from optimal policy can be attained by the appropriate choice of coefficients in a simple rule that conditions only on current states. Full commitment to the entire sequence of shocks adds only little.[40]

Fiscal policy. In Appendix C.2, we show that similar results hold for optimal fiscal policy. First, even under flexible prices and no cyclical fiscal policy, there are large welfare losses relative to the first-best, which are largely driven by depressed TFP. Second, in response to countercyclical capital market distortions, optimal policy is strongly countercyclical (i.e., sets a countercyclical output gap via a procyclical labor income tax). The welfare gains from such a policy are significant and stem largely from increasing TFP (by roughly 0.5%). Last, we find that the gains from fiscal/monetary coordination are modest: once monetary policy is optimally determined accounting for heterogeneity, there is only small scope for further welfare improvements from labor market fiscal policies.

4.3 Financial Frictions and Capital Wedges

In this section, we delve into two detailed examples of financial frictions and directly evaluate their role in contributing to the capital market distortion we infer from the firm-level micro-data. Specifically, we build on the examples outlined in Section 2 and study (i) a model of frictional financial intermediation and (ii) a TANK model featuring limited asset market participation.

Frictional intermediation. We consider the GK model extension laid out in Section 2 (full details of the intermediary block are in Appendix A.1). From (7), the capital distortion reflects the shadow marginal value of net worth to financial intermediaries, $\frac{\partial V_t}{\partial N_t}$. Due to homogeneity of the intermediary value function, we can show $\frac{\partial V_t}{\partial N_t} = \theta \frac{Q_t}{N_t}$, i.e., the unobservable shadow value can be related to the observable (book) leverage ratio, $\frac{Q_t}{N_t}$, scaled by the limit on collateral, θ. Using this result and log-linearizing, we can write the unexpected shock to the distortion as[41]

$$\tau_{\Lambda t} - \mathbb{E}_{t-1}\left[\tau_{\Lambda t}\right] = -\tau_{\Lambda a}\varepsilon_t \ ,$$

where the key elasticity, $\tau_{\Lambda a}$, depends on the parameters of the financial sector. In Appendix A we show how we can pin down these parameters using two steady state moments: the observed

[40] The result echoes previous findings of small gains from commitment vs. optimal simple rules in representative agent New Keynesian models (e.g., Taylor (2007) and Taylor and Williams (2010)).

[41] From (10), the anticipated piece of the distortion will cancel and thus has no effect on the allocation.

book leverage ratio of the financial sector and the price/earnings ratio of the aggregate market. Thus, we can calculate a value for $\tau_{\Lambda a}$ (as implied by this specific channel) using a wholly different approach and different data types than in our baseline analysis, which relied on a more indirect (but likely more general) method from firm-level micro-data. Using data from the Federal Reserve Flow of Funds, the average book leverage ratio in recent years has been about 12. Using data from Robert Shiller's website, the average price/earnings ratio has been about 23.[42] Calibrating the financial sector parameters so that the steady state matches these moments, we solve the model including the GK block numerically and obtain a value for $\tau_{\Lambda a}$ of about 3.6, which represents about one-quarter of the estimate in Table 2.[43] The result suggests that frictions in the financial intermediation process along the lines of the GK model imply a quantitatively significant capital wedge.

Limited participation. Next, we consider the TANK model laid out in Section 2 (full details in Appendix A.2). From (8), the capital distortion is a function of two new objects, namely, the share of households that are financially constrained and do not participate in capital markets, θ, and shocks to the price markup, τ_t^p. Denoting the cyclicality of these shocks as τ_a^p (so that $\tau_t^p = \tau_a^p a_t$), we can use (8) to write the unexpected shock to the distortion as

$$\tau_{\Lambda t} - \mathbb{E}_{t-1}\left[\tau_{\Lambda t}\right] = -\underbrace{\frac{\alpha_2}{1-\alpha_2}\gamma\theta\tau_a^p}_{\tau_{\Lambda a}}\varepsilon_t \ . \tag{32}$$

Expression (32) directly relates the capital wedge to two observables – the share of households who participate in capital markets and the cyclicality of price markups – and thus suggests a calibration strategy matching these two moments. To place a value on θ, we use data from the Survey of Consumer Finances, which reports that over the past few decades, approximately 50% of households held stock either directly or indirectly through actively managed mutual funds, index funds and retirement plans.[44] This statistic implies a value of θ of 0.50.

The parameter τ_a^p represents the elasticity of the price markup to the exogenous shock, a_t. Although this shock is unobservable, the stochastic processes of the markup and endogenous TFP (expression (28)) imply that a regression of the former on the latter yields a coefficient $\beta_{\tau^p,\psi} = \tau_a^p \frac{\delta^2(1-\psi_a)+\psi_a}{\delta^2(1-\psi_a^2)+\psi_a^2}$. As is standard in this class of model, the markup is equal to the inverse of labor's share of income.[45] Hence, we can estimate the parameter τ_a^p from a regression of the

[42] Flow of funds data were obtained from https://www.federalreserve.gov/releases/z1/ and price/earnings data from http://www.econ.yale.edu/~shiller/data.htm.

[43] Note that since the aggregate capital stock is fixed in our model, the dynamics of the intermediary sector only affect the other aggregate variables through the wedge $\tau_{\Lambda t}$.

[44] Data were obtained from https://www.federalreserve.gov/econres/scfindex.htm, Table 7.

[45] Indeed, it is variation in labor's share that is important, not the markup interpretation per se.

(log) inverse labor share on endogenous TFP and applying this expression, which translates the coefficient into the elasticity on a_t. The regression of inverse labor share on TFP yields a statistically significant coefficient of 0.41, which translates into a value of τ_a^p of 0.36.[46] Using this value in (32), we obtain a value of $\tau_{\Lambda a}$ of 3.7, slightly over one-quarter of our baseline estimate in Table 2. Thus, the combination of limited asset market participation and fluctuations in labor's share of income lead to a significant capital wedge. Taken together, the two detailed frictions studied here account for about one-half of the total wedge that we measure from the firm-level data, which illustrates their important role, but also suggests other factors are likely at play in the data as well.

5 Conclusion

In this paper, we have studied the implications of firm heterogeneity – specifically, differences in cyclicality – for business cycle dynamics and optimal monetary stabilization. The heterogeneous firm economy can be recast in a representative firm formulation but where the resource allocation and hence aggregate TFP are endogenous. The monetary policy regime determines the nature of aggregate risk and in part shapes the allocation and dynamic behavior of TFP, i.e., its long-run level and cyclical volatility. Empirically estimated capital market frictions lead optimal policy to be more aggressively countercyclical than in an observationally equivalent representative firm model. Thus, firm heterogeneity tends to strengthen the rationale for such stabilization policies. A quantitative exercise suggests that the welfare gains from implementing policies that account for allocational considerations can be significant.

We have deliberately kept our framework simple in order to highlight the new insights while taking only small departures from textbook business cycle models. A fruitful, though challenging next step would be to add additional ingredients that enable the model to match a wider set of business cycle and micro-level moments, e.g., adjustment costs, financial frictions, more complicated preferences, etc., and evaluate the effects of heterogeneity in a state-of-the-art quantitative DSGE model. Of particular interest would be the implications for capital accumulation and the dynamics of aggregate investment, as well as the inclusion of additional distortions/shocks that have been highlighted in the literature (for example, such as those studied qualitatively in Section 3.1). One broader lesson of our paper is that understanding the properties of inefficiencies – in particular, their heterogeneous effects – is crucial to reaching accurate conclusions regarding effective macroeconomic policies.

[46] We use data on labor share for employees in the non-financial corporate sector from the BEA (downloaded from the St. Louis FRED database). The coefficient estimate is likely conservative, as a similar regression in growth rates (with a suitably modified correction term) implies a larger value for τ_a^p.

References

Acemoglu, D. and F. Zilibotti (1997): "Was Prometheus unbound by chance? Risk, diversification, and growth," *Journal of Political Economy*, 105, 709–751.

Acharya, S., E. Challe, and K. Dogra (2020): "Optimal Monetary Policy According to HANK," CEPR Discussion Paper No. DP14429.

Aghion, P., G.-M. Angeletos, A. Banerjee, and K. Manova (2010): "Volatility and growth: Credit constraints and the composition of investment," *Journal of Monetary Economics*, 57, 246–265.

Angeletos, G.-M. and J. La'O (2020): "Optimal monetary policy with informational frictions," *Journal of Political Economy*, 128, 1027–1064.

Atkeson, A. and A. Burstein (2010): "Innovation, Firm Dynamics, and International Trade," *Journal of Political Economy*, 118, 433–484.

Auclert, A. (2019): "Monetary Policy and the Redistribution Channel," *American Economic Review*, 109, 2333–67.

Auclert, A., M. Rognlie, and L. Straub (2018): "The intertemporal keynesian cross," Tech. rep., National Bureau of Economic Research.

Bansal, R. and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59.

Baqaee, D., E. Farhi, and K. Sangani (2021): "The Supply-Side Effects of Monetary Policy," Tech. rep., National Bureau of Economic Research.

Barlevy, G. (2004): "The cost of business cycles under endogenous growth," *American Economic Review*, 94, 964–990.

——— (2005): "The cost of business cycles and the benefits of stabilization," *Economic Perspectives-Federal Reserve Bank of Chicago*, 29, 32.

Barone-Adesi, G., L. Mancini, and H. Shefrin (2012): "Behavioral finance and the Pricing Kernel Puzzle: Estimating risk aversion, optimism, and overconfidence," *Working paper*.

Bhandari, A., D. Evans, M. Golosov, and T. J. Sargent (2021): "Inequality, Business Cycles, and Monetary-Fiscal Policy," *Econometrica*, 89, 2559–2599.

Bilbiie, F. O. (2008): "Limited asset markets participation, monetary policy and (inverted) aggregate demand logic," *Journal of Economic Theory*, 140, 162–196.

——— (2018): "Monetary policy and heterogeneity: An analytical framework," CEPR Discussion Paper No. DP12601.

Bilbiie, F. O. and X. Ragot (2020): "Optimal monetary policy and liquidity with heterogeneous households," *Review of Economic Dynamics*.

Broer, T., N.-J. H. Hansen, P. Krusell, and E. Oberg (2020): "The New Keynesian Transmission Mechanism: A Heterogeneous-Agent Perspective," *The Review of Economic Studies*, 87, 77–101.

Caramp, N. and D. H. Silva (2021): "Monetary Policy and Wealth Effects: The Role of Risk and Heterogeneity," Tech. rep., Working Paper.

Challe, E. (2020): "Uninsured unemployment risk and optimal monetary policy in a zero-liquidity economy," *American Economic Journal: Macroeconomics*, 12, 241–83.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2007): "Business Cycle Accounting," *Econometrica*, 75, 781–836.

Cieslak, A. and A. Vissing-Jorgensen (2021): "The economics of the Fed put," *The Review of Financial Studies*, 34, 4045–4089.

Clarida, R., J. Gali, and M. Gertler (2000): "Monetary policy rules and macroeconomic stability: evidence and some theory," *The Quarterly Journal of Economics*, 115, 147–180.

David, J. M., L. Schmid, and D. Zeke (2021): "Risk-Adjusted Capital Allocation and Misallocation," Working paper, University of Southern California.

David, J. M. and V. Venkateswaran (2019): "The sources of capital misallocation," *American Economic Review*, 109, 2531–67.

Debortoli, D. and J. Galí (2018): "Monetary policy with heterogeneous agents: Insights from TANK models," Universitat Pompeu Fabra Economics Working Paper No. 1686.

Gali, J. (1994): "Keeping up with the Joneses: Consumption externalities, portfolio choice, and asset prices," *Journal of Money, Credit and Banking*, 26, 1–8.

Galí, J. (2015): *Monetary Policy, Inflation, and the Business Cycle: an Introduction to the New Keynesian Framework and its Applications*, Princeton University Press.

Garga, V. and S. R. Singh (2021): "Output hysteresis and optimal monetary policy," *Journal of Monetary Economics*, 117, 871–886.

Gertler, M. and P. Karadi (2011): "A model of unconventional monetary policy," *Journal of Monetary Economics*, 58, 17–34.

González, B., G. Nuño, D. Thaler, and S. Albrizio (2022): "Firm heterogeneity, capital misallocation and optimal monetary policy," Banco de espana working paper.

Hall, R. E. (2011): "The high sensitivity of economic activity to financial frictions," *The Economic Journal*, 121, 351–378.

Hazell, J., J. Herreno, E. Nakamura, and J. Steinsson (2020): "The slope of the Phillips Curve: evidence from US states," Tech. rep., National Bureau of Economic Research.

Hsieh, C.-T. and P. J. Klenow (2009): "Misallocation and manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 124, 1403–1448.

Jones, L. E., R. E. Manuelli, H. E. Siu, and E. Stacchetti (2005): "Fluctuations in convex models of endogenous growth, I: Growth effects," *Review of Economic Dynamics*, 8, 780–804.

Jordà, Ò., S. R. Singh, and A. M. Taylor (2020): "The long-run effects of monetary policy," Tech. rep., National Bureau of Economic Research.

Kaplan, G. (2021): "What has the HANK Literature Taught Us About Monetary Policy," Heterogeneity in Macroeconomics: Implications for Policy, Federal Reserve Bank of New York.

Kaplan, G., B. Moll, and G. L. Violante (2018): "Monetary Policy According to HANK," *American Economic Review*, 108, 697–743.

Kekre, R. and M. Lenel (2021): "Monetary policy, redistribution, and risk premia," Tech. rep., National Bureau of Economic Research.

Koren, M. and S. Tenreyro (2007): "Volatility and development," *The Quarterly Journal of Economics*, 122, 243–287.

——— (2013): "Technological diversification," *American Economic Review*, 103, 378–414.

Kormendi, R. C. and P. G. Meguire (1985): "Macroeconomic determinants of growth: cross-country evidence," *Journal of Monetary Economics*, 16, 141–163.

Kurtzman, R. J. and D. Zeke (2020): "Misallocation Costs of Digging Deeper Into the Central Bank Toolkit," *Review of Economic Dynamics*.

La'O, J. and A. Tahbaz-Salehi (2020): "Optimal Monetary Policy in Production Networks," Tech. rep., National Bureau of Economic Research.

Longstaff, F. A. and M. Piazzesi (2004): "Corporate earnings and the equity premium," *Journal of Financial Economics*, 74, 401–421.

Mehra, R. and E. C. Prescott (1985): "The equity premium: A puzzle," *Journal of Monetary Economics*, 15, 145–161.

Miller, M., P. Weller, and L. Zhang (2002): "Moral Hazard and The US Stock Market: Analysing the 'Greenspan Put'," *The Economic Journal*, 112, C171–C186.

Moran, P. and A. Queralto (2018): "Innovation, productivity, and monetary policy," *Journal of Monetary Economics*, 93, 24–41.

Obstfeld, M. (1994): "Risk-Taking, Global Diversification, and Growth," *American Economic Review*, 84, 1310–1329.

Ottonello, P. and T. Winberry (2020): "Financial heterogeneity and the investment channel of monetary policy," *Econometrica*, 88, 2473–2502.

Poole, W. (2008): "Market bailouts and the "Fed Put"," *Federal Reserve Bank of St. Louis Review*, 90.

Ramey, G. and V. A. Ramey (1991): "Technology commitment and the cost of economic fluctuations," *NBER working paper*.

——— (1995): "Cross-Country Evidence on the Link Between Volatility and Growth," *The American Economic Review*, 85, 1138–1151.

Restuccia, D. and R. Rogerson (2008): "Policy distortions and aggregate productivity with heterogeneous establishments," *Review of Economic Dynamics*, 11, 707–720.

Rubbo, E. (2020): "Networks, Phillips Curves and Monetary Policy," Tech. rep., mimeo, Harvard University.

Shefrin, H. (2008): *A behavioral approach to asset pricing*, Elsevier.

Taylor, J. B. (2007): *Monetary policy rules*, vol. 31, University of Chicago Press.

Taylor, J. B. and J. C. Williams (2010): "Simple and robust rules for monetary policy," in *Handbook of monetary economics*, Elsevier, vol. 3, 829–859.

Woodford, M. (2003): *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press: Princeton, NJ.

——— (2010): "Optimal Monetary Stabilization Policy," in *Handbook of Monetary Economics*, Elsevier, vol. 3, 723–828.

Appendix: For Online Publication

A Capital Wedges – Examples

This appendix provides detailed examples of frictions that lead to a capital wedge of the form described in Sections 2 and 4.3.

A.1 Frictional Financial Intermediation

The setup builds closely on that in Gertler and Karadi (2011). At any point in time, a fraction f of households work as bankers that manage a financial intermediary. Intermediary earnings are transferred back to households and thus households effectively own the intermediaries, i.e., they hold all claims on intermediary equity and liabilities. Bankers exit exogenously with probability $1-\sigma$ and are immediately replaced by new bankers, keeping the relative proportion of bankers fixed. The average survival time of a banker is equal to $\frac{1}{1-\sigma}$.

Intermediaries own the capital stock, which they rent to firms for use in production. Firms finance their capital by issuing equity claims to intermediaries equal to the amount of capital to be financed, K_{it+1}, at price Q_{it}. Thus intermediaries receive both the return on capital and any economic profits that stem from firm monopoly power, which sum to firm operating profits Π_{it+1} as defined in (5). Intermediaries borrow from households at the risk-free rate in order to provide funding to firms.

The balance sheet of intermediary j is given by

$$\int Q_{it} S_{ijt} di = N_{jt} + D_{jt} \tag{33}$$

Intermediary assets consist of the total market value of its ownership claims on firms, where Q_{it} and S_{ijt} denote the price and quantity of claims held by intermediary j on firm i. Intermediary liabilities consist of deposits, D_{jt}, and book equity, or net worth, N_{jt}. Net worth is equal to the gross return on assets less costs of borrowing:

$$N_{jt} = \int R^e_{it} Q_{it-1} S_{ijt-1} di - R_t D_{jt-1} \tag{34}$$

Combining (33) and (34), we have

$$N_{jt+1} = \int R^e_{it+1} Q_{it} S_{ijt} di + \left(N_{jt} - \int Q_{it} S_{ijt} di \right) R_{t+1} \tag{35}$$

The intermediary acts to maximize the expected discounted stream of dividends payed out to households:

$$V_{jt} = \mathbb{E}_t \left[\Lambda_{t+1} \left((1-\sigma) N_{jt+1} + \sigma V_{jt+1} \right) \right] \tag{36}$$

Due to a moral hazard/costly enforcement problem, intermediaries face collateral constraints that limit their ability to obtain deposits. Specifically, the intermediary can divert a fraction θ of its assets, which leads to the the following incentive constraint that limits its collateral:

$$V_{jt} \geq \theta \int Q_{it} S_{ijt} di \tag{37}$$

The intermediary chooses its holdings of firms' securities, S_{ijt}, to maximize (36) subject to (35)

and (37). Substituting for N_{jt+1}, we can set up the Lagrangian as

$$\mathcal{L} = \max_{S_{ijt}} \; V_{jt} + \zeta_t \left(V_{jt} - \theta \int Q_{it} S_{ijt} di \right)$$

which yields a first order condition

$$\frac{\partial V_{jt}}{\partial S_{ijt}} = \theta \frac{\zeta_t}{1+\zeta_t} Q_{it}$$

and using the fact that

$$\frac{\partial V_{jt}}{\partial S_{ijt}} = Q_{it} \mathbb{E}_t \left[\Lambda_{t+1} \left(1 - \sigma + \sigma \frac{\partial V_{jt+1}}{\partial N_{jt+1}} \right) \left(R^e_{it+1} - R_{t+1} \right) \right]$$

we can write the first order condition as

$$\mathbb{E}_t \left[\Lambda_{t+1} \left(1 - \sigma + \sigma \frac{\partial V_{jt+1}}{\partial N_{jt+1}} \right) \left(R^e_{it+1} - R_{t+1} \right) \right] = \theta \frac{\zeta_t}{1+\zeta_t}$$

As in GK, we can show that the intermediary problem is homogeneous of degree one in N_{jt} and thus $\frac{\partial V_{jt+1}}{\partial N_{jt+1}} = \frac{\partial V_{t+1}}{\partial N_{t+1}} = \frac{V_{t+1}}{N_{t+1}}$, where variables without j subscripts denote aggregates of the entire intermediary sector. Using this, we have

$$\mathbb{E}_t \left[\Lambda_{t+1} T_{\Lambda t+1} \left(R^e_{it+1} - R_{t+1} \right) \right] = \theta \frac{\zeta_t}{1+\zeta_t}$$

where

$$T_{\Lambda t} = 1 - \sigma + \sigma \frac{\partial V_t}{\partial N_t}$$

which shows that the relevant SDF pricing assets takes the form described in the text and is equal to the household SDF, Λ_t, multiplied by a wedge, $T_{\Lambda t}$, that captures the shadow marginal value of net worth to the intermediary sector. Combining this result with the optimality conditions for K_{it+1} from firm profit maximization yields an equation analogous to (6).

Aggregation and equilibrium dynamics of $T_{\Lambda t}$. The remainder of this section pertains to the analysis in Section 4.3.

The aggregate net worth of intermediaries satisfies

$$N_t = \sigma \left(\int R^e_{it} Q_{it-1} S_{it-1} di + \left(N_{t-1} - \int Q_{it-1} S_{it-1} di \right) R_t \right) + N_{et}$$

where N_{et} denotes the wealth of new bankers. Defining $Q_t = \int Q_{it} di$ and the total return as $R^e_t = \frac{\int R^e_{it} Q_{it-1} di}{\int Q_{it-1} di}$ and noting that since intermediaries hold all financial claims, $S_{it} = 1$, we can write this as

$$N_t = \sigma \left(R^e_t Q_{t-1} + \left(N_{t-1} - Q_{t-1} \right) R_t \right) + N_{et}$$

and defining aggregate book leverage as $\phi_t = \frac{Q_t}{N_t}$,

$$N_t = \sigma N_{t-1} \left(\left(R^e_t - R_t \right) \phi_{t-1} + R_t \right) + N_{et}$$

Following GK, we assume that $N_{et} = n_e Q_t$, i.e., a fraction n_e of the total value of the assets held by

the intermediary sector is held by new bankers.

Using homogeneity of the value function, we can write the aggregate value of intermediaries and the shadow value of net worth as

$$\begin{aligned} V_t &= N_t \mathbb{E}_t \left[\Lambda_{t+1} T_{\Lambda t+1} \left(\left(R^e_{t+1} - R_{t+1} \right) \phi_t + R_{t+1} \right) \right] \\ T_{\Lambda t} &= 1 - \sigma + \sigma \frac{V_t}{N_t} \end{aligned}$$

The return on the aggregate portfolio is equal to $R^e_{t+1} = \frac{\Pi_{t+1} + Q_{t+1}}{Q_t}$, where $\Pi_{t+1} = \int \Pi_{it+1} di$ is the aggregate over all firm profits.

Combining these expressions and assuming the collateral constraint binds so that $V_t = \theta Q_t$, the intermediary sector is characterized by

$$\begin{aligned} \mathbb{E}_t \left[\Lambda_{t+1} T_{\Lambda t+1} R_{t+1} \right] &= \frac{Q_t}{N_t} \left(\theta - \mathbb{E}_t \left[\Lambda_{t+1} T_{\Lambda t+1} \left(R^e_{t+1} - R_{t+1} \right) \right] \right) \\ T_{\Lambda t} &= 1 - \sigma + \sigma \theta \frac{Q_t}{N_t} \\ R^e_t &= \frac{\Pi_t + Q_t}{Q_{t-1}} \\ N_t &= \sigma \left(\left(R^e_t - R_t \right) Q_{t-1} + N_{t-1} R_t \right) + n_e Q_t \end{aligned}$$

which is a system of four equations in four unknowns – $N_t, Q_t, T_{\Lambda t}$ and R^e_t – given Π_t, R_t and Λ_t, which are pinned down from the firm and household problems. As discussed in the text, since the aggregate capital stock is fixed in our model, the dynamics of the intermediary sector only affect the other aggregate variables through $T_{\Lambda t}$.

Log-linearizing these equations, we obtain an expression for $\tau_{\Lambda t} \equiv \log T_{\Lambda t}$ of the form

$$\tau_{\Lambda t} - \mathbb{E}_{t-1} \left[\tau_{\Lambda t} \right] = \tau_{\Lambda a} \varepsilon_t$$

The elasticity to the shock, $\tau_{\Lambda a}$, depends on the parameters of the baseline model in the text and the three new parameters of the financial sector, θ, σ and n_e. First, we follow GK and calibrate σ directly to target an average horizon of bankers of a decade, which yields $\sigma = 0.972^4$. We pin down the remaining two parameters so that the steady state of the model matches (i) the average book leverage ratio of the intermediary sector, which corresponds to $\frac{Q}{N}$ and (ii) the price/earnings ratio on the aggregate stock market, which corresponds to $\frac{Q}{\Pi}$. We obtain data on the book leverage ratio from the Federal Reserve Flow of Funds Table F.108 ('Domestic Financial Sectors'), calculated as the value of financial assets divided by bank book equity. We use the average value over the period 2000-2019 of 12.3. This value is conservative – the post World War II average is 18.9 and the post-1980 average is even higher at 20.6 (the aggregate leverage ratio was extremely high in the 1980's and fell gradually through about 2000, the initial year we consider). We obtain data on the price/earnings ratio from Robert Shiller's website and use the average value since 1980 of about 23.

To see how we use these two moments, note that the steady state equations of the financial sector

are given by

$$\begin{aligned}
\theta &= \frac{\Lambda T_\Lambda R}{\frac{Q}{N}} + \Lambda T_\Lambda \left(R^e - R\right) \\
T_\Lambda &= 1 - \sigma + \sigma\theta\frac{Q}{N} \\
R^e &= \frac{1}{\frac{Q}{\Pi}} + 1 \\
1 &= \sigma\left(\left(R^e - R\right)\frac{Q}{N} + R\right) + n_e\frac{Q}{N}
\end{aligned}$$

which can be solved for the two unknowns, θ and n_e, as a function of known parameters and the two moments. Using these values, we solve the log-linearized system for $\tau_{\Lambda a}$.

A.2 Limited Asset Market Participation

The setup follows closely the two agent New Keynesian (TANK) model developed in Debortoli and Galí (2018). There are two types of households of time invariant measures θ and $1-\theta$, respectively. The first type – 'constrained' households – do not participate in financial markets and simply consume their labor income in each period. The second type – 'unconstrained' – hold all capital and equity shares in firms. Under Rotemberg wage setting frictions, wages are common across households. We assume that employment is uniformly distributed across households.

Because only unconstrained households own capital, the relevant SDF for pricing capital returns is given by $\Lambda_t^U = \rho\left(\frac{C_t^U}{C_{t-1}^U}\right)^{-\gamma}$, where C_t^U denotes the consumption of an unconstrained household. By definition, aggregate consumption is the sum of consumption across constrained and unconstrained households:

$$C_t = \theta C_T^K + (1-\theta)\, C_T^U$$

Define $G_t = \frac{C_t^U - C_t^K}{C_t^U}$ as the gap between the consumption of unconstrained and constrained households. We can rewrite this as

$$C_t = C_t^U\left(1 - \theta G_t\right)$$

and log-linearizing and rearranging yields an expression for unconstrained consumption as a function of aggregate consumption and the consumption gap

$$c_t^U = c_t + \frac{\theta}{1-\theta\overline{G}}g_t \tag{38}$$

where $\overline{G}$ denotes the gap in the non-stochastic steady state.

Next, we can also write the gap as a function of the ratio of profits to labor income:

$$G_t = 1 - \frac{C_t^K}{C_t^U} = \frac{\frac{\Pi_t}{W_t L_t}}{1 - \theta + \frac{\Pi_t}{W_t L_t}}$$

Firms face a common shock to their price markup, denoted T_t^p, and choose labor to satisfy

$$\max_{L_{it}} \; P_{it}Y_{it} - T_t^p W_t L_{it}$$

Taking first-order conditions, aggregating and rearranging yields

$$\frac{\Pi_t}{W_t L_t} = \frac{Y_t}{W_t L_t} - 1 = \frac{1}{\alpha_2} T_t^p - 1$$

where we assume a constant subsidy is in place to correct the steady state markup. Use this expression to rewrite the consumption gap as

$$G_t = \frac{\frac{1}{\alpha_2} T_t^p - 1}{\frac{1}{\alpha_2} T_t^p - \theta}, \qquad \overline{G} = \frac{\frac{1}{\alpha_2} - 1}{\frac{1}{\alpha_2} - \theta}$$

and log-linearizing,

$$g_t = \Theta \tau_t^p \quad \text{where} \quad \Theta = \frac{\frac{1}{\alpha_2}(1-\theta)}{\left(\frac{1}{\alpha_2} - \theta\right)\left(\frac{1}{\alpha_2} - 1\right)} > 0$$

Substituting into (38) gives

$$c_t^U = c_t + \frac{\alpha_2}{1-\alpha_2} \theta \tau_t^p$$

and from here, we can obtain an expression for the relevant SDF:

$$\begin{aligned} \log \Lambda_t^U &= \lambda_t + \log T_{\Lambda t} \\ &\text{where} \\ \lambda_t &= -\gamma (y_t - y_{t-1}) \\ \log T_{\Lambda t} &= -\frac{\alpha_2}{1-\alpha_2} \gamma \theta \Delta \tau_t^p \end{aligned}$$

which is expression (8).

A.3 Externalities in Preferences

Assume that the period utility function over consumption is given by

$$\frac{C_{jt}^{1-\gamma} C_t^{\gamma\lambda}}{1-\gamma}$$

a la Gali (1994), where C_{jt} is own consumption of individual j and C_t aggregate consumption.

The SDF for individual j is

$$\tilde{\Lambda}_t = \rho \frac{C_{jt}^{-\gamma} C_t^{\gamma\lambda}}{C_{jt-1}^{-\gamma} C_{t-1}^{\gamma\lambda}}$$

and since $C_{jt} = C_t$ in equilibrium, we have the relevant SDF as

$$\begin{aligned} \tilde{\Lambda}_t &= \rho \left(\frac{C_t}{C_{t-1}}\right)^{-\gamma} \left(\frac{C_t}{C_{t-1}}\right)^{\gamma\lambda} \\ &= \Lambda_t T_{\Lambda t} \end{aligned}$$

where $T_{\Lambda t} = \left(\frac{C_t}{C_{t-1}}\right)^{\gamma\lambda}$. Here, the wedge depends on the dynamics of aggregate consumption and the extent of the externality, i.e., λ.

A.4 Expectational Biases

Following Shefrin (2008) and Barone-Adesi et al. (2012), we define 'bias' or 'sentiment' as the ratio of a biased subjective probability distribution over aggregate shocks used by agents to the true probability distribution, i.e.,

$$\xi\left(\varepsilon_t\right) = \frac{f^b\left(\varepsilon_t\right)}{f\left(\varepsilon_t\right)}$$

where $f^b\left(\varepsilon_t\right)$ is the biased distribution and $f\left(\varepsilon_t\right)$ the true one. We parameterize the bias as

$$\xi\left(\varepsilon_t\right) = e^{-\frac{\mu}{\sigma_\varepsilon^2}\varepsilon_t - \frac{\mu^2}{2\sigma_\varepsilon^2}}$$

Then,

$$f^b\left(\varepsilon_t\right) = f\left(\varepsilon_t\right)\xi\left(\varepsilon_t\right) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} e^{-\frac{\left(\varepsilon_t+\mu\right)^2}{2\sigma_\varepsilon^2}}$$

i.e., the subjective distribution is normal with a shifted mean equal to $-\mu$ and unbiased variance σ_ε^2. If μ is positive, the bias is positive for ε_t small enough and negative for ε_t large enough implying that agents overweight bad states and underweight good ones. This is a case of over-pessimism. The opposite holds if μ is negative, which is a case of over-optimism.

The Euler equation for firm-level investment is then

$$R_{t-1}^K = \int \rho\left(\frac{C_{t+1}\left(\varepsilon_t\right)}{C_{t-1}}\right)^{-\gamma} MRPK_{it}\left(\varepsilon_t\right) f^b\left(\varepsilon_t\right) d\varepsilon_t$$

and substituting,

$$\begin{aligned} R_{t-1}^K &= \int \rho\left(\frac{C_{t+1}\left(\varepsilon_t\right)}{C_{t-1}}\right)^{-\gamma} MRPK_{it}\left(\varepsilon_t\right)\xi\left(\varepsilon_t\right) f\left(\varepsilon_t\right) d\varepsilon_t \\ &= \mathbb{E}_{t-1}\left[\rho\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma} T_{\Lambda t} MRPK_{it}\right] \end{aligned}$$

where the expectation is taken using the true probability distribution and the wedge is given by $T_{\Lambda t} = \xi\left(\varepsilon_t\right)$.

B Derivations and Proofs

This appendix provides detailed derivations and proofs for the results in Section 2.

B.1 Aggregation and Exact Solution for TFP

Here we prove Propositions 1 and 2. The proof of Proposition 2 details our assumptions on firm-level TFP in (11) and derives exact expressions for the terms κ, ψ_a and $\overline{\psi}$.

Proof of Proposition 1. To derive expression (10), aggregate the labor demand condition (4) to obtain

$$\alpha_2 \nu Y_t = W_t L_t \tag{39}$$

which implies $\frac{L_{it}}{L_t} = \left(\frac{Y_{it}}{Y_t}\right)^{\nu}$ and substituting into the firm-level production function,

$$Y_{it} = A_{it} K_{it}^{\alpha_1} \left(\left(\frac{Y_{it}}{Y_t} \right)^{\nu} L_t \right)^{\alpha_2}$$

so that

$$Y_{it}^{\nu} = \left(A_{it} K_{it}^{\alpha_1} \left(Y_t^{-\nu} L_t \right)^{\alpha_2} \right)^{\frac{\nu}{1-\alpha_2 \nu}}$$

Substituting into the final good production function (3) and rearranging yields

$$Y_t = \Psi_t K_t^{\alpha_1} L_t^{\alpha_2}$$

where

$$\Psi_t = \left(\int \left(A_{it}^{\frac{\nu}{1-\alpha_2 \nu}} \left(\frac{K_{it}}{K_t} \right)^{\alpha} \right) di \right)^{\frac{1-\alpha_2 \nu}{\nu}}$$

To solve for relative capital, rewrite (6) as

$$\alpha \mathbb{E}_{t-1} \left[\Lambda_t T_{\Lambda t} \Pi_{it} \right] = R_{t-1}^K K_{it}$$

and integrating and imposing market clearing,

$$\alpha \int \mathbb{E}_{t-1} \left[\Lambda_t T_{\Lambda t} \Pi_{it} \right] di = R_{t-1}^K K_t$$

Combining yields the right-hand expression in (10).

To derive (9), combine (39) with the labor supply condition (2) and resource constraint to obtain

$$L_t = \left(\frac{\alpha_2 \nu}{\chi} \right)^{\frac{1}{1+\varphi}} \mathcal{M}_t^{\frac{1}{1+\varphi}} Y_t^{\frac{1-\gamma}{1+\varphi}} \tag{40}$$

Substituting into the aggregate production function in (10) gives

$$Y_t = \left(\frac{\alpha_2 \nu}{\chi} \right)^{\phi_l} \Psi_t^{\phi_\psi} \mathcal{M}_t^{\phi_l} K_t^{\alpha_1 \phi_\psi} \tag{41}$$

where

$$\begin{aligned} \phi_\psi &= \frac{1}{1 - \alpha_2 \frac{1-\gamma}{1+\varphi}} \\ \phi_l &= \frac{\alpha_2}{1+\varphi} \frac{1}{1 - \alpha_2 \frac{1-\gamma}{1+\varphi}} \end{aligned}$$

Taking logs (and suppressing constants) yields (9).

The equilibrium is completed with the Phillips curve and consumption Euler equation. The house-

hold problem is standard. The household solves:

$$\max_{C_{jt},W_{jt}^n,B_{jt},K_{jt}} \mathbb{E}_{-1}\sum_{t=0}^{\infty}\rho^t\left[\frac{C_{jt}^{1-\gamma}}{1-\gamma}-\chi\frac{L_{jt}^{1+\varphi}}{1+\varphi}-\frac{\theta_w}{2}\left(\frac{W_{jt}^n}{W_{jt-1}^n}-1\right)^2 Y_t^{1-\gamma}\right]$$

$$s.t.$$

$$L_{jt}=\left(\frac{W_{jt}}{W_t}\right)^{-\nu_w}L_t$$

$$P_tC_{jt}+B_{jt}+P_t^K\left(K_{jt+1}-K_{jt}\right)=\left(1-\tau_{nt}\right)W_{jt}^nL_{jt}+\left(1+i_{t-1}\right)B_{jt-1}+R_{t-1}^KK_{jt}+\Pi_{jt}$$

where P_t denotes the price of the final good, W_{jt}^n the nominal wage, P_t^K the price of capital and Π_{jt} distributed profits. Note that we have scaled the wage adjustment costs by the marginal utility of consumption, $C_t^{-\gamma}$, in order to express them in utility units.

The Phillips curve is given by:

$$0=\nu_w\left(\frac{MRS_t}{W_t}-1\right)-\theta_w\left(\Pi_t^w-1\right)\frac{Y_t}{N_t}\frac{P_t}{P_{t-1}}\frac{1}{W_{t-1}}+\mathbb{E}_t\left[\Lambda_{t+1}\theta_w\left(\Pi_{t+1}^w-1\right)\Pi_{t+1}^w\frac{Y_{t+1}}{N_tW_t}\right] \tag{42}$$

where we have used the assumption of a constant labor subsidy $1-\bar{\tau}_n=\frac{\nu_w}{\nu_w-1}$ to correct the monopoly distortion and $MRS_t\equiv\frac{\chi L_t^{\varphi}}{C_t^{-\gamma}}$. Without wage stickiness, i.e., when $\theta_w=0$, we have $MRS_t=W_t$, i.e., the MRS always equals the real wage.

The Euler equation is given by:

$$C_t^{-\gamma}=\left(1+i_t\right)\rho\mathbb{E}_t\left[C_{t+1}^{-\gamma}\frac{1}{\Pi_{t+1}^p}\right] \tag{43}$$

where i_t denotes the nominal interest rate and Π_{t+1}^p gross price inflation.

□

Proof of Proposition 2. We prove that TFP takes the form in expression (28), which nests (14) in the case of i.i.d. shocks, i.e., $\delta=0$.

We assume firm-level productivity takes the form of (11), specifically,

$$a_{it}=\hat{\beta}_ia_t+\mathcal{O}_{it}$$

where $\mathcal{O}_{it}$ is defined as

$$\mathcal{O}_{it}=\xi_0+\xi_\beta\hat{\beta}_i^2+\xi_{\varepsilon^2}\varepsilon_t^2+\xi_{a_{-1}^2}\mathbb{E}_{t-1}\left[a_t\right]^2+\xi_{a_{-1},\varepsilon}\mathbb{E}_{t-1}\left[a_t\right]\varepsilon_t+\xi_{a_{-1}}\mathbb{E}_{t-1}\left[a_t\right] \tag{44}$$

with

$$\xi_0=\frac{\alpha_1}{2}\log\left(1+\sigma_\beta^2\sigma_\varepsilon^2\right)+\alpha_1\left(\frac{\nu}{1-\alpha_2\nu}\right)^2\frac{\sigma_\varepsilon^2}{2},\qquad \xi_\beta=-\alpha_1\frac{\left(\frac{\nu}{1-\alpha_2\nu}\right)^2\frac{\sigma_\varepsilon^2}{2}}{1+\sigma_\beta^2\sigma_\varepsilon^2}$$

$$\xi_{\varepsilon^2}=-\frac{1-\alpha_2\nu}{\nu}\frac{\sigma_\beta^2}{2},\qquad \xi_{a_{-1}^2}=-\frac{1-\left(\alpha_2+\alpha_2\right)\nu}{\nu}\left(\frac{1+\sigma_\beta^2\sigma_\varepsilon^2}{1-\alpha+\sigma_\beta^2\sigma_\varepsilon^2}\right)^2\frac{\sigma_\beta^2}{2}$$

$$\xi_{a_{-1},\varepsilon}=-\frac{1-\alpha_2\nu}{\nu}\frac{1+\sigma_\beta^2\sigma_\varepsilon^2}{1-\alpha+\sigma_\beta^2\sigma_\varepsilon^2}\sigma_\beta^2,\qquad \xi_{a_{-1}}=\frac{\alpha\sigma_\beta^2\sigma_\varepsilon^2}{1-\alpha+\sigma_\beta^2\sigma_\varepsilon^2}$$

We assume the following forms for the capital wedge and the output gap:

$$\tau_{\Lambda t} = -\tau_{\Lambda a}\varepsilon_t + \mathbb{E}_{t-1}\left[\tau_{\Lambda t}\right] \tag{45}$$
$$\mu_t = \mu_a\varepsilon_t + \mathbb{E}_{t-1}\left[\mu_t\right] \tag{46}$$

which are more general versions of (12) and (13). These generalizations ensure that the result holds across all classes of monetary policy studied in the paper (e.g., the simple rules in (13) and (30) and the full commitment optimal policy in Section 4).

We conjecture (and later verify) the following log-linear form for TFP:

$$\psi_t = \mathbb{E}_{t-1}\left[\psi_t\right] + \psi_a\varepsilon_t \tag{47}$$

Substituting the form of the profit function in (5) into (6) and rearranging yields the following expression for the capital choice

$$K_{it} = \left(\alpha G\left(R^K_{t-1}\right)^{-1}\mathbb{E}_{t-1}\left[\Lambda_t T_{\Lambda t} A_{it}^{\frac{\nu}{1-\alpha_2\nu}} Y_t^{\frac{1-\nu}{1-\alpha_2\nu}} W_t^{-\frac{\alpha_2\nu}{1-\alpha_2\nu}}\right]\right)^{\frac{1}{1-\alpha}}. \tag{48}$$

and relative capital stocks:

$$\frac{K_{it}}{K_t} = \frac{\left(\mathbb{E}_{t-1}\left[\Lambda_t T_{\Lambda t} A_{it}^{\frac{\nu}{1-\alpha_2\nu}} Y_t^{\frac{1-\nu}{1-\alpha_2\nu}} W_t^{-\frac{\alpha_2\nu}{1-\alpha_2\nu}}\right]\right)^{\frac{1}{1-\alpha}}}{\int\left(\mathbb{E}_{t-1}\left[\Lambda_t T_{\Lambda t} A_{it}^{\frac{\nu}{1-\alpha_2\nu}} Y_t^{\frac{1-\nu}{1-\alpha_2\nu}} W_t^{-\frac{\alpha_2\nu}{1-\alpha_2\nu}}\right]\right)^{\frac{1}{1-\alpha}} di} \tag{49}$$

Next, combine (39) with the labor supply condition (2) and resource constraint to obtain

$$W_t = (\alpha_2\nu)^{1-\frac{1}{1+\varphi}}\chi^{\frac{1}{1+\varphi}} Y_t^{1-\frac{1-\gamma}{1+\varphi}}\mathcal{M}_t^{-\frac{1}{1+\varphi}} \tag{50}$$

Substituting for Y_t and W_t using (41) and (50) and rearranging, we can write the expectation in (48) as

$$\mathbb{E}_{t-1}\left[\Lambda_t T_{\Lambda t} A_{it}^{\frac{\nu}{1-\alpha_2\nu}} Y_t^{\frac{1-\nu}{1-\alpha_2\nu}} W_t^{-\frac{\alpha_2\nu}{1-\alpha_2\nu}}\right] = \text{Const.} \times \mathbb{E}_{t-1}\left[T_{\Lambda t} A_{it}^{\frac{\nu}{1-\alpha_2\nu}} \mathcal{M}_t^{-\kappa_l\phi_l}\Psi_t^{-\kappa_\psi}\right] \tag{51}$$

where Const. is a homogeneous term that picks up terms that may vary through time but are common across firms and are known at time $t-1$ and

$$\begin{aligned}\kappa_\psi &= -\left(\frac{1-\nu}{1-\alpha_2\nu} - \gamma - \frac{\alpha_2\nu}{1-\alpha_2\nu}\left(1-\frac{1-\gamma}{1+\varphi}\right)\right)\phi_\psi \\ &= (\gamma-1)\,\phi_\psi + \frac{\nu}{1-\alpha_2\nu} \qquad (52)\\ \kappa_l\phi_l &= -\frac{1}{1+\varphi}\frac{\alpha_2\nu}{1-\alpha_2\nu} - \left(\frac{1-\nu}{1-\alpha_2\nu} - \gamma - \frac{\alpha_2\nu}{1-\alpha_2\nu}\left(1-\frac{1-\gamma}{1+\varphi}\right)\right)\phi_l \\ &= (\gamma-1)\,\phi_l \\ \Rightarrow\ \kappa_l &= \gamma - 1 \qquad (53)\end{aligned}$$

We provide further intuition for these terms in Appendix B.4.

Using (45), (46), (47) and the definition of μ_t in (9), we can substitute to write

$$\mathbb{E}_{t-1}\left[T_{\Lambda t}A_{it}^{\frac{\nu}{1-\alpha_2\nu}}\mathcal{M}_t^{-\kappa_l\phi_l}\Psi_t^{-\kappa_\psi}\right] = \text{Const.} \times \mathbb{E}_{t-1}\left[A_{it}^{\frac{\nu}{1-\alpha_2\nu}}e^{-\kappa\varepsilon_t}\right] \tag{54}$$

where κ is as defined in equation (21) and again Const. picks up terms that may vary through time but are common across firms and known at time $t-1$. Substituting into (49),

$$\frac{K_{it}}{K_t} = \frac{\left(\mathbb{E}_{t-1}\left[A_{it}^{\frac{\nu}{1-\alpha_2\nu}}e^{-\kappa\varepsilon_t}\right]\right)^{\frac{1}{1-\alpha}}}{\int\left(\mathbb{E}_{t-1}\left[A_{it}^{\frac{\nu}{1-\alpha_2\nu}}e^{-\kappa\varepsilon_t}\right]\right)^{\frac{1}{1-\alpha}}di} \tag{55}$$

and into the expression for TFP in (10), we obtain:

$$\Psi_t = \frac{\left(\int\left(A_{it}\left(\mathbb{E}_{t-1}\left[A_{it}^{\frac{\nu}{1-\alpha_2\nu}}e^{-\kappa\varepsilon_t}\right]\right)^{\alpha_1\frac{1-\alpha_2\nu}{1-(\alpha_1+\alpha_2)\nu}}\right)^{\frac{\nu}{1-\alpha_2\nu}}di\right)^{\frac{1-\alpha_2\nu}{\nu}}}{\left(\int\left(\mathbb{E}_{t-1}\left[A_{it}^{\frac{\nu}{1-\alpha_2\nu}}e^{-\kappa\varepsilon_t}\right]\right)^{\frac{1-\alpha_2\nu}{1-(\alpha_1+\alpha_2)\nu}}di\right)^{\alpha_1}} \tag{56}$$

Substituting for a_{it} from (44), we can write

$$\begin{aligned}\mathbb{E}_{t-1}\left[A_{it}^{\frac{\nu}{1-\alpha_2\nu}}e^{-\kappa\varepsilon_t}\right] &= e^{\frac{\nu}{1-\alpha_2\nu}\left(\hat{\beta}_i\mathbb{E}_{t-1}[a_t]+\xi_0+\xi_\beta\beta_i^2+\xi_{a_{-1}^2}\mathbb{E}_{t-1}[a_t]^2+\xi_{a_{-1}}\mathbb{E}_{t-1}[a_t]\right)} \\ &\times \mathbb{E}_{t-1}\left[e^{\varepsilon_t\left(\frac{\nu}{1-\alpha_2\nu}\left(\hat{\beta}_i+\xi_{a_{-1},\varepsilon}\mathbb{E}_{t-1}[a_t]\right)-\kappa\right)+\frac{\nu}{1-\alpha_2\nu}\xi_{\varepsilon^2}\varepsilon_t^2}\right]\end{aligned} \tag{57}$$

To explicitly evaluate the integral, we make use of the fact that for a normal random variable $x \sim \mathcal{N}\left(\bar{x}, \sigma_x^2\right)$ where $1-2b\sigma_x^2 > 0$, the properties of Gaussian integrals imply

$$\mathbb{E}\left[e^{ax+bx^2}\right] = \frac{1}{\sqrt{1-2b\sigma_x^2}}e^{\frac{a\bar{x}+\frac{a^2}{2}\sigma_x^2+b\bar{x}^2}{1-2b\sigma_x^2}} \tag{58}$$

Therefore, the expectational term in (57) can be written as:

$$\mathbb{E}_{t-1}\left[e^{\varepsilon_t\left(\frac{\nu}{1-\alpha_2\nu}\left(\hat{\beta}_i+\xi_{a_{-1},\varepsilon}\mathbb{E}_{t-1}[a_t]\right)-\kappa\right)+\frac{\nu}{1-\alpha_2\nu}\xi_{\varepsilon^2}\varepsilon_t^2}\right] = \frac{e^{\frac{1}{2}\frac{\left(\frac{\nu}{1-\alpha_2\nu}\left(\hat{\beta}_i+\xi_{a_{-1},\varepsilon}\mathbb{E}_{t-1}[a_t]\right)-\kappa\right)^2\sigma_\varepsilon^2}{1-2\sigma_\varepsilon^2\frac{\nu}{1-\alpha_2\nu}\xi_{\varepsilon^2}}}}{\sqrt{1-2\sigma_\varepsilon^2\frac{\nu}{1-\alpha_2\nu}\xi_{\varepsilon^2}}} \tag{59}$$

and substituting, together with (57), into (56) and simplifying yields:

$$\Psi_t = \frac{\left(\int \left(A_{it}\left(e^{\frac{\nu}{1-\alpha_2\nu}\left(\hat{\beta}_i\mathbb{E}_{t-1}[a_t]+\xi_\beta\hat{\beta}_i^2\right)+\frac{1}{2}\frac{\left(\frac{\nu}{1-\alpha_2\nu}\left(\hat{\beta}_i+\xi_{a_{-1},\varepsilon}\mathbb{E}_{t-1}[a_t]\right)-\kappa\right)^2\sigma_\varepsilon^2}{1-2\sigma_\varepsilon^2\frac{\nu}{1-\alpha_2\nu}\xi_{\varepsilon^2}}}\right)^{\frac{\alpha_1(1-\alpha_2\nu)}{1-(\alpha_1+\alpha_2)\nu}}\right)^{\frac{\nu}{1-\alpha_2\nu}} di\right)^{\frac{1-\alpha_2\nu}{\nu}}}{\left(\int\left(e^{\frac{\nu}{1-\alpha_2\nu}\left(\hat{\beta}_i\mathbb{E}_{t-1}[a_t]+\xi_\beta\hat{\beta}_i^2\right)+\frac{1}{2}\frac{\left(\frac{\nu}{1-\alpha_2\nu}\left(\hat{\beta}_i+\xi_{a_{-1},\varepsilon}\mathbb{E}_{t-1}[a_t]\right)-\kappa\right)^2\sigma_\varepsilon^2}{1-2\sigma_\varepsilon^2\frac{\nu}{1-\alpha_2\nu}\xi_{\varepsilon^2}}}\right)^{\frac{1-\alpha_2\nu}{1-(\alpha_1+\alpha_2)\nu}} di\right)^{\alpha_1}}$$

Substituting for A_{it} and the definitions of ξ_β, ξ_{ε^2} and $\xi_{a_{-1},\varepsilon}$ and rearranging yields

$$\Psi_t = A_t^h \frac{\left(\int\left(e^{\hat{\beta}_i\frac{\nu}{1-\alpha_2\nu}\left(\varepsilon_t+\mathbb{E}_{t-1}[a_t]\left(\frac{1-\alpha_2\nu}{1-(\alpha_1+\alpha_2)\nu}\right)-\frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\kappa\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}-\frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\mathbb{E}_{t-1}[a_t]\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\right)}\right) di\right)^{\frac{1-\alpha_2\nu}{\nu}}}{\left(\int\left(e^{\hat{\beta}_i\frac{\nu}{1-(\alpha_1+\alpha_2)\nu}\left(\mathbb{E}_{t-1}[a_t]-\frac{\kappa\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}-\frac{\mathbb{E}_{t-1}[a_t]\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\right)+\frac{\frac{1}{2}\left(\frac{\nu}{1-\alpha_2\nu}\right)^2\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}\hat{\beta}_i^2}\right) di\right)^{\alpha_1}}$$

where $A_t^h = e^{\xi_0+\xi_{\varepsilon^2}\varepsilon_t^2+\xi_{a_{-1}^2}\mathbb{E}_{t-1}[a_t]^2+\xi_{a_{-1}\varepsilon}\mathbb{E}_{t-1}[a_t]\varepsilon_t+\xi_{a_{-1}}\mathbb{E}_{t-1}[a_t]}$ consists of terms that are common across firms.

Thus, we can write

$$\log\Psi_t = \log A_t^h + \mathcal{I}_n - \mathcal{I}_d \tag{60}$$

where

$$\mathcal{I}_n = \frac{1-\alpha_2\nu}{\nu}\log\left(\int e^{\hat{\beta}_i\frac{\nu}{1-\alpha_2\nu}\left(\varepsilon_t+\mathbb{E}_{t-1}[a_t]\left(\frac{1-\alpha_2\nu}{1-(\alpha_1+\alpha_2)\nu}\right)-\frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\kappa\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}-\frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\mathbb{E}_{t-1}[a_t]\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\right)} di\right)$$

$$\mathcal{I}_d = \alpha_1\log\left(\int e^{\hat{\beta}_i\frac{\nu}{1-(\alpha_1+\alpha_2)\nu}\left(\mathbb{E}_{t-1}[a_t]-\frac{\kappa\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}-\frac{\mathbb{E}_{t-1}[a_t]\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\right)+\frac{\frac{1}{2}\left(\frac{\nu}{1-\alpha_2\nu}\right)^2\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}\hat{\beta}_i^2} di\right)$$

Both terms are Gaussian integrals over $\hat{\beta}_i$ which is normally distributed, and therefore we can apply result (58) to obtain

$$\begin{aligned}\mathcal{I}_n &= \varepsilon_t + \mathbb{E}_{t-1}[a_t]\frac{1-\alpha_2\nu}{1-(\alpha_1+\alpha_2)\nu} - \frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\kappa\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2} - \frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\mathbb{E}_{t-1}[a_t]\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2} \\ &+ \left(\frac{\nu}{1-\alpha_2\nu}\right)\left(\varepsilon_t + \mathbb{E}_{t-1}[a_t]\frac{1-\alpha_2\nu}{1-(\alpha_1+\alpha_2)\nu} - \frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\kappa\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2} - \frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\mathbb{E}_{t-1}[a_t]\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\right)^2\frac{\sigma_{\hat{\beta}}^2}{2}\end{aligned}$$

$$\mathcal{I}_d = \frac{\alpha_1}{2}\log\left(1+\sigma_\varepsilon^2\sigma_\beta^2\right) + \alpha_1 \frac{\frac{\nu}{1-(\alpha_1+\alpha_2)\nu}\left(\mathbb{E}_{t-1}\left[a_t\right] - \frac{\kappa\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2} - \frac{\mathbb{E}_{t-1}[a_t]\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\right)}{\frac{1}{1+\sigma_\varepsilon^2\sigma_\beta^2}}$$
$$+ \alpha_1 \frac{\left(\frac{\nu}{1-(\alpha_1+\alpha_2)\nu}\left(\mathbb{E}_{t-1}\left[a_t\right] - \frac{\kappa\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2} - \frac{\mathbb{E}_{t-1}[a_t]\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\right)\right)^2 \frac{\sigma_{\hat{\beta}}^2}{2} + \frac{\frac{1}{2}\left(\frac{\nu}{1-\alpha_2\nu}\right)^2\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}}{\frac{1}{1+\sigma_\varepsilon^2\sigma_\beta^2}}$$

Substituting into (60) yields an expression for TFP of the form

$$\psi_t = \overline{\psi} + \psi_a\varepsilon_t + \psi_{a_{-1}}\mathbb{E}_{t-1}\left[a_t\right] + \psi_{a_{-1},\varepsilon}\mathbb{E}_{t-1}\left[a_t\right]\varepsilon_t + \psi_{\varepsilon^2}\varepsilon_t^2 + \psi_{a_{-1}^2}\mathbb{E}_{t-1}\left[a_t\right]^2$$

When $\psi_{a_{-1},\varepsilon} = \psi_{\varepsilon^2} = \psi_{a_{-1}^2} = 0$ and $\psi_{a_{-1}} = 1$, we have

$$\psi_t = \overline{\psi} + \psi_a\varepsilon_t + \mathbb{E}_{t-1}\left[a_t\right] \tag{61}$$

where

$$\psi_a = 1 - \frac{\alpha_1\left(1-\alpha_2\nu\right)}{1-\left(\alpha_1+\alpha_2\right)\nu}\frac{\kappa\sigma_\varepsilon^2\sigma_\beta^2}{1+\sigma_\varepsilon^2\sigma_\beta^2} \tag{62}$$

$$\overline{\psi} = \kappa\sigma_\varepsilon^2\frac{\alpha_1\nu}{1-\left(\alpha_1+\alpha_2\right)\nu}\frac{\sigma_\varepsilon^2\sigma_\beta^2}{1+\sigma_\varepsilon^2\sigma_\beta^2} - \frac{\sigma_\beta^2}{2}\kappa^2\left(\frac{1}{1-\alpha}\frac{\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}\right)^2\alpha_1\left(1-\alpha+\sigma_\varepsilon^2\sigma_\beta^2\right) \tag{63}$$

which verifies our conjecture in (47) and yields (14) in the i.i.d. case and (28) when a_t follows an AR(1) process with persistence δ. From here, we can substitute for κ from (21) to solve for

$$\psi_a = \frac{1-\left(\tau_{\Lambda a}+\kappa_l\mu_a\right)\omega\Upsilon}{1+\kappa_\psi\omega\Upsilon} \tag{64}$$

and

$$\kappa = \frac{\kappa_\psi+\tau_{\Lambda a}+\kappa_l\mu_a}{1+\kappa_\psi\omega\Upsilon} \tag{65}$$

where we have used the definition of $\omega = \frac{\alpha_1(1-\alpha_2\nu)}{1-(\alpha_1+\alpha_2)\nu}$ and to ease notation, we define $\Upsilon \equiv \frac{\sigma_\varepsilon^2\sigma_\beta^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}$.

The remainder of the proof consists in showing that $\psi_{a_{-1},\varepsilon} = \psi_{\varepsilon^2} = \psi_{a_{-1}^2} = 0$ and $\psi_{a_{-1}} = 1$ so that (61) holds, and derives (62) and (63).

Proof that $\psi_{a_{-1},\varepsilon} = 0$: Combining the terms in (60),

$$\psi_{a_{-1},\varepsilon} = \xi_{\varepsilon,a_{-1}} + \frac{\nu}{1-\alpha_2\nu}\left(\frac{1-\alpha_2\nu}{1-\left(\alpha_1+\alpha_2\right)\nu} - \frac{\alpha_1\nu}{1-\left(\alpha_1+\alpha_2\right)\nu}\frac{\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\right)\sigma_{\hat{\beta}}^2$$

and substituting for the definition of $\xi_{a_{-1},\varepsilon}$:

$$\begin{aligned}
\psi_{a_{-1},\varepsilon} &= -\frac{\nu}{(1-\alpha_2\nu)}\frac{1+\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\sigma_{\hat{\beta}}^2 \\
&+ \frac{\nu}{1-\alpha_2\nu}\left(\frac{1-\alpha_2\nu}{1-(\alpha_1+\alpha_2)\nu}-\frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\right)\sigma_{\hat{\beta}}^2 \\
&= \frac{\nu}{1-\alpha_2\nu}\sigma_{\hat{\beta}}^2\left(-\frac{(1-\alpha_2\nu)}{1-(\alpha_1+\alpha_2)\nu}+\frac{1-\alpha_2\nu}{1-(\alpha_1+\alpha_2)\nu}\right) \\
&= 0
\end{aligned}$$

Proof that $\psi_{\varepsilon^2}=0$: Combining the terms in (60),

$$\psi_{\varepsilon^2}=\xi_{\varepsilon^2}+\left(\frac{\nu}{1-\alpha_2\nu}\right)\frac{\sigma_{\hat{\beta}}^2}{2}$$

and substituting for the definition of ξ_{ε^2}:

$$\psi_{\varepsilon^2}=-\frac{1-\alpha_2\nu}{\nu}\frac{\sigma_\beta^2}{2}+\frac{1-\alpha_2\nu}{\nu}\frac{\sigma_\beta^2}{2}=0$$

Proof that $\psi_{a_{-1}^2}=0$: Combining the terms in (60),

$$\begin{aligned}
\psi_{a_{-1}^2} &= \xi_{a_{-1}^2}+\frac{\nu}{1-\alpha_2\nu}\left(\frac{1-\alpha_2\nu}{1-(\alpha_1+\alpha_2)\nu}-\frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\right)^2\frac{\sigma_{\hat{\beta}}^2}{2} \\
&- \alpha_1\frac{\left(\frac{\nu}{1-(\alpha_1+\alpha_2)\nu}\left(1-\frac{\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\right)\right)^2\frac{\sigma_{\hat{\beta}}^2}{2}}{\frac{1}{1+\sigma_\varepsilon^2\sigma_\beta^2}}
\end{aligned}$$

and substituting for the definition of $\xi_{a_{-1}^2}$:

$$\begin{aligned}
\psi_{a_{-1}^2} &= -\frac{\sigma_\beta^2}{2}\frac{1-\nu(\alpha_1+\alpha_2)}{\nu}\left(\frac{1+\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\right)^2-\alpha_1\frac{\left(\frac{\nu}{1-(\alpha_1+\alpha_2)\nu}\left(1-\frac{\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\right)\right)^2\left(\frac{1-\alpha_2\nu}{\nu}\right)^2\frac{\sigma_\beta^2}{2}}{\frac{1}{1+\sigma_\varepsilon^2\sigma_\beta^2}} \\
&+ \frac{1-\alpha_2\nu}{\nu}\left(\frac{1-\alpha_2\nu}{1-(\alpha_1+\alpha_2)\nu}-\frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\right)^2\frac{\sigma_\beta^2}{2} \\
&= \frac{\sigma_\beta^2}{2}\left(\frac{1+\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\right)^2\left(-\frac{1-\nu(\alpha_1+\alpha_2)}{\nu}+\frac{1-\nu(\alpha_1+\alpha_2)}{\nu}\right) \\
&= 0
\end{aligned}$$

Proof that $\psi_{a_{-1}} = 1$: Combining the terms in (60),

$$\begin{aligned}
\psi_{a_{-1}} &= \xi_{a_{-1}} + \frac{1-\alpha_2\nu}{1-(\alpha_1+\alpha_2)\nu} - \frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2} \\
&- \frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\kappa\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}\frac{\nu}{1-\alpha_2\nu}\left(\frac{1-\alpha_2\nu}{1-(\alpha_1+\alpha_2)\nu} - \frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\right)\sigma_{\hat{\beta}}^2 \\
&- \alpha_1\frac{\frac{\nu}{1-(\alpha_1+\alpha_2)\nu}\left(1-\frac{\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\right)}{\frac{1}{1+\sigma_\varepsilon^2\sigma_\beta^2}} - \alpha_1\frac{-\frac{\kappa\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}\left(\frac{\nu}{1-(\alpha_1+\alpha_2)\nu}\right)^2\left(1-\frac{\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\right)\sigma_{\hat{\beta}}^2}{\frac{1}{1+\sigma_\varepsilon^2\sigma_\beta^2}}
\end{aligned}$$

and simplifying,

$$\psi_{a_{-1}} = \xi_{a_{-1}} + \frac{1+\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\frac{1-\nu(\alpha_1+\alpha_2)}{1-\alpha_2\nu}$$

and substituting for the definition of $\xi_{a_{-1}}$:

$$\begin{aligned}
\psi_{a_{-1}} &= \frac{\frac{1-\nu(\alpha_1+\alpha_2)}{1-\alpha_2\nu}+\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2} \\
&= 1
\end{aligned}$$

Derivation of ψ_a: Combining the terms in (60),

$$\begin{aligned}
\psi_a &= 1 - \frac{\nu}{1-\alpha_2\nu}\frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\kappa\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}\sigma_{\hat{\beta}}^2 \\
&= 1 - \frac{\alpha_1(1-\alpha_2\nu)}{1-(\alpha_1+\alpha_2)\nu}\frac{\kappa\sigma_\varepsilon^2\sigma_\beta^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}
\end{aligned}$$

Derivation of $\overline{\psi}$: Combining the terms in (60),

$$\begin{aligned}
\overline{\psi} &= \xi_0 - \frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\kappa\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2} + \frac{\nu}{1-\alpha_2\nu}\left(\frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\kappa\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}\right)^2\frac{\sigma_{\hat{\beta}}^2}{2} \\
&- \frac{\alpha_1}{2}\log\left(1+\sigma_\varepsilon^2\sigma_\beta^2\right) + \alpha_1\frac{\left(\frac{\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\kappa\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}\right)}{\frac{1}{1+\sigma_\varepsilon^2\sigma_\beta^2}} - \alpha_1\frac{\left(\frac{\nu}{1-(\alpha_1+\alpha_2)\nu}\left(\frac{\kappa\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}\right)\right)^2\frac{\sigma_{\hat{\beta}}^2}{2} + \frac{\frac{1}{2}\left(\frac{\nu}{1-\alpha_2\nu}\right)^2\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}}{\frac{1}{1+\sigma_\varepsilon^2\sigma_\beta^2}}
\end{aligned}$$

and substituting for the definition of ξ_0:

$$\begin{aligned}
\overline{\psi} &= \kappa\sigma_\varepsilon^2\frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\sigma_\varepsilon^2\sigma_\beta^2}{1+\sigma_\varepsilon^2\sigma_\beta^2} \\
&+ \frac{1}{2}\sigma_\varepsilon^2\kappa^2\frac{\sigma_\beta^2\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}\left(\frac{1-\alpha_2\nu}{\nu}\right)^2\left(\left(\frac{\nu}{1-\alpha_2\nu}\right)\left(\frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\right)^2\frac{1}{1+\sigma_\varepsilon^2\sigma_\beta^2} - \alpha_1\left(\frac{\nu}{1-(\alpha_1+\alpha_2)\nu}\right)^2\right) \\
&= \kappa\sigma_\varepsilon^2\frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\sigma_\varepsilon^2\sigma_\beta^2}{1+\sigma_\varepsilon^2\sigma_\beta^2} - \frac{\sigma_{\hat{\beta}}^2}{2}\kappa^2\left(\frac{1}{1-\alpha}\frac{\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}\right)^2\alpha_1\left(1-\alpha+\sigma_\varepsilon^2\sigma_\beta^2\right)
\end{aligned}$$

which completes the proof of (14).

The proof of Proposition 2 is completed with the linearized Phillips curve and Euler equation. Log-linearizing (42) yields expression (15):

$$\pi_t^w = \rho\mathbb{E}_t\left[\pi_{t+1}^w\right] + \lambda_w\mu_t$$

where $\lambda_w \equiv \frac{\alpha_2 \nu_w}{\theta_w \phi_l}$ and $\frac{1}{\phi_l}\mu_t = mrs_t - w_t = \varphi l_t + \gamma y_t - w_t$.

Log-linearizing (43) and combining with the resource constraint, $c_t = y_t$, yields (16):

$$y_t = \mathbb{E}_t\left[y_{t+1}\right] - \frac{1}{\gamma}\left(i_t - \mathbb{E}_t\left[\pi^p_{t+1}\right]\right)$$

□

B.2 Monetary Policy Implementation

Interest rate rule. Here we show that a policy of the form (13) can be implemented with a nominal interest rate rule that sets the nominal rate as a function of the exogenous shock. We stay within the i.i.d. case for simplicity, but a similar result holds with persistence.

From (16), the log-linearized Euler equation is

$$y_t = \mathbb{E}_t\left[y_{t+1}\right] - \frac{1}{\gamma}\left(i_t - \mathbb{E}_t\left[\pi^p_{t+1}\right]\right)$$

From (15) the log-linearized Phillips curve is

$$\pi^w_t = \rho \mathbb{E}_t\left[\pi^w_{t+1}\right] + \lambda_w \mu_t$$

and using this along with (13), we can guess and verify that

$$\pi^w_t = \zeta^w_a \varepsilon_t, \quad \zeta^w_a = \lambda_w \mu_a$$

From (50), in deviations,

$$w_t = \left(1 - \frac{1-\gamma}{1+\varphi}\right) y_t - \frac{1}{1+\varphi}\frac{1}{\phi_l}\mu_t$$

and from (9) and (14), in deviations,

$$y_t = y_a \varepsilon_t, \quad y_a = \phi_\psi \psi_a + \mu_a$$

Combining,

$$w_t = w_a \varepsilon_t, \quad w_a = \left(1 - \frac{1-\gamma}{1+\varphi}\right)\phi_\psi \psi_a - \frac{1-\alpha_2}{\alpha_2}\mu_a$$

By definition

$$\pi^p_{t+1} = w_t - w_{t+1} + \pi^w_{t+1} = w_a\left(\varepsilon_t - \varepsilon_{t+1}\right) + \zeta^w_a \varepsilon_{t+1}$$

so

$$\mathbb{E}_t\left[\pi^p_{t+1}\right] = w_a \varepsilon_t$$

and substituting into the Euler equation:

$$\begin{aligned} i_t &= \left(w_a - \gamma y_a\right)\varepsilon_t \\ &= \left(\frac{\varphi\left(1-\gamma\right)}{1+\varphi}\phi_\psi \psi_a - \frac{1-\alpha_2\left(1-\gamma\right)}{\alpha_2}\mu_a\right)\varepsilon_t \\ &= i_a \varepsilon_t \end{aligned}$$

Last, substitute for ψ_a from (64) to obtain

$$i_a = \frac{\varphi(1-\gamma)}{1+\varphi}\phi_\psi \frac{1-\tau_{\Lambda a}\omega\Upsilon}{1+\kappa_\psi\omega\Upsilon} - \left(\frac{\varphi(1-\gamma)}{1+\varphi}\phi_\psi \frac{\kappa_l\omega\Upsilon}{1+\kappa_\psi\omega\Upsilon} + \frac{1-\alpha_2(1-\gamma)}{\alpha_2}\right)\mu_a \quad (66)$$

$$= \Phi^i - \Phi^i_\mu \mu_a \quad (67)$$

which shows that any choice of μ_a in (13) can be implemented by a nominal interest rate rule, i_a, that satisfies (66).

To prove that $\Phi^i_\mu > 0$, we can use the fact that $\kappa_\psi = \phi_\psi\kappa_l + \frac{\nu}{1-\alpha_2\nu}$ to write

$$\Phi^i_\mu = (1-\gamma)\left(\frac{\varphi}{1+\varphi}\frac{\phi_\psi\kappa_l\omega\Upsilon}{1+\phi_\psi\kappa_l\omega\Upsilon + \frac{\nu}{1-\alpha_2\nu}\omega\Upsilon} - 1\right) + \frac{1}{\alpha_2}$$

Because both fractions in braces are less than one, the entire term in braces is negative and because $\gamma > 1$, the entire term is unambiguously positive. Thus, $\Phi^i_\mu > 0$ and a more procyclical output gap is associated with a less procyclical nominal interest rate.

Taylor Rule. Here we show that a policy of the form (13) can be implemented with standard formulations of a Taylor Rule. First, we consider a rule in the output gap and expected price inflation which is what we use in our quantitative exercise in Section 4. Next we show that a similar mapping holds when the rule is formulated in terms of the output gap and realized wage inflation.

Consider a rule of the form

$$i_t = \phi_y\mu_t + \phi_\pi \mathbb{E}_t\left[\pi^p_{t+1}\right]$$

From the Euler equation, expression (66) gives one expression for i_a, the responsiveness of the nominal rate to the exogenous shock. Substituting for μ_t and $\mathbb{E}_t\left[\pi^p_{t+1}\right]$ in the Taylor rule, we can derive

$$i_t = \left(\phi_\pi\left(1-\frac{1-\gamma}{1+\varphi}\right)\phi_\psi\frac{1-\tau_{\Lambda a}\omega\Upsilon}{1+\kappa_\psi\omega\Upsilon} + \left(\phi_y - \phi_\pi\left(\left(1-\frac{1-\gamma}{1+\varphi}\right)\phi_\psi\frac{\kappa_l\omega\Upsilon}{1+\kappa_\psi\omega\Upsilon} + \frac{1-\alpha_2}{\alpha_2}\right)\right)\mu_a\right)\varepsilon_t$$

$$= i^{TR}_a\varepsilon_t$$

which is a second representation of the responsiveness of the nominal rate. Equating coefficients yields

$$\mu_a = \frac{\frac{1-\tau_{\Lambda a}\omega\Upsilon}{1+\kappa_\psi\omega\Upsilon}\phi_\psi\left(\frac{\varphi(1-\gamma)}{1+\varphi} - \phi_\pi\left(1-\frac{1-\gamma}{1+\varphi}\right)\right)}{\frac{\varphi(1-\gamma)}{1+\varphi}\phi_\psi\frac{\kappa_l\omega\Upsilon}{1+\kappa_\psi\omega\Upsilon} + \frac{1-\alpha_2(1-\gamma)}{\alpha_2} + \phi_y - \phi_\pi\left(\left(1-\frac{1-\gamma}{1+\varphi}\right)\phi_\psi\frac{\kappa_l\omega\Upsilon}{1+\kappa_\psi\omega\Upsilon} + \frac{1-\alpha_2}{\alpha_2}\right)}$$

which shows that any μ_a can be implemented via an appropriate choice of coefficients in the Taylor rule, ϕ_y and ϕ_π.

Taylor rule in wage inflation. As a second example, consider a rule that reacts to the output gap and wage inflation, i.e.,

$$i_t = \phi_y\mu_t + \phi_\pi\pi^w_t$$

Following similar steps as above, we can derive

$$\mu_a = \frac{\frac{\varphi(1-\gamma)}{1+\varphi}\phi_\psi\frac{1-\tau_{\Lambda a}\omega\Upsilon}{1+\kappa_\psi\omega\Upsilon}}{\frac{\varphi(1-\gamma)}{1+\varphi}\phi_\psi\frac{\kappa_l\omega\Upsilon}{1+\kappa_\psi\omega\Upsilon} + \frac{1-\alpha_2(1-\gamma)}{\alpha_2} + \phi_y + \phi_\pi\lambda_w}$$

B.3 Approximate Solution

Here we derive the approximate expressions in Section 2.1.

To derive (17), we use (10):

$$\Psi_t = \left(\int A_{it}^{\frac{\nu}{1-\alpha_2\nu}} \left(\frac{K_{it}}{K_t} \right)^{\alpha} di \right)^{\frac{1-\alpha_2\nu}{\nu}}$$

For purposes of the approximation, assume $\mathcal{O}_{it} = 0$ in (11) so that $a_{it} = \hat{\beta}_i a_t$. Substituting and approximating the TFP expression around the ergodic mean where $K_{it} = K_i$ and $a_t = 0$ yields

$$\Psi_t = \left(\int \left(\frac{K_i}{\overline{K}} \right)^{\alpha} \left(1 + \frac{\nu}{1-\alpha_2\nu} \hat{\beta}_i a_t \right) di \right)^{\frac{1-\alpha_2\nu}{\nu}}$$

Taking logs,

$$\psi_t = \frac{1-\alpha_2\nu}{\nu} \log \left(\int \left(\frac{K_i}{\overline{K}} \right)^{\alpha} \left(1 + \frac{\nu}{1-\alpha_2\nu} \hat{\beta}_i a_t \right) di \right)$$

and approximating using $\log x \approx x - 1$,

$$\begin{aligned}
\psi_t &= \frac{1-\alpha_2\nu}{\nu} \left(\int \left(\frac{K_i}{\overline{K}} \right)^{\alpha} \left(1 + \frac{\nu}{1-\alpha_2\nu} \hat{\beta}_i a_t \right) di - 1 \right) \\
&= \frac{1-\alpha_2\nu}{\nu} \left(\int \left(\frac{K_{it}}{\overline{K}} \right)^{\alpha} di - 1 \right) + \int \left(\frac{K_i}{\overline{K}} \right)^{\alpha} \hat{\beta}_i di a_t \\
&= \frac{1-\alpha_2\nu}{\nu} \log \left(\int \left(\frac{K_{it}}{\overline{K}} \right)^{\alpha} di \right) + \int \left(\frac{K_i}{\overline{K}} \right)^{\alpha} \hat{\beta}_i di a_t \\
&= \overline{\psi} + \psi_a a_t
\end{aligned} \tag{68}$$

Turning to the first term, we have

$$\overline{\psi} = \frac{1-\alpha_2\nu}{\nu} \log \left(\int \left(\frac{K_{it}}{\overline{K}} \right)^{\alpha} di \right)$$

At the ergodic mean, $MRPK_i = \text{Const.} \times \alpha K_i^{\alpha-1}$ where Const. is constant across firms, and rearranging,

$$K_i = \left(\frac{1}{\alpha} \frac{MRPK_i}{\text{Const.}} \right)^{\frac{1}{\alpha-1}}$$

Capital market clearing gives

$$\overline{K} = \left(\frac{1}{\alpha} \frac{1}{\text{Const.}} \right)^{\frac{1}{\alpha-1}} \int MRPK_i^{\frac{1}{\alpha-1}} di$$

and combining,

$$\frac{K_i}{\overline{K}} = \frac{MRPK_i^{\frac{1}{\alpha-1}}}{\int MRPK_i^{\frac{1}{\alpha-1}} di}$$

so that

$$\begin{aligned}\overline{\psi} &= \frac{1-\alpha_2\nu}{\nu}\log\left(\frac{\int MRPK_i^{\frac{\alpha}{\alpha-1}}di}{\left(\int MRPK_i^{\frac{1}{\alpha-1}}di\right)^{\alpha}}\right) \\ &= \frac{1-\alpha_2\nu}{\nu}\log\left(\int MRPK_i^{-\frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}}di\right)-\alpha_1\log\left(\int MRPK_i^{-\frac{1-\alpha_2\nu}{1-(\alpha_1+\alpha_2)\nu}}di\right)\end{aligned}$$

and taking a second-order approximation to the integral terms:

$$\overline{\psi} = -\frac{1}{2}\frac{\alpha_1\left(1-\alpha_2\nu\right)}{1-\left(\alpha_1+\alpha_2\right)\nu}\text{var}\left(mrpk_i\right)$$

Turning to the second term in (68), we have

$$\psi_a = \int\left(\frac{K_i}{\overline{K}}\right)^{\alpha}\hat{\beta}_i di$$

and rewriting and approximating,

$$\begin{aligned}\psi_a &= \int e^{\alpha\left(k_i-\overline{k}\right)}\hat{\beta}_i di \\ &= \int\left(1+\alpha\left(k_i-\overline{k}\right)\right)\hat{\beta}_i di\end{aligned}$$

Notice that

$$\begin{aligned}\int\left(k_i-\bar{k}\right)\hat{\beta}_i di &= \mathbb{E}\left[\left(k_i-\bar{k}\right)\hat{\beta}_i\right]=\mathbb{E}\left[k_i-\bar{k}\right]\mathbb{E}\left[\hat{\beta}_i\right]+\text{cov}\left(k_i-\bar{k},\hat{\beta}_i\right) \\ &= \text{cov}\left(k_i,\hat{\beta}_i\right)\end{aligned}$$

since $\mathbb{E}_t\left[k_i-\bar{k}\right]\approx 0$. Thus,

$$\psi_a = 1+\alpha\text{cov}\left(k_i,\hat{\beta}_i\right)=1+\alpha_1\text{cov}\left(k_i,\beta_i\right)$$

To derive expression (18), write (6) as

$$\mathbb{E}_{t-1}\left[e^{\tilde{\lambda}_t+mrpk_{it}}\right]=R_{t-1}^K$$

Taking a second-order approximation yields

$$\mathbb{E}_{t-1}\left[mrpk_{it}\right]+\frac{1}{2}\text{var}_{t-1}\left(mrpk_{it}\right)+\text{cov}_{t-1}\left(mrpk_{it},\tilde{\lambda}_t\right)+\text{Const.}=\log R_{t-1}^K$$

where the constant picks up terms that are common across firms. Suppressing constants gives (18).

Proof of Proposition 3. To derive (19), use expression (55) to write

$$\begin{aligned} K_{it} &= \text{Const.} \times \left(\mathbb{E}_{t-1}\left[e^{\frac{\nu}{1-\alpha_2\nu}a_{it}-\kappa\varepsilon_t}\right]\right)^{\frac{1}{1-\alpha}} \\ &= \text{Const.} \times \left(e^{\frac{\nu}{1-\alpha_2\nu}\mathbb{E}_{t-1}[a_{it}]+\frac{1}{2}\left(\frac{\nu}{1-\alpha_2\nu}\right)^2\text{var}_{t-1}(a_{it})+\frac{1}{2}\kappa^2\sigma_\varepsilon^2-\frac{\nu}{1-\alpha_2\nu}\kappa\text{cov}_{t-1}(a_{it},\varepsilon_t)}\right)^{\frac{1}{1-\alpha}} \end{aligned}$$

where we have used $a_{it} = \hat{\beta}_i a_t$ and the properties of the log-normal distribution. Taking logs, rearranging and suppressing constants,

$$k_{it} = \frac{1}{1-\alpha}\left(\log \mathbb{E}_{t-1}\left[A_{it}^{\frac{\nu}{1-\alpha_2\nu}}\right] - \beta_i\kappa\sigma_\varepsilon^2\right)$$

From (5) and (6),

$$mrpk_{it} = \text{Const.} + \frac{\nu}{1-\alpha_2\nu}a_{it} + (\alpha-1)\,k_{it}$$

and substituting for k_{it}, rearranging and suppressing constants,

$$mrpk_{it} = \log A_{it}^{\frac{\nu}{1-\alpha_2\nu}} - \log \mathbb{E}_{t-1}\left[A_{it}^{\frac{\nu}{1-\alpha_2\nu}}\right] + \beta_i\kappa\sigma_\varepsilon^2$$

To derive (20), approximate $\mathbb{E}_{t-1}\left[e^{\frac{\nu}{1-\alpha_2\nu}a_{it}}\right] \approx \mathbb{E}_{t-1}\left[1+\frac{\nu}{1-\alpha_2\nu}a_{it}\right]$ so that at the ergodic mean where $\mathbb{E}_t[a_{it}] = a_{it} = 0$,

$$\begin{aligned} k_i &\approx -\frac{1}{1-\alpha}\beta_i\kappa\sigma_\varepsilon^2 \quad \Rightarrow \quad \text{cov}(k_i,\beta_i) = -\frac{1}{1-\alpha}\kappa\sigma_\varepsilon^2\sigma_\beta^2 \\ mrpk_i &\approx \beta_i\kappa\sigma_\varepsilon^2 \quad \Rightarrow \quad \text{var}(mrpk_i) = \left(\kappa\sigma_\varepsilon^2\right)^2\sigma_\beta^2 \end{aligned}$$

and substituting into (17) yields (20).

The derivation of (21) is the same as in expression (54), which shows that that expression is exact. □

B.4 Interpretation of κ

To gain intuition for κ, consider again the optimality condition for capital in (48):

$$K_{it} = \left(\alpha G\left(R_{t-1}^K\right)^{-1}\mathbb{E}_{t-1}\left[\tilde{\Lambda}_t A_{it}^{\frac{\nu}{1-\alpha_2\nu}}Y_t^{\frac{1-\nu}{1-\alpha_2\nu}}W_t^{-\frac{\alpha_2\nu}{1-\alpha_2\nu}}\right]\right)^{\frac{1}{1-\alpha}}$$

where $\tilde{\Lambda}_t = \Lambda_t T_{\Lambda t}$. Taking logs and suppressing constants,

$$k_{it} = \frac{1}{1-\alpha}\log\left(\mathbb{E}_{t-1}\left[\tilde{\Lambda}_t A_{it}^{\frac{\nu}{1-\alpha_2\nu}}Y_t^{\frac{1-\nu}{1-\alpha_2\nu}}W_t^{-\frac{\alpha_2\nu}{1-\alpha_2\nu}}\right]\right)$$

and taking a second-order approximation,

$$\begin{aligned} k_{it} &= \frac{1}{1-\alpha}\left(\log \mathbb{E}_{t-1}\left[A_{it}^{\frac{\nu}{1-\alpha_2\nu}}\right] + \frac{\nu}{1-\alpha_2\nu}\text{cov}_{t-1}\left(a_{it}, \tilde{\lambda}_t + \frac{1-\nu}{1-\alpha_2\nu}y_t - \frac{\alpha_2\nu}{1-\alpha_2\nu}w_t\right)\right) \\ &= \frac{1}{1-\alpha}\left(\log \mathbb{E}_{t-1}\left[A_{it}^{\frac{\nu}{1-\alpha_2\nu}}\right] + \beta_i\frac{\partial}{\partial\varepsilon_t}\left(\tilde{\lambda}_t + \frac{1-\nu}{1-\alpha_2\nu}y_t - \frac{\alpha_2\nu}{1-\alpha_2\nu}w_t\right)\sigma_\varepsilon^2\right) \end{aligned}$$

and comparing to (19), we can see

$$\kappa = -\frac{\partial}{\partial\varepsilon_t}\left(\tilde{\lambda}_t + \frac{1-\nu}{1-\alpha_2\nu}y_t - \frac{\alpha_2\nu}{1-\alpha_2\nu}w_t\right)$$

i.e., κ measures the negative of the elasticity of the discounted profitability of capital to unexpected shocks operating through the (distorted) discount factor, wage and aggregate demand.

Expressions (51) and (54) show that using the equilibrium conditions to substitute, we obtain

$$\kappa = \kappa_\psi\psi_a + \tau_{\Lambda a} + \kappa_l\mu_a$$

where

$$\begin{aligned} \kappa_\psi &= -\left(\frac{1-\nu}{1-\alpha_2\nu} - \gamma - \frac{\alpha_2\nu}{1-\alpha_2\nu}\left(1 - \frac{1-\gamma}{1+\varphi}\right)\right)\phi_\psi \\ \kappa_l &= -\frac{1}{1+\varphi}\frac{\alpha_2\nu}{1-\alpha_2\nu}\frac{1}{\phi_l} - \left(\frac{1-\nu}{1-\alpha_2\nu} - \gamma - \frac{\alpha_2\nu}{1-\alpha_2\nu}\left(1 - \frac{1-\gamma}{1+\varphi}\right)\right) \end{aligned}$$

First, κ_ψ captures the effects of ε_t through changes in endogenous aggregate TFP, ψ_t. The term in parentheses measures the (negative) elasticity of discounted profitability to movements in aggregate output, Y_t. These come through movements in aggregate demand and so firm-level prices (the term $\frac{1-\nu}{1-\alpha_2\nu}$), in the (undistorted) SDF (γ) and in the wage ($-\frac{\alpha_2\nu}{1-\alpha_2\nu}\left(1-\frac{1-\gamma}{1+\varphi}\right)$). Multiplying by ϕ_ψ translates this elasticity with respect to output into the elasticity with respect to TFP, ψ_t, and multiplying through by ψ_a translates the entire term into the elasticity with respect to the exogenous shock, ε_t.

Second, movements in ε_t affect the capital wedge with elasticity $\tau_{\Lambda a}$.

Third, κ_l capture the effects of ε_t through the output gap. The first term in the expression captures the direct effect of μ_t on profitability through changes in the wage, which is the product of the elasticity of profits with respect to the wage, $-\frac{\alpha_2\nu}{1-\alpha_2\nu}$ and the elasticity of the wage with respect to the output gap, $-\frac{1}{1+\varphi}\frac{1}{\phi_l}$. The second term captures the effect of the wedge on profitability through changes in output. Multiplying through by μ_a translates this term into an elasticity with respect to ε_t.

As shown in (52) and (53), these expressions reduce to

$$\begin{aligned} \kappa_\psi &= (\gamma-1)\,\phi_\psi + \frac{\nu}{1-\alpha_2\nu} \\ \kappa_l &= \gamma - 1 \end{aligned}$$

C Welfare and Optimal Policy

Welfare criterion. The period utility function inclusive of the costs of adjusting wages is equal to

$$U_t = \frac{C_t^{1-\gamma}}{1-\gamma} - \chi\frac{L^{1+\varphi}}{1+\varphi} - \frac{\theta_w}{2}\left(\pi_t^w\right)^2 Y_t^{1-\gamma}$$

and a second-order approximation around the ergodic mean (denoted subscript m) yields

$$U_t - U_m = C_m^{1-\gamma}\left(c_t + \frac{1-\gamma}{2}c_t^2 - \frac{\theta_w}{2}\left(\pi_t^w\right)^2\right) - \chi L_m^{1+\varphi}\left(l_t + \frac{1+\varphi}{2}l_t^2\right)$$

Under our assumption that the optimal time-invariant production subsidy, $\frac{1}{\nu}$, is in place, we have $\chi L_m^{1+\varphi} = \alpha_2 C_m^{1-\gamma}$ so that

$$U_t - U_m = C_m^{1-\gamma}\left(c_t - \alpha_2 l_t + \frac{1-\gamma}{2}c_t^2 - \alpha_2\frac{1+\varphi}{2}l_t^2 - \frac{\theta_w}{2}\left(\pi_t^w\right)^2\right)$$

and using the fact that

$$U_m = \frac{C_m^{1-\gamma}}{1-\gamma} - \chi\frac{L_m^{1+\varphi}}{1+\varphi} = \frac{1}{\phi_\psi}\frac{1}{1-\gamma}C_m^{1-\gamma}$$

we can write

$$\log U_t - \log U_m \approx \frac{U_t - U_m}{U_m} = (1-\gamma)\,\phi_\psi\left(c_t - \alpha_2 l_t + \frac{1-\gamma}{2}c_t^2 - \alpha_2\frac{1+\varphi}{2}l_t^2 - \frac{\theta_w}{2}\left(\pi_t^w\right)^2\right)$$

Next, from (41), in the non-stochastic steady state (denoted with bars),

$$\bar{C} = \bar{Y} = \left(\frac{\alpha_2}{\chi}\right)^{\phi_l}\bar{K}^{\alpha_1\phi_\psi}$$

where we have used the fact $\bar{A} = 1$. At the ergodic mean,

$$C_m = Y_m = \left(\frac{\alpha_2}{\chi}\right)^{\phi_l}\bar{K}^{\alpha_1\phi_\psi}\bar{\Psi}^{\phi_\psi}$$

so that

$$\frac{U_m}{\bar{U}} = \left(\frac{C_m}{\bar{C}}\right)^{1-\gamma} = \bar{\Psi}^{(1-\gamma)\phi_\psi} \quad \Rightarrow \quad \log U_m - \log\bar{U} = (1-\gamma)\,\phi_\psi\overline{\psi}$$

and combining,

$$\log U_t - \log\bar{U} = (1-\gamma)\,\phi_\psi\left(\overline{\psi} + c_t - \alpha_2 l_t + \frac{1-\gamma}{2}c_t^2 - \alpha_2\frac{1+\varphi}{2}l_t^2 - \frac{\theta_w}{2}\left(\pi_t^w\right)^2\right)$$

Define

$$\mathbb{U}_t = \frac{1}{1-\gamma}\frac{1}{\phi_\psi}\log\left(\frac{U_t}{\bar{U}}\right) = \overline{\psi} + c_t - \alpha_2 l_t + \frac{1-\gamma}{2}c_t^2 - \alpha_2\frac{1+\varphi}{2}l_t^2 - \frac{\theta_w}{2}\left(\pi_t^w\right)^2$$

Using

$$\begin{aligned} c_t &= \phi_\psi\psi_t + \mu_t \\ l_t &= \frac{1-\gamma}{1+\varphi}\phi_\psi\psi_t + \frac{1}{\alpha_2}\mu_t \end{aligned}$$

we have

$$\begin{aligned}\frac{1-\gamma}{2}c_t^2-\alpha_2\frac{1+\varphi}{2}l_t^2 &= \frac{1-\gamma}{2}\left(\phi_\psi^2\psi_t^2+2\phi_\psi\psi_t\mu_t+\mu_t^2\right)\\ &- \alpha_2\frac{1+\varphi}{2}\left(\left(\frac{1-\gamma}{1+\varphi}\right)^2\phi_\psi^2\psi_t^2+\frac{2\left(1-\gamma\right)}{1+\varphi}\frac{1}{\alpha_2}\phi_\psi\psi_t\mu_t+\left(\frac{1}{\alpha_2}\right)^2\mu_t^2\right)\\ &= \frac{1-\gamma}{2}\left(\phi_\psi^2\psi_t^2+\mu_t^2\right)-\alpha_2\frac{1+\varphi}{2}\left(\left(\frac{1-\gamma}{1+\varphi}\right)^2\phi_\psi^2\psi_t^2+\left(\frac{1}{\alpha_2}\right)^2\mu_t^2\right)\end{aligned}$$

and

$$\begin{aligned}\frac{1-\gamma}{2}\phi_\psi^2\psi_t^2\left(1-\alpha_2\frac{1-\gamma}{1+\varphi}\right) &= \frac{1}{2}\left(1-\gamma\right)\phi_\psi\psi_t^2\\ \frac{1}{2}\mu_t^2\left(1-\gamma-\frac{1+\varphi}{\alpha_2}\right) &= -\frac{1}{2}\frac{1}{\phi_l}\mu_t^2\end{aligned}$$

Last,

$$\begin{aligned}c_t-\alpha_2 l_t &= \left(1-\alpha_2\frac{1-\gamma}{1+\varphi}\right)c_t-\frac{\alpha_2}{1+\varphi}\frac{1}{\phi_l}\mu_t\\ &= \frac{1}{\phi_\psi}c_t-\frac{\alpha_2}{1+\varphi}\frac{1}{\phi_l}\mu_t\\ &= \frac{1}{\phi_\psi}\left(\phi_\psi\psi_t+\mu_t\right)-\frac{\alpha_2}{1+\varphi}\frac{1}{\phi_l}\mu_t\\ &= \psi_t-\left(\frac{1}{\phi_\psi}-\frac{1}{\phi_\psi}\right)\mu_t\\ &= \psi_t\end{aligned}$$

and substituting these expressions back into $\mathbb{U}_t$ and using the definition of λ_w,

$$\mathbb{U}_t=\psi_t-\frac{1}{2}\left(\gamma-1\right)\phi_\psi\left(\psi_t-\overline{\psi}\right)^2-\frac{1}{2}\frac{1}{\phi_l}\mu_t^2-\frac{1}{2}\frac{\alpha_2\nu_w}{\lambda_w\phi_l}\left(\pi_t^w\right)^2$$

Since $\mathbb{E}_{-1}\left[\psi_t\right]=\overline{\psi}$, the negative of the expression is the period loss function in (25).

C.1 I.I.D. Shocks

In the i.i.d. case we have

$$\left(1-\rho\right)\mathbb{E}_{-1}\left[\sum\rho^t\psi_t\right]=\overline{\psi}$$

since $\mathbb{E}_{-1}\left[\psi_t\right]=\overline{\psi}$. Next,

$$\begin{aligned}\left(1-\rho\right)\mathbb{E}_{-1}\left[\sum\rho^t\frac{1}{2}\left(\gamma-1\right)\phi_\psi\left(\psi_t-\overline{\psi}\right)^2\right] &= \left(1-\rho\right)\frac{1}{2}\left(\gamma-1\right)\phi_\psi\psi_a^2\mathbb{E}_{-1}\left[\sum\rho^t\varepsilon_t^2\right]\\ &= \frac{1}{2}\left(\gamma-1\right)\phi_\psi\psi_a^2\sigma_\varepsilon^2\end{aligned}$$

since $\mathbb{E}_{-1}\left[\varepsilon_t^2\right] = \sigma_\varepsilon^2$. Similarly,

$$(1-\rho)\,\mathbb{E}_{-1}\left[\sum \rho^t \frac{1}{2}\frac{1}{\phi_l}\mu_t^2\right] = \frac{1}{2}\frac{1}{\phi_l}\mu_a^2\sigma_\varepsilon^2$$

To solve for π_t^w in terms of μ_t we use the Phillips Curve. Conjecture that

$$\pi_t^w = \zeta_a^w \varepsilon_t$$

Then,

$$\zeta_a^w \varepsilon_t = \lambda_w \mu_a \varepsilon_t \quad \Rightarrow \quad \zeta_a^w = \lambda_w \mu_a$$

and thus,

$$\begin{aligned}(1-\rho)\,\mathbb{E}_{-1}\left[\sum \rho^t \frac{1}{2}\frac{\alpha_2\nu_w}{\lambda_w\phi_l}\left(\pi_t^w\right)^2\right] &= (1-\rho)\,\frac{1}{2}\frac{\alpha_2\nu_w}{\lambda_w\phi_l}\lambda_w^2\mu_a^2\mathbb{E}_{-1}\left[\rho^t\varepsilon_t^2\right] \\ &= \frac{1}{2}\frac{\alpha_2\nu_w\lambda_w}{\phi_l}\mu_a^2\sigma_\varepsilon^2\end{aligned}$$

Finally, substituting into (25),

$$\mathbb{W} = -\overline{\psi} + (\gamma-1)\,\phi_\psi\psi_a^2\frac{\sigma_\varepsilon^2}{2} + \frac{1}{\phi_l}\mu_a^2\frac{\sigma_\varepsilon^2}{2} + \frac{\alpha_2\nu_w\lambda_w}{\phi_l}\mu_a^2\frac{\sigma_\varepsilon^2}{2} \tag{69}$$

Optimal risk adjustment. The optimal risk adjustment solves the problem of a planner who chooses κ to minimize (69), taking all else as given. The first order condition yields

$$-\frac{\partial\overline{\psi}}{\partial\kappa} + (\gamma-1)\,\phi_\psi\psi_a\sigma_\varepsilon^2\frac{\partial\psi_a}{\partial\kappa} = 0$$

From (62) and (63), the derivatives of ψ_t with respect to κ are

$$\frac{\partial\psi_a}{\partial\kappa} = -\omega\Upsilon \tag{70}$$

$$\frac{\partial\overline{\psi}}{\partial\kappa} = \omega\frac{\nu}{1-\alpha_2\nu}\sigma_\varepsilon^2\Upsilon - \alpha_1\kappa\left(\frac{1}{1-\alpha}\frac{\sigma_\varepsilon^2}{1+\sigma_\beta^2\sigma_\varepsilon^2}\right)^2\left(1-\alpha+\sigma_\beta^2\sigma_\varepsilon^2\right)\sigma_\beta^2 \tag{71}$$

Using these expressions (along with the definitions of ω, Υ and α) and equation (62), some lengthy but straightforward algebra yields

$$-\frac{\partial\overline{\psi}}{\partial\kappa} + (\gamma-1)\,\phi_\psi\psi_a\sigma_\varepsilon^2\frac{\partial\psi_a}{\partial\kappa} = -\omega\Upsilon\sigma_\varepsilon^2\left(\kappa_\psi - \kappa\left(1+\omega\Upsilon\kappa_\psi\right)\right) \tag{72}$$

and setting equal to zero and rearranging,

$$\kappa^* = \frac{\kappa_\psi}{1+\kappa_\psi\omega\Upsilon}$$

which satisfies

$$\kappa^* = \psi_a^*\kappa_\psi$$

where ψ_a^* denotes the value of ψ_a at κ^*.

Proof of Proposition 4. Optimal policy in the i.i.d case within the class (13) is characterized by a value μ_a^* that minimizes (69), accounting for the effects on $\overline{\psi}$ and ψ_a. The first order condition gives

$$0 = -\frac{\partial \overline{\psi}}{\partial \mu_a} + (\gamma - 1)\,\phi_\psi \sigma_\varepsilon^2 \psi_a \frac{\partial \psi_a}{\partial \mu_a} + \frac{1}{\phi_l}\sigma_\varepsilon^2 \mu_a + \frac{\alpha_2 \nu_w \lambda_w}{\phi_l}\sigma_\varepsilon^2 \mu_a$$

which we can rewrite as

$$0 = \left(-\frac{\partial \overline{\psi}}{\partial \kappa} + \frac{\gamma - 1}{1 - \alpha_2 \frac{1-\gamma}{1+\varphi}} \sigma_\varepsilon^2 \psi_a \frac{\partial \psi_a}{\partial \kappa} \right) \frac{\partial \kappa}{\partial \mu_a} + \frac{1}{\phi_l}\sigma_\varepsilon^2 \mu_a + \frac{\alpha_2 \nu_w \lambda_w}{\phi_l}\sigma_\varepsilon^2 \mu_a$$

Using (72) along with the definition of κ in (65) and the fact that

$$\frac{\partial \kappa}{\partial \mu_a} = \frac{\kappa_l}{1 + \kappa_\psi \omega \Upsilon}$$

we can substitute into the first order condition to derive

$$0 = (\tau_{\Lambda a} + \kappa_l \mu_a) \frac{\kappa_l \omega \Upsilon \sigma_\varepsilon^2}{1 + \kappa_\psi \omega \Upsilon} + \frac{1}{\phi_l}\sigma_\varepsilon^2 \mu_a + \frac{\alpha_2 \nu_w \lambda_w}{\phi_l}\sigma_\varepsilon^2 \mu_a \tag{73}$$

and rearranging,

$$\mu_a^* = -\frac{\kappa_l \Omega \tau_{\Lambda a}}{\frac{1}{\phi_l} + \frac{\alpha_2 \nu_w \lambda_w}{\phi_l} + \kappa_l^2 \Omega} = -\tau_{\Lambda a} \sigma_\beta^2 \Phi_\Lambda, \quad \text{where} \quad \Phi_\Lambda \equiv \frac{\frac{\kappa_l \omega}{1+\kappa_\psi \omega \Upsilon} \frac{\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2 \sigma_\beta^2}}{\frac{1}{\phi_l} + \frac{\alpha_2 \nu_w \lambda_w}{\phi_l} + \kappa_l^2 \Omega} \tag{74}$$

and $\Omega \equiv \frac{\omega \Upsilon}{1+\kappa_\psi \omega \Upsilon}$. Note that σ_β^2 also appears in the composite Φ_Λ, but it is straightforward to verify that $\frac{\partial \mu_a^*}{\partial \sigma_\beta^2}$ takes on the signs discussed in the text.

□

Costs of policy mistakes. From (73), we have the first derivative of the welfare function with respect to μ_a:

$$\frac{\partial \mathbb{W}}{\partial \mu_a} = (\tau_{\Lambda a} + \kappa_l \mu_a) \frac{\kappa_l \omega \Upsilon \sigma_\varepsilon^2}{1 + \kappa_\psi \omega \Upsilon} + \frac{1}{\phi_l}\sigma_\varepsilon^2 \mu_a + \frac{\alpha_2 \nu_w \lambda_w}{\phi_l}\sigma_\varepsilon^2 \mu_a$$

so that the second derivative is

$$\frac{\partial^2 \mathbb{W}}{\partial \mu_a^2} = \frac{\kappa_l^2 \omega \Upsilon \sigma_\varepsilon^2}{1 + \kappa_\psi \omega \Upsilon} + \frac{1}{\phi_l}\sigma_\varepsilon^2 + \frac{\alpha_2 \nu_w \lambda_w}{\phi_l}\sigma_\varepsilon^2$$

which is a constant. By the fundamental theorem of calculus, can write

$$\begin{aligned} \mathbb{W}(\mu_a) &= \int_{x=\mu_a^*}^{\mu_a} \mathbb{W}'(x)\,dx + \mathbb{W}(\mu_a^*) \\ \mathbb{W}'(\mu_a) &= \int_{x=\mu_a^*}^{\mu_a} \mathbb{W}''(x)\,dx \end{aligned}$$

where there is no constant in the second line since $\mathbb{W}'(\mu_a^*) = 0$.

Then,

$$\begin{aligned}
\mathbb{W}(\mu_a) &= \int_{x=\mu_a^*}^{\mu_a}\int_{y=\mu_a^*}^{x}\mathbb{W}''(y)\,dydx + \mathbb{W}(\mu_a^*) \\
&= \left(\frac{\kappa_l^2\omega\Upsilon\sigma_\varepsilon^2}{1+\kappa_\psi\omega\Upsilon} + \frac{1}{\phi_l}\sigma_\varepsilon^2 + \frac{\alpha_2\nu_w\lambda_w}{\phi_l}\sigma_\varepsilon^2\right)\int_{x=\mu_a^*}^{\mu_a}\int_{y=\mu_a^*}^{x} dydx + \mathbb{W}(\mu_a^*) \\
&= \left(\frac{\kappa_l^2\omega\Upsilon\sigma_\varepsilon^2}{1+\kappa_\psi\omega\Upsilon} + \frac{1}{\phi_l}\sigma_\varepsilon^2 + \frac{\alpha_2\nu_w\lambda_w}{\phi_l}\sigma_\varepsilon^2\right)\int_{x=\mu_a^*}^{\mu_a}(x-\mu_a^*)\,dx + \mathbb{W}(\mu_a^*) \\
&= \left(\frac{\kappa_l^2\omega\Upsilon\sigma_\varepsilon^2}{1+\kappa_\psi\omega\Upsilon} + \frac{1}{\phi_l}\sigma_\varepsilon^2 + \frac{\alpha_2\nu_w\lambda_w}{\phi_l}\sigma_\varepsilon^2\right)\frac{(\mu_a-\mu_a^*)^2}{2} + \mathbb{W}(\mu_a^*)
\end{aligned}$$

so that

$$\mathbb{W}(\mu_a) - \mathbb{W}(\mu_a^*) = \left(\frac{\kappa_l^2\omega\Upsilon\sigma_\varepsilon^2}{1+\kappa_\psi\omega\Upsilon} + \frac{1}{\phi_l}\sigma_\varepsilon^2 + \frac{\alpha_2\nu_w\lambda_w}{\phi_l}\sigma_\varepsilon^2\right)\frac{(\mu_a-\mu_a^*)^2}{2}$$

Since the first term is increasing in σ_β^2, the expression is increasing in the extent of heterogeneity.

Optimal nominal interest rate. Combining (26) and (67), we obtain an equation for the optimal nominal rate of the form $i_t = i_a^* a_t$, where

$$i_a^* = \Phi^i + \tau_{\Lambda a}\sigma_\beta^2\Phi_\Lambda^i, \qquad \Phi_\Lambda^i = \Phi_\Lambda\Phi_\mu^i$$

Substituting for Φ_Λ and Φ_μ^i from those expressions, we can show $\Phi_\Lambda^i > 0$ and $\frac{\partial i_a^*}{\partial\sigma_\beta^2} > (<)$ if $\tau_{\Lambda a} > (<)\,0$.

Optimal fiscal policy. Under fiscal policy in a flexible price environment, there are no costs of inflation and the output gap is given by $\phi_l\tau_{la}a_t$ where τ_{la} captures the cyclicality of the labor income tax. Following similar steps as above, we can derive the welfare loss function as

$$\mathbb{W} = -\overline{\psi} + (\gamma-1)\,\phi_\psi\psi_a^2\frac{\sigma_\varepsilon^2}{2} + \phi_l\tau_{la}^2\frac{\sigma_\varepsilon^2}{2} \tag{75}$$

which is the analog to (69). Again following similar steps as with optimal monetary policy, we can take the first order condition to obtain the optimal fiscal policy:

$$\tau_{la} = -\frac{\kappa_l\Omega\tau_{\Lambda a}}{1+\kappa_l^2\phi_l\Omega} = -\tau_{\Lambda a}\sigma_\beta^2\Phi_\Lambda^f, \qquad \text{where} \qquad \Phi_\Lambda^f \equiv \frac{\frac{\kappa_l\omega}{1+\kappa_\psi\omega\Upsilon}\frac{\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}}{1+\kappa_l^2\phi_l\Omega} \tag{76}$$

Fiscal-monetary coordination. With both fiscal and monetary policy active in the sticky price economy, the welfare function is given by

$$\mathbb{W} = -\overline{\psi} + (\gamma-1)\,\phi_\psi\psi_a^2\frac{\sigma_\varepsilon^2}{2} + \frac{1}{\phi_l}(\mu_a+\phi_l\tau_{la})^2\frac{\sigma_\varepsilon^2}{2} + \frac{\alpha_2\nu_w\lambda_w}{\phi_l}\mu_a^2\frac{\sigma_\varepsilon^2}{2}$$

Taking first order conditions with respect to μ_a and τ_{la}, we can show that $\mu_a^* = 0$ and τ_{la}^* satisfies (76), i.e., monetary policy restores the flexible price economy by completely stabilizing inflation and fiscal policy is set to the same value as in a flexible price economy.

Additional distortions. With labor market distortions, the output gap is given by $(\mu_a + \phi_l \tau_{la})\,\varepsilon_t$. With cost push shocks, we can use the Phillips curve and the method of undetermined coefficients to show that wage inflation satisfies

$$\pi_t^w = (\lambda_w \mu_a + \eta_a)\,\varepsilon_t$$

Using these, the welfare function is given by

$$\mathbb{W} = -\overline{\psi} + (\gamma - 1)\,\phi_\psi \psi_a^2 \frac{\sigma_\varepsilon^2}{2} + \frac{1}{\phi_l}\left(\mu_a + \phi_l \tau_{la}\right)^2 \frac{\sigma_\varepsilon^2}{2} + \frac{\alpha_2 \nu_w}{\lambda_w \phi_l}\left(\lambda_w \mu_a + \eta_a\right)^2 \frac{\sigma_\varepsilon^2}{2}$$

Taking the first order condition with respect to μ_a, we can derive

$$\mu_a^* = -\tau_{\Lambda a} \sigma_\beta^2 \Phi_\Lambda - \eta_a \Phi_\eta - \tau_{la} \Phi_l$$

where

$$\begin{aligned} \Phi_l &= \frac{1 + \kappa_l^2 \phi_l \Omega}{\frac{1}{\phi_l} + \frac{\alpha_2 \nu_w \lambda_w}{\phi_l} + \kappa_l^2 \Omega} \\ \Phi_\eta &= \frac{\frac{\alpha_2 \nu_w}{\phi_l}}{\frac{1}{\phi_l} + \frac{\alpha_2 \nu_w \lambda_w}{\phi_l} + \kappa_l^2 \Omega} \end{aligned}$$

and Φ_Λ is as defined in (74). Taking derivatives with respect to σ_β^2 verifies the signs in the text.

C.2 Persistent Shocks

Full commitment optimal policy. We assume that the policy maker has full commitment and can specify any policy that is linear in current and past state variables. We can write such a policy as an output gap that loads on arbitrary – and potentially infinite – lags of the exogenous shocks, i.e.,

$$\mu_t = \sum_{s=0}^{\infty} \zeta_s^\mu \varepsilon_{t-s}$$

Through the Phillips curve, these loadings have a one-to-one mapping into loadings of (wage) inflation on the shocks, i.e., we can write

$$\pi_t^w = \sum_{s=0}^{\infty} \zeta_s^\pi \varepsilon_{t-s}$$

We work with this latter formulation for much of our derivations below.

Substituting this policy function and the Phillips curve into the welfare loss function (25) and evaluating the summations yields

$$\mathbb{W} = -\overline{\psi} + (\gamma - 1)\,\phi_\psi \psi_a^2 \frac{\sigma_\varepsilon^2}{2} + \frac{1}{\phi_l} \frac{\sigma_\varepsilon^2}{2} \sum_{s=0}^{\infty} \rho^s \left(\frac{\zeta_s^\pi - \rho \zeta_{s+1}^\pi}{\lambda_w} \right)^2 + \frac{\alpha_2 \nu_w}{\lambda_w \phi_l} \frac{\sigma_\varepsilon^2}{2} \sum_{s=0}^{\infty} \rho^s \left(\zeta_s^\pi\right)^2 \tag{77}$$

The risk adjustment, κ, and thus TFP (i.e., $\overline{\psi}$ and ψ_a) depend on the response of the output gap to the shock at time t, ζ_0^μ. Using the Phillips curve, we can write

$$\zeta_0^\mu = \frac{\zeta_0^\pi - \rho \zeta_1^\pi}{\lambda_w}$$

which shows that the TFP terms are affected by ζ_0^π and ζ_1^π, but not by any ζ_s^π where $s \geq 2$. Using (72) and (65), we have

$$-\frac{\partial\overline{\psi}}{\partial\kappa} + (\gamma-1)\,\phi_\psi\psi_a\sigma_\varepsilon^2\frac{\partial\psi_a}{\partial\kappa} = \omega\Upsilon\sigma_\varepsilon^2\left(\tau_{\Lambda a} + \kappa_l\frac{\zeta_0^\pi - \rho\zeta_1^\pi}{\lambda_w}\right)$$

and

$$\begin{aligned}\frac{\partial\kappa}{\partial\zeta_0^\pi} &= \frac{1}{\lambda_w}\frac{\kappa_l}{1+\kappa_\psi\omega\Upsilon}\\ \frac{\partial\kappa}{\partial\zeta_1^\pi} &= -\frac{\rho}{\lambda_w}\frac{\kappa_l}{1+\kappa_\psi\omega\Upsilon}\end{aligned}$$

Using these, the first order conditions from (77) yield

$$\begin{aligned}0 &= \frac{\zeta_s^\pi - \rho\zeta_{s+1}^\pi}{\lambda_w} - \frac{\zeta_{s-1}^\pi - \rho\zeta_s^\pi}{\lambda_w} + \alpha_2\nu_w\zeta_s^\pi, \qquad s \geq 2\\ 0 &= -\left(\tau_{\Lambda a} + \kappa_l\frac{\zeta_0^\pi - \rho\zeta_1^\pi}{\lambda_w}\right)\frac{\kappa_l\omega\Upsilon}{1+\kappa_\psi\omega\Upsilon} + \frac{1}{\phi_l}\frac{\zeta_1^\pi - \rho\zeta_2^\pi}{\lambda_w} - \frac{1}{\phi_l}\frac{\zeta_0^\pi - \rho\zeta_1^\pi}{\lambda_w} + \frac{\alpha_2\nu_w}{\phi_l}\zeta_1^\pi, \qquad s = 1\\ 0 &= \left(\tau_{\Lambda a} + \kappa_l\frac{\zeta_0^\pi - \rho\zeta_1^\pi}{\lambda_w}\right)\frac{\kappa_l\omega\Upsilon}{1+\kappa_\psi\omega\Upsilon} + \frac{1}{\phi_l}\frac{\zeta_0^\pi - \rho\zeta_1^\pi}{\lambda_w} + \frac{\alpha_2\nu_w}{\phi_l}\zeta_0^\pi, \qquad s = 0\end{aligned}$$

The first order condition for $s \geq 2$ can be rearranged to obtain

$$0 = \zeta_s^\pi\left(1 + \rho + \alpha_2\nu_w\lambda_w\right) - \rho\zeta_{s+1}^\pi - \zeta_{s-1}^\pi$$

which is a second order difference equation with a solution of the form

$$\zeta_s^\pi = \sum_{j=1}^{2} c_j b_j^s$$

The method of undetermined coefficients then gives

$$b_j = \frac{\frac{1+\rho+\alpha_2\nu_w\lambda_w}{\rho} \pm \sqrt{\left(\frac{1+\rho+\alpha_2\nu_w\lambda_w}{\rho}\right)^2 - \frac{4}{\rho}}}{2}$$

The larger root is bigger than one which would imply unbounded inflation and thus the relevant solution is given by

$$\zeta_s^\pi = cb^s, \qquad b = \frac{\frac{1+\rho+\alpha_2\nu_w\lambda_w}{\rho} - \sqrt{\left(\frac{1+\rho+\alpha_2\nu_w\lambda_w}{\rho}\right)^2 - \frac{4}{\rho}}}{2} \qquad \text{for } s \geq 1$$

Finally, substituting into the first order conditions for $s = 1, 0$ yields two equations in the two unknowns, c and ζ_0^π:

$$\begin{aligned}0 &= -\left(\tau_{\Lambda a} + \kappa_l\frac{\zeta_0^\pi - \rho cb}{\lambda_w}\right)\frac{\kappa_l\omega\Upsilon}{1+\kappa_\psi\omega\Upsilon} + \frac{1}{\phi_l}\frac{cb - \rho cb^2}{\lambda_w} - \frac{1}{\phi_l}\frac{\zeta_0^\pi - \rho cb}{\lambda_w} + \frac{\alpha_2\nu_w}{\phi_l}cb\\ 0 &= \left(\tau_{\Lambda a} + \kappa_l\frac{\zeta_0^\pi - \rho cb}{\lambda_w}\right)\frac{\kappa_l\omega\Upsilon}{1+\kappa_\psi\omega\Upsilon} + \frac{1}{\phi_l}\frac{\zeta_0^\pi - \rho cb}{\lambda_w} + \frac{\alpha_2\nu_w}{\phi_l}\zeta_0^\pi\end{aligned}$$

Table 4: Full Commitment vs. Optimal Simple Rule

	Full Commitment (1)	Simple Rule (2)
Welfare loss (%)		
Total	1.1462	1.1463
TFP level	0.9245	0.9248
TFP volatility	0.0258	0.0258
Output gap volatility	0.1709	0.1702
Inflation volatility	0.0249	0.0256
Equilibrium statistics		
$\Delta\sigma(\psi_t)$ (%)	-26.723	-26.724
$\sigma(\mu_t)$	1.690	1.686
$\sigma(\pi_t^w)$	0.076	0.077
ϵ_{i_t,ψ_t}	0.784	0.808

which completes the characterization of optimal policy. We can map this back to obtain the optimal loadings of the output gap on past and current shocks using the Phillips curve:

$$\zeta_s^\mu = \frac{\zeta_s^\pi - \rho\zeta_{s+1}^\pi}{\lambda_w}$$

Optimal simple rule. Assume that monetary policy follows (30). Because κ only depends on μ_a, the definitions of $\overline{\psi}$ and ψ_a are unchanged. Substituting the policy rule into the Phillips curve, the method of undetermined coefficients yields

$$\pi_t^w = \frac{\lambda_w\mu_{a_{-1}}}{1-\rho\delta}a_{t-1} + \left(\frac{\rho\lambda_w\mu_{a_{-1}}}{1-\rho\delta} + \lambda_w\mu_a\right)\varepsilon_t$$

Substituting in (25) and evaluating the summations, we can express the welfare loss as

$$\begin{aligned}\mathbb{W} &= -\overline{\psi} + (\gamma-1)\,\phi_\psi\psi_a^2\frac{\sigma_\varepsilon^2}{2} + \frac{1}{\phi_l}\left(\frac{\rho}{1-\rho\delta^2}\mu_{a_{-1}}^2 + \mu_a^2\right)\frac{\sigma_\varepsilon^2}{2} \\ &+ \frac{\alpha_2\nu}{\lambda_w\phi_l}\left(\frac{\rho}{1-\rho\delta^2}\left(\frac{\lambda_w\mu_{a_{-1}}}{1-\rho\delta}\right)^2 + \left(\frac{\rho\lambda_w\mu_{a_{-1}}}{1-\rho\delta} + \lambda_w\mu_a\right)^2\right)\frac{\sigma_\varepsilon^2}{2}\end{aligned}$$

Taking first order conditions with respect to $\mu_{a_{-1}}$ and μ_a gives two equations in two unknowns that characterize the optimal policy.

Table 4 compares the optimal simple rule to the unconstrained full commitment optimal policy from Table 3. Equilibrium outcomes and welfare are extremely close under the two policies, implying that an appropriately chosen rule can approximate very well the unconstrained policy.

Full commitment fiscal policy. As with monetary policy, under a general fiscal policy with full commitment, the fiscal authority chooses a labor market tax that loads on the full sequence of

exogenous shocks, i.e.,

$$\tau_{lt} = \sum_{s=0}^{\infty} \zeta_s \varepsilon_{t-s}$$

Following similar steps as above and assuming flexible prices, we can derive the welfare loss as

$$\mathbb{W} = -\overline{\psi} + (\gamma - 1)\, \phi_\psi \psi_a^2 \frac{\sigma_\varepsilon^2}{2} + \phi_l \frac{\sigma_\varepsilon^2}{2} \sum_{s=0}^{\infty} \rho^s \zeta_s^2$$

which is the analog of (77). It is easily seen that the first order conditions give $\zeta_s = 0$ for $s > 0$, i.e., the optimal tax depends only on the current shock and is completely independent of lagged shocks. Using this result and noting that $\zeta_0 = \tau_{la}$ in (75), we can see that the welfare loss is the same as in that equation and optimal policy satisfies (76).

Table 5 presents the effects of optimal fiscal policy under perfectly flexible wages. Column (1) reports results from a baseline equilibrium where we assume that there are no cyclical labor income taxes, i.e., $\tau_{la} = 0$. In column (2), we show outcomes under the optimal fiscal policy and in column (3) under the optimal policy ignoring heterogeneity.[47] In this case, since the capital wedge is the only distortion (in contrast to the environment above with the pricing friction), the optimal fiscal policy when not accounting for heterogeneity is a laissez-faire one, i.e., $\tau_{la} = 0$. Thus, columns (1) and (3) coincide. The first-best allocation is the same as in Table 3 so we do not repeat it here. We use $\tilde{y}_t$ to denote the output gap, which is equal to $\tilde{y}_t = \phi_l \tau_{la} \varepsilon_t$.

The results are qualitatively (and quantitatively) similar to those in Table 3. In the baseline equilibrium, long-run TFP is 1.4% lower than in the case with no risk adjustment in the allocation and TFP is 29% less volatile. Notice that these outcomes are identical to column (4) in Table 3 – the flexible price economy with laissez-faire fiscal policy is the same as the sticky price economy with complete stabilization of inflation and the output gap (and inactive fiscal policy). Optimal fiscal policy works to reduce both of these effects: long run TFP increases by over 0.5% and TFP volatility also rises, though quite modestly by about 3%. In total, welfare under the optimal policy is 0.28% higher than in the baseline equilibrium. The value of τ_{la} shows that these gains are achieved through aggressive countercyclical tax policy (a procyclical tax): the elasticity of the tax with respect to TFP shocks is large and positive (the negative of τ_{la}). Further, the flexible price economy gives a particularly sharp illustration of the importance of accounting for heterogeneity when setting policy – optimal policy when ignoring heterogeneity corresponds to the policy in the baseline equilibrium and thus, the entirety of the welfare gains from the true optimal policy stem from addressing allocational considerations.

We can also use Table 5 to gauge the benefits from monetary-fiscal coordination. In this case, it is straightforward to verify that just as with i.i.d. shocks, monetary policy completely stabilizes inflation and optimal fiscal policy is then set as it would be in a flexible price economy, i.e., the optimal fiscal policy continues to satisfy (76). This is exactly the scenario in column (2) in Table 5. Thus, comparing the results in that column to column (3) in Table 3 gives the incremental gains of monetary-fiscal coordination over monetary policy alone. It turns out that these gains are modest, about 0.03% of lifetime steady state consumption. The TFP gain is also only modest, about 0.05%. In other words, once monetary policy is optimally determined accounting for heterogeneity, the scope for additional improvements from labor market fiscal policies is small.[48]

[47] For purposes of comparison, we do not recalibrate the value of $\tau_{\Lambda a}$ in Table 5 versus Table 3. However, doing so leads to only small changes in the results.

[48] In the simple environment here, the result largely follows from our focus on labor income taxes. If the fiscal authority had access to two distinct cyclical taxes that did not have exactly proportional effects on labor supply and the capital allocation (e.g., a cyclical tax on firm profits), the first-best allocation could be achieved.

Table 5: Heterogeneity and Optimal Fiscal Policy

	Baseline (1)	Optimal Policy (2)	Ignoring Hetero. (3)
Welfare loss (%)			
Total	1.393	1.115	1.393
TFP level	1.385	0.873	1.385
TFP volatility	0.008	0.029	0.008
Output gap volatility	0.000	0.213	0.000
Equilibrium statistics			
$\Delta\sigma(\psi_t)$ (%)	-28.97	-26.35	-28.97
$\sigma(\tilde{y}_t)$	0.00	1.89	0.00
τ_{la}	0.00	-4.52	0.00

D Calibration

Given the behavior of the output gap in (30), there are five parameters to calibrate internally, namely, the persistence and volatility of exogenous aggregate shocks, δ and σ^2_ε, the dispersion in firm exposures to these shocks, $\sigma^2_{\hat{\beta}}$, the cost of changing wages, θ_w and the capital wedge, $\tau_{\Lambda a}$. Here, we first show that the Taylor rule in (29) maps to an output gap that satisfies (30). We then detail how we set the five remaining parameters jointly to match five moments.

Monetary policy rule. From (9), (28) and assuming (30) holds,

$$\begin{aligned} y_t &= \left(\delta\phi_\psi + \mu_{a_{-1}}\right) a_{t-1} + \left(\phi_\psi\psi_a + \mu_a\right)\varepsilon_t \\ &= y_{a_{-1}} a_{t-1} + y_a \varepsilon_t \end{aligned}$$

In deviations,

$$\begin{aligned} w_t &= \left(1 - \frac{1-\gamma}{1+\varphi}\right) y_t - \frac{1}{1+\varphi}\frac{1}{\phi_l}\mu_t \\ &= \left(1 - \frac{1-\gamma}{1+\varphi}\right)\left(y_{a_{-1}} a_{t-1} + y_a\varepsilon_t\right) - \frac{1}{1+\varphi}\frac{1}{\phi_l}\left(\mu_{a_{-1}} a_{t-1} + \mu_a\varepsilon_t\right) \\ &= \left(\left(1 - \frac{1-\gamma}{1+\varphi}\right)\phi_\psi\delta - \frac{1-\alpha_2}{\alpha_2}\mu_{a_{-1}}\right) a_{t-1} + \left(\left(1 - \frac{1-\gamma}{1+\varphi}\right)\phi_\psi\psi_a - \frac{1-\alpha_2}{\alpha_2}\mu_a\right)\varepsilon_t \\ &= w_{a_{-1}} a_{t-1} + w_a\varepsilon_t \end{aligned}$$

Using the Phillips curve (15) along with (30), we can guess and verify that

$$\pi^w_t = \zeta^w_{a_{-1}} a_{t-1} + \zeta^w_a \varepsilon_t \tag{78}$$

However, this would not be the case in a richer environment with additional distortions, such as the labor market distortions and cost-push shocks we study in Section 3.

where

$$\begin{aligned}\zeta_{a_{-1}}^{w} &= \frac{\lambda_w \mu_a}{1-\rho\delta} \\ \zeta_{a}^{w} &= \frac{\rho\lambda_w \mu_{a_{-1}}}{1-\rho\delta} + \lambda_w \mu_a\end{aligned}$$

By definition

$$\begin{aligned}\pi_{t+1}^{p} &= w_t - w_{t+1} + \pi_{t+1}^{w} \\ &= w_{a_{-1}}\left(a_{t-1} - a_t\right) + w_a\left(\varepsilon_t - \varepsilon_{t+1}\right) + \zeta_{a_{-1}}^{w} a_t + \zeta_{a}^{w}\varepsilon_{t+1} \\ &= \left(\left(1 - \frac{1-\gamma}{1+\varphi}\right)\phi_\psi \delta\left(1-\delta\right) - \frac{1-\alpha_2}{\alpha_2}\left(1-\delta\right)\mu_{a_{-1}} + \frac{\delta\lambda_w\mu_{a_{-1}}}{1-\rho\delta}\right) a_{t-1} \\ &+ \left(\left(1 - \frac{1-\gamma}{1+\varphi}\right)\phi_\psi\left(\psi_a - \delta\right) - \frac{1-\alpha_2}{\alpha_2}\left(\mu_a - \mu_{a_{-1}}\right) + \frac{\lambda_w\mu_{a_{-1}}}{1-\rho\delta}\right)\varepsilon_t \\ &- \left(\left(1 - \frac{1-\gamma}{1+\varphi}\right)\phi_\psi\psi_a - \frac{1-\alpha_2}{\alpha_2}\mu_a - \frac{\rho\lambda_w\mu_{a_{-1}}}{1-\rho\delta} - \lambda_w\mu_a\right)\varepsilon_{t+1} \\ &= \zeta_{a_{-1}}^{p} a_{t-1} + \zeta_{a}^{p}\varepsilon_t + \zeta_{\varepsilon}^{p}\varepsilon_{t+1}\end{aligned}$$

so

$$\mathbb{E}_t\left[\pi_{t+1}^{p}\right] = \zeta_{a_{-1}}^{p} a_{t-1} + \zeta_{a}^{p}\varepsilon_t$$

Substituting into the Taylor Rule, along with the definition of ψ_a gives one representation of the nominal rate:

$$\begin{aligned}i_t &= \phi_y\left(\mu_{a_{-1}} a_{t-1} + \mu_a\varepsilon_t\right) + \phi_\pi\left(\zeta_{a_{-1}}^{p} a_{t-1} + \zeta_{a}^{p}\varepsilon_t\right) \\ &= \left(\phi_y + \phi_\pi\left(1 - \frac{1-\gamma}{1+\varphi}\right)\phi_\psi\delta\left(1-\delta\right) + \phi_\pi\left(\frac{\delta\lambda_w}{1-\rho\delta} - \frac{1-\alpha_2}{\alpha_2}\left(1-\delta\right)\right)\mu_{a_{-1}}\right) a_{t-1} \\ &+ \left(\phi_\pi\left(1 - \frac{1-\gamma}{1+\varphi}\right)\phi_\psi\left(\frac{1-\tau_{\Lambda a}\omega\Upsilon}{1+\kappa_\psi\omega\Upsilon} - \delta\right) + \phi_\pi\left(\frac{\lambda_w}{1-\rho\delta} + \frac{1-\alpha_2}{\alpha_2}\right)\mu_{a_{-1}}\right. \\ &+ \left.\left(\phi_y - \phi_\pi\left(\left(1 - \frac{1-\gamma}{1+\varphi}\right)\phi_\psi\frac{\kappa_l\omega\Upsilon}{1+\kappa_\psi\omega\Upsilon} + \frac{1-\alpha_2}{\alpha_2}\right)\right)\mu_a\right)\varepsilon_t \\ &= i_{a_{-1}}^{TR} a_{t-1} + i_{a}^{TR}\varepsilon_t\end{aligned}$$

Next, using the Euler equation (16),

$$\begin{aligned}i_t &= \gamma\left(\mathbb{E}_t\left[y_{t+1}\right] - y_t\right) + \mathbb{E}_t\left[\pi_{t+1}^{p}\right] \\ &= \gamma\left(y_{a_{-1}} a_t - \left(y_{a_{-1}} a_{t-1} + y_a\varepsilon_t\right)\right) + \zeta_{a_{-1}}^{p} a_{t-1} + \zeta_{a}^{p}\varepsilon_t \\ &= \left(\frac{\varphi\left(1-\gamma\right)}{1+\varphi}\phi_\psi\delta\left(1-\delta\right) + \left(\frac{\delta\lambda_w}{1-\rho\delta} - \frac{1-\alpha_2\left(1-\gamma\right)}{\alpha_2}\left(1-\delta\right)\right)\mu_{a_{-1}}\right) a_{t-1} \\ &+ \left(\frac{\varphi\left(1-\gamma\right)}{1+\varphi}\phi_\psi\left(\frac{1-\tau_{\Lambda a}\omega\Upsilon}{1+\kappa_\psi\omega\Upsilon} - \delta\right) + \left(\frac{\lambda_w}{1-\rho\delta} + \frac{1-\alpha_2\left(1-\gamma\right)}{\alpha_2}\right)\mu_{a_{-1}}\right. \\ &- \left.\left(\frac{\varphi\left(1-\gamma\right)}{1+\varphi}\phi_\psi\frac{\kappa_l\omega\Upsilon}{1+\kappa_\psi\omega\Upsilon} + \frac{1-\alpha_2\left(1-\gamma\right)}{\alpha_2}\right)\mu_a\right)\varepsilon_t \\ &= i_{a_{-1}}^{EE} a_{t-1} + i_{a}^{EE}\varepsilon_t\end{aligned}$$

which is a second representation of the nominal rate. Equating coefficients and grouping terms yields an equation of the form

$$A \begin{bmatrix} \mu_{a_{-1}} \\ \mu_a \end{bmatrix} = B \tag{79}$$

where the elements of the matrices A and B are functions of the Taylor rule coefficients, which shows that any pair ϕ_π and ϕ_y maps to a pair of values $\mu_{a_{-1}}$ and μ_a.

Aggregate shock process. From expression (28), we can derive the serial correlation of aggregate TFP and the standard deviation of TFP growth as

$$\text{corr}\left(\psi_{t+1}, \psi_t\right) = \frac{\delta^3 \sigma_a^2 + \delta \psi_a \sigma_\varepsilon^2}{\delta^2 \sigma_a^2 + \psi_a^2 \sigma_\varepsilon^2} \tag{80}$$

$$\text{std. dev.}\left(\Delta\psi_t\right) = \left(\delta^2 \left(\delta - 1\right)^2 \sigma_a^2 + \left(\left(\delta - \psi_a\right)^2 + \psi_a^2\right) \sigma_\varepsilon^2\right)^{\frac{1}{2}} \tag{81}$$

where $\sigma_a^2 = \frac{\sigma_\varepsilon^2}{1-\delta^2}$. For a given value of ψ_a (which is a function of the other parameters of the model), these two expressions pin down the two parameters δ and σ_ε^2.

Firm exposures to aggregate shocks. To estimate firm-level exposures to aggregate shocks, we first use expression (5) to calculate firm-level capital productivity (up to a term that is time-varying but constant across firms) as

$$\begin{aligned} z_{it} = \log\left(P_{it} Y_{it}\right) - \alpha k_{it} &= \text{const.} + \frac{\nu}{1 - \alpha_2 \nu} a_{it} \\ &= \text{const.} + \beta_i a_t \end{aligned}$$

where, as in the text, $\beta_i = \frac{\nu}{1-\alpha_2\nu}\hat{\beta}_i$, and the constant term captures the effects of movements in aggregate output and wages.

Next, for each firm we estimate a time-series regression of Δz_{it} on aggregate TFP growth, $\Delta\psi_t$. Using expressions (11) and (28), the coefficient from this regression is equal to

$$\beta_i^{obs} = \beta_i \frac{\delta^2 - \delta + 2\psi_a + \psi_a\delta - \psi_a\delta^2}{2\psi_a^2 + 2\delta^2 - 2\psi_a\delta + 2\psi_a^2\delta - 2\psi_a\delta^2} + \text{const.} \tag{82}$$

From here, we can derive the cross-sectional standard deviation of $\hat{\beta}$ as

$$\sigma\left(\hat{\beta}_i\right) = \sigma\left(\beta_i^{obs}\right) \frac{1 - \alpha_2\nu}{\nu} \frac{2\psi_a^2 + 2\delta^2 - 2\psi_a\delta + 2\psi_a^2\delta - 2\psi_a\delta^2}{\delta^2 - \delta + 2\psi_a + \psi_a\delta - \psi_a\delta^2} \tag{83}$$

We trim the 0.5% tails of the estimates and adjust for sampling error by applying the following procedure: we assume that we estimate the betas with error so that

$$\beta_i^{obs} = \beta_i^{true} + e_i$$

where e_i denotes the error. Assuming the error is uncorrelated with the true beta, we can write the cross-sectional variance of the true betas as

$$\sigma^2\left(\beta_i^{true}\right) = \sigma^2\left(\beta_i^{obs}\right) - \sigma^2\left(e_i\right)$$

and since $\mathbb{E}[e_i] = 0$, we have

$$\mathbb{E}\left[\sigma^2(e_i)\right] = \frac{1}{N}\sum_i \mathbb{E}\left[e_i^2\right] = \frac{1}{N}\sum_i \widehat{\sigma^2(e_i)}$$

where $\widehat{\sigma^2(e_i)}$ denotes an unbiased estimator of the variance of the estimation error for firm i.

Finally, using the fact that the mean squared standard error of the beta estimates represents just such an unbiased estimator, i.e., $\frac{1}{N}\sum_i \widehat{\sigma^2(e_i)} = \frac{1}{N}\sum_i \left(s.e.\left(\beta_i^{obs}\right)\right)^2$, we correct the estimate of the cross-sectional dispersion as

$$\mathbb{E}\left[\sigma^2\left(\beta_i^{true}\right)\right] = \sigma^2\left(\beta_i^{obs}\right) - \frac{1}{N}\sum_i \left(s.e.\left(\beta_i^{obs}\right)\right)^2$$

i.e., we substitute $\mathbb{E}\left[\sigma^2\left(\beta_i^{true}\right)\right]$ for $\sigma^2\left(\beta_i^{obs}\right)$ in equation (83).

Wage adjustment cost. From (30), the output gap is equal to

$$\mu_t = \mu_{a_{-1}} a_{t-1} + \mu_a \varepsilon_t$$

and from (78), wage inflation is given by

$$\pi_t^w = \frac{\lambda_w \mu_{a_{-1}}}{1-\rho\delta} a_{t-1} + \left(\frac{\rho\lambda_w \mu_{a_{-1}}}{1-\rho\delta} + \lambda_w \mu_a\right)\varepsilon_t$$

From here, we can derive the coefficient from the regression of wage inflation on the output gap, e.g., the 'slope' of the Phillips curve as

$$\frac{\mathrm{cov}\left(\pi_t^w, \mu_t\right)}{\mathrm{var}\left(\mu_t\right)} = \lambda_w \frac{\frac{\mu_{a_{-1}}^2}{(1-\rho\delta)(1-\delta^2)} + \mu_a\left(\frac{\rho\mu_{a_{-1}}}{1-\rho\delta} + \mu_a\right)}{\frac{\mu_{a_{-1}}^2}{1-\delta^2} + \mu_a^2} \tag{84}$$

Recalling that $\lambda_w = \frac{\alpha_2 \nu_w}{\theta_w \phi_l}$, the coefficient is decreasing in the adjustment cost, θ_w. Note that $\mu_{a_{-1}}$ and μ_a are functions of the other model parameters, including the coefficients in the Taylor rule.

Capital wedge. Using (82), we can calculate β_i (up to a constant) from the regressions of firm-level on aggregate productivity growth. To derive an exact solution for $mrpk_{it}$, use (5) and (6) to write

$$MRPK_{it} = \alpha G Y_t^{\frac{1-\nu}{1-\alpha_2\nu}} A_{it}^{\frac{\nu}{1-\alpha_2\nu}} W_t^{-\frac{\alpha_2\nu}{1-\alpha_2\nu}} K_t^{\alpha-1}\left(\frac{K_{it}}{K_t}\right)^{\alpha-1}$$

Substituting from (55),

$$MRPK_{it} = \alpha G Y_t^{\frac{1-\nu}{1-\alpha_2\nu}} A_{it}^{\frac{\nu}{1-\alpha_2\nu}} W_t^{-\frac{\alpha_2\nu}{1-\alpha_2\nu}} K_t^{\alpha-1}\left(\frac{\left(\mathbb{E}_{t-1}\left[A_{it}^{\frac{\nu}{1-\alpha_2\nu}} e^{-\kappa\varepsilon_t}\right]\right)^{\frac{1}{1-\alpha}}}{\int\left(\mathbb{E}_{t-1}\left[A_{it}^{\frac{\nu}{1-\alpha_2\nu}} e^{-\kappa\varepsilon_t}\right]\right)^{\frac{1}{1-\alpha}} di}\right)^{\alpha-1}$$

Using (59), collecting terms and taking logs,

$$mrpk_{it} = H_t + \beta_i \left(\varepsilon_t - \frac{\xi_{a-1,\varepsilon} \frac{\nu}{1-\alpha_2 \nu} \mathbb{E}_{t-1}\left[a_t\right]}{1 + \sigma_\beta^2 \sigma_\varepsilon^2} + \frac{\kappa \sigma_\varepsilon^2}{1 + \sigma_\beta^2 \sigma_\varepsilon^2} \right) - \frac{1}{2} \beta_i^2 \frac{\sigma_\varepsilon^2}{1 + \sigma_\beta^2 \sigma_\varepsilon^2}$$

where H_t collects all terms that are common across firms.

As a normally distributed random variable with mean $\frac{\nu}{1-\alpha_2 \nu}$, we have $\text{cov}\left(\beta_i^2, \beta_i\right) = 2 \frac{\nu}{1-\alpha_2 \nu} \sigma_\beta^2$ and hence,

$$\text{cov}\left(mrpk_{it}, \beta_i\right) = \frac{\kappa - \frac{\nu}{1-\alpha_2 \nu}}{1 + \sigma_\beta^2 \sigma_\varepsilon^2} \sigma_\varepsilon^2 \sigma_\beta^2$$

where we have used the fact that $\mathbb{E}_{t-1}\left[a_t\right]$ and ε_t are mean zero and so do not affect the expected covariance. Thus, the coefficient from a regression of $mrpk_{it}$ on β_i as specified in (31) is equal to

$$\lambda_\beta = \frac{\kappa - \frac{\nu}{1-\alpha_2 \nu}}{1 + \sigma_\varepsilon^2 \sigma_\beta^2} \sigma_\varepsilon^2$$

which is the exact analog of the approximate slope described in the text. Substituting for κ from (65) and rearranging,

$$\tau_{\Lambda a} = \left(1 + \sigma_\varepsilon^2 \sigma_\beta^2 \left(1 + \kappa_\psi \omega\right)\right) \left(\frac{\lambda_\beta}{\sigma_\varepsilon^2} + \frac{1}{1 + \sigma_\varepsilon^2 \sigma_\beta^2} \frac{\nu}{1 - \alpha_2 \nu} \right) - \kappa_\psi - \kappa_l \mu_a \tag{85}$$

Expressions (80), (81), (83), (84), (85) and the mapping between the two Taylor rule and output gap coefficients derived in (79) yield seven equations in the seven parameters to be calibrated that can be jointly solved for their values.