Risk-Taking, Capital Allocation and Monetary Policy*

Joel M. David† David Zeke‡
FRB Chicago Federal Reserve Board

October 26, 2025

Abstract

We study the implications of firm heterogeneity and financial frictions 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. Financial frictions distort the pricing of risk and hence firm-level costs of capital and the resource allocation. 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 thus 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, financial frictions, aggregate TFP, reallocation, risk premia

*We thank the editor and anonymous referees, Andy Atkeson, Gadi Barlevy, Florin Bilbiie, Mark Bils, Jaroslav Borovicka, Nicolas Caramp (discussant), Sebastian Di Tella, Jonas Fisher, Mark Gertler, Mikhail Golosov, Francois Gourio, Tarek Hassan, Rohan Kekre (discussant), Pete Klenow, Narayana Kocherlakota, Pablo Kurlat, Moritz Lenel, Virgiliu Midrigan, Alessandra Peter, Monika Piazzesi, Elisa Rubbo, Martin Schneider, Stephen Terry, Chris Tonetti, Tom Winberry (discussant), and various seminar and conference participants for helpful comments and suggestions, and Adalyn Schommer for excellent research assistance. The views expressed here are those of the authors and not necessarily those of the Federal Reserve Bank of Chicago, the Federal Reserve Board of Governors, or the Federal Reserve System.

†joel.david@chi.frb.org.
‡david.zeke@frb.gov.

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] Frictions in financial markets distort the pricing of risk and accordingly, firm-level costs of capital and the capital allocation. 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) business cycle dynamics and (ii) the effects and optimal conduct of monetary stabilization policy.

Our point of departure is the workhorse New Keynesian (NK) 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) frictions in financial intermediation that distort the pricing of risk and hence lead to inefficiencies in the allocation of capital. The heterogeneous 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. Our analysis consists of three main parts.

In the first part of the paper, we develop an analytically tractable model that sharply illustrates our key theoretical results. The main simplifications are (i) that the aggregate capital stock is an exogenous fixed endowment and (ii) frictions in financial intermediation that depend on intermediary cash flows rather than on the stock of intermediary net worth.[2] Assumption (i) allows us to hone in on the new allocational considerations in our framework while maintaining the usual linear structure of 'small-scale' NK models (which typically omit capital altogether); assumption (ii) captures the intuition that the marginal value of firm payoffs depends on the household's discount factor and also on financial sector outcomes, while avoiding introducing additional state variables into the economy. The model exactly nests the textbook three equation NK model (e.g., Woodford, 2003 and Galí, 2015) and yields a tractable representation of the equilibrium system, enabling us to formally prove our main results.

At the heart of the 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,

[1] We illustrate these facts in Figure 5 in Appendix C.1. See also David et al. (2022) for related findings.

[2] Our formulation of these frictions captures an earnings-based variant of the canonical Gertler and Karadi (2011) (GK) model, extended to a heterogeneous firm environment. The GK setup has been widely employed to study the implications of financial market imperfections for monetary policy in NK models and so places our paper squarely in this context; we show, however, that other types of frictions can lead to similar effects.

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 effective stochastic discount factor (SDF), which depends on aggregate consumption/output and the severity of financial frictions. Hence, the micro-allocation is in part determined by the economy's macroeconomic dynamics. In turn, aggregate TFP depends on the micro-allocation. We derive a sharp expression for 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. Thus, the micro-allocation and aggregate TFP are both endogenous objects jointly determined in equilibrium.

The risk-return relationship at the micro-level implies a macro-level tradeoff between the 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, such reallocation introduces a wedge between firm-level capital and productivity, which leads to dispersion in MRPK and reduces the level of TFP. The result can be understood as a form of self-insurance – although there are no aggregate savings in this simple economy, insurance against business cycle risk is attained by reallocating capital to less cyclical firms, which reduces aggregate volatility. The cost of this insurance is the foregone output caused by lower TFP due to the misalignment of firm-level capital and productivity.[3]

We study both the positive and normative implications for monetary policy. The monetary policy rule determines the degree of macroeconomic risk and thus plays a key role in shaping the micro-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., reallocating 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] The result implies that (i) monetary policy influences the distribution of capital across firms and thus shifts the economy closer to/further from its production possibilities frontier, captured by the level of TFP, and (ii) firm heterogeneity and reallocation opportunities dampen the direct effects of stabilization policy by causing an offsetting increase in TFP volatility. These 'positive' effects of monetary policy rely only on the presence of nominal rigidities and firm heterogeneity, irrespective of whether there are other distortions (e.g., financial frictions) at work in the economy or not.

Because they distort the pricing of risk and hence the micro-allocation and TFP, financial

[3]Private insurance contracts between firms do not undo these dynamics; the mechanism relies on *systematic* risk, which cannot be diversified away.

[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.

frictions play a key role for the normative implications of monetary policy. The cyclical effects of these distortions call for a cyclical policy in response. We derive a representation of the welfare criterion that can be decomposed into four terms – volatility in inflation and the output gap as in the standard representative firm setup, and the level and volatility of TFP, which together 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 TFP and so both are relevant for understanding the welfare implications of policy. The level of TFP enters the welfare criterion as a first-order term while the other terms are second-order, suggesting that allocational effects may be a key consideration for optimal policy.

We analytically characterize the optimal policy and reach three main results: first, 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 prices and restoring the flexible price outcome in the usual manner; despite the rich heterogeneity in the economy, the monetary authority can safely ignore that heterogeneity, and set policy as if there were a representative firm. Second, when the allocation is distorted by financial frictions, complete price stabilization is no longer optimal and the divine coincidence fails. Financial frictions lead firm owners to put excessively high (low) value on payoffs in economic downturns (expansions) such that the effective discount factor is excessively countercyclical and the allocation overly conservative. Optimal policy entails a countercyclical output gap that raises (reduces) output in downturns (expansions) relative to the natural rate – such a policy incentivizes risk-taking and more closely aligns capital and productivity, increasing TFP and output. Third, 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.[5]

In the second part of the paper, we provide novel empirical evidence for the key implications of the theory, namely, that more countercyclical monetary policy reduces the price of risk, which (i) induces a reallocation of capital towards riskier firms, increasing the covariance of capital and risk exposures and reducing dispersion in MRPK, and accordingly (ii) leads to an increase in both the level and volatility of TFP. Empirically exploring these predictions is challenging: in the theory, monetary policy affects risk premia and hence the capital allocation and TFP via its anticipated state conditionality, i.e., through the expected monetary policy rule. In contrast, identified unexpected monetary policy shocks have the same effects in our model as in the analogous representative firm model, but do not affect risk premia or the allocation; thus

[5]The results do not hinge on monetary policy as the tool of stabilization or the financial origins of distortions per se. For example, similar results hold when fiscal policy is the tool of stabilization, with non-financial distortions to capital choices, and with labor rather than capital market distortions. Further, heterogeneity dampens the optimal response to inflationary cost-push shocks, and this effect is augmented by financial frictions.

such shocks are not informative about the strength of the mechanism.

To address this challenge, we build on recent work in, e.g., Bauer and Swanson (2023a,b) and Bauer et al. (2024), which (1) argues that monetary policy surprises measured from high-frequency data around FOMC announcements in part reflect changing perceptions about the monetary policy rule, and (2) demonstrates that under rational learning about the monetary policy rule, an empirical proxy for shocks to these perceptions is given by the interaction of the monetary policy surprises with a measure of the state of the business cycle. Incorporating this approach into our framework, we construct precisely this proxy and estimate a set of local projections to show that, as predicted by the theory, an increase in the perceived countercyclicality of the monetary policy rule (i) induces a reallocation of capital towards riskier firms, which implies (ii) a fall in MRPK dispersion and an increase in the level and volatility of TFP.

In the third part of the paper, we conduct a numerical evaluation of our findings in the context of a quantitative medium-scale NK model. The two main ingredients relative to the simpler analytic model are (1) aggregate investment subject to adjustment costs, which provides an additional vehicle for insurance against aggregate risk, and (2) frictions in financial intermediation that depend on intermediary net worth as is standard in the extant literature, e.g., Gertler and Karadi (2011). Importantly, our aggregation results from the simpler model continue to hold, such that the equilibrium conditions for macro aggregates are identical to those in a representative firm environment with the addition of an equation characterizing the dynamics of TFP as a function of the capital allocation. We calibrate the model to match a set of standard business cycle moments from aggregate data, e.g., the persistence and volatility of TFP and the relative volatility of investment, as well as a key moment from the micro-data, namely, the dispersion in cyclicality across firms. The calibrated model matches well the relationships between firm cyclicality, MRPK and capital, and generates an impulse response to changes in the perceived monetary policy rule in line with our empirical estimates.

We evaluate equilibrium outcomes and welfare under various policy regimes and find three main results: first, heterogeneity and risk can have significant effects on TFP and welfare. For example, when monetary policy follows a standard Taylor rule, the level of TFP is lower by 0.9% (relative to the non-stochastic steady state), and its volatility is 3.5%, matching the data. In contrast, in the first-best, these values are 0.01% and 4.9%. As in the analytic model, the equilibrium allocation is inefficiently conservative – less cyclical firms hold too much capital, reducing both the level and volatility of TFP. Strikingly, of the total equilibrium welfare losses of 1.4%, the lion's share – roughly 1.2 percentage points (p.p.), or 85% – stems from the reduction in TFP due to the distorted allocation. Second, we find an important role for policy to improve on equilibrium outcomes by correcting allocative inefficiencies: relative to the Taylor rule, optimal monetary policy increases welfare by 1.1 p.p., almost exclusively due to the gains

in TFP. Thus, abstracting from allocational considerations may miss important welfare effects of business cycle fluctuations and, as formally proved in the analytic model, the allocational effects of stabilization policies can be first-order. Third, ignoring heterogeneity can be costly: if the central bank sets policy to the optimal one ignoring heterogeneity, the TFP loss is 0.3% and the welfare gain relative to the Taylor rule is 0.9 p.p. – thus, accounting for heterogeneity increases the potential gains from policy by a significant amount, about 20% by our estimates.

Related literature. Our paper relates 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), and for the implementation of optimal policy, e.g., Bilbiie (2018), Challe (2020), Acharya, Challe, and Dogra (2023), Bilbiie and Ragot (2020) and Bhandari et al. (2021).[6] Recent papers studying the transmission of policy through risk premia and asset revaluations include Kekre and Lenel (2022) and Caramp and Silva (2021). This work has largely focused on 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) optimal policy can entail a redistribution towards agents (in our case, firms) that are more exposed to the business cycle.

We connect closely 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), La'O and Tahbaz-Salehi (2022) and Rubbo (2023), who study optimal policy in environments with dispersed private information and input-output linkages, respectively. Baqaee et al. (2021) and González et al. (2022) link monetary policy to capital (mis)allocation and TFP in models with markup dispersion and firm-level borrowing constraints, respectively. Kurtzman and Zeke (2020) highlight the mis-allocative effects of central bank purchases of corporate securities and Piazzesi et al. (2021) study the climate impact of such purchases via firm-level risk premia and costs of capital. Relatedly, David et al. (2022) link marginal product dispersion to aggregate risk exposures in a partial equilibrium setting and calculate the implications for measures of misallocation and TFP. 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

[6] 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.

of the micro-data. We incorporate heterogeneous risk premia into a workhorse business cycle environment and find an important role for this form of heterogeneity in determining the effects and optimal conduct of monetary policy via its impact on the resource allocation and TFP.[7]

Our empirical analysis adds to an active and growing line of research documenting the heterogeneous impact of monetary policy across firms with differing characteristics, salient examples of which include Ottonello and Winberry (2020), Cloyne et al. (2023) and Jeenas (2023), who study the response to monetary policy shocks across firm leverage/distance to default, age, and liquidity, respectively. Complementary to these papers, we introduce a novel dimension of firm heterogeneity, namely, exposure to aggregate risk as measured through productivity cyclicality, and incorporate newly developed empirical techniques to estimate the reallocation of capital across firms of differing risk induced by shocks to the perceived monetary policy rule.

Finally, our paper bridges two branches of literature that thus far have studied distinct implications of frictions in the financial intermediation sector, namely, (1) a large body of work in macroeconomics examining the direct effects of intermediary balance sheets for aggregate dynamics and monetary policy transmission, but which typically focuses on first-order solutions and thus does not incorporate risk premia, e.g., Bernanke et al. (1999), Gertler and Karadi (2011) and Gertler and Kiyotaki (2010), and (2) the literature in financial economics on 'intermediary-based asset pricing' that instead focuses on the effects of these frictions on the pricing of risk, e.g., He and Krishnamurthy (2013) and Adrian and Shin (2010), and concisely reviewed in He and Krishnamurthy (2018). Our framework merges the insights from these literatures in an otherwise standard NK model featuring heterogeneous firms and reveals how the risk premia effects of intermediation frictions influence the real capital allocation, with important implications for aggregate TFP and macro dynamics, optimal policy and welfare.

2 Analytic 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, $\gamma > 1$.

[7] Also related to our work are recent papers linking monetary policy and TFP growth in models with endogenous innovation, e.g., Moran and Queralto (2018), Garga and Singh (2021) and Jordà et al. (2020).

We assume nominal rigidities in the form of sticky wages. The setup is standard and we provide only a brief overview, with full details in Appendix A.4. 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.[8] Households receive labor income, distributions from ownership of firms and capital via a financial intermediary, and the return on holdings of nominally risk-free bonds (in zero net supply).

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 denotes the real wage, L_t and C_t aggregate labor supply and consumption, and $\mathcal{M}_t > 0$ 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 stochastic discount factor (SDF) is given by:

$$\Lambda_t = \rho \left(\frac{C_t}{C_{t-1}} \right)^{-\gamma} \ . \tag{3}$$

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:

$$Y_t = \left(\int Y_{it}^{\nu} di \right)^{\frac{1}{\nu}} \ , \tag{4}$$

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

$$Y_{it} = A_{it} K_{it}^{\alpha_1} L_{it}^{\alpha_2}, \qquad \alpha_1 + \alpha_2 \leq 1 \ . \tag{5}$$

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

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

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

[8] Following Auclert et al. (2018), we assume these costs enter as an extra additive disutility term in (1). 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 can also accommodate Calvo frictions with a modified slope of the Phillips Curve, or sticky prices rather than wages.

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

$$\alpha_2\nu\frac{Y_t^{1-\nu}Y_{it}^{\nu}}{L_{it}} = W_t \ , \tag{6}$$

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

$$\Pi_{it} = P_{it}Y_{it} - W_tL_{it} = (1-\alpha_2\nu)\,P_{it}Y_{it} = GY_t^{\frac{1-\nu}{1-\alpha_2\nu}}A_{it}^{\frac{\nu}{1-\alpha_2\nu}}W_t^{-\frac{\alpha_2\nu}{1-\alpha_2\nu}}K_{it}^{\alpha} \ , \tag{7}$$

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 an effective discount factor $\Lambda_tT_{\Lambda t}$, where $T_{\Lambda t}$ denotes a capital or financial 'wedge' stemming from frictions in financial markets (described next), which distorts firms' investment choices and hence the allocation of capital.

The firm optimality condition for capital satisfies

$$\mathbb{E}_{t-1}\left[\Lambda_tT_{\Lambda t}MRPK_{it}\right] = R_{t-1}^K, \qquad MRPK_{it} \equiv \alpha\frac{\Pi_{it}}{K_{it}} \ , \tag{8}$$

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

For the remainder of this section, we abstract from aggregate capital accumulation and assume that the total capital stock is an exogenous fixed endowment, i.e., $K_t = \overline{K} \ \forall \ t$. This simplification allows us to hone in on the new allocational effects in our framework while retaining sharp analytic tractability, and ensures that the model here exactly nests the standard NK model without capital. We relax this assumption in our quantitative work in Section 4, but show that the same intuition from this simpler case carries through.

The resource constraints in the economy are then given by

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

Frictional financial intermediation. We consider an analytically tractable, earnings-based variant of the canonical frictional financial intermediation model of Gertler and Karadi (2011) (GK). Here we provide an overview of the setup and key conditions, with full details in Appendix A.1. There is a representative financial intermediary, owned by households. The intermediary owns the capital stock, which it rents to firms for use in production. Firms finance their capital by issuing equity claims to the intermediary. Thus, the intermediary receives both the return on capital and any economic profits, which sum to firm operating profits, Π_{it}. The intermediary issues equity and tax-advantaged debt to households to provide this financing.[9]

The intermediary faces two frictions: first, a convex cost of equity payouts, which implies that the marginal value of intermediary cash flows is equal to $\Phi(CF_t)$, where $\Phi'(CF_t) > 0$ and $\Phi''(CF_t) < 0$;[10] second, a leverage constraint limiting its debt as a proportion of operating earnings: $B_t \leq \xi \Pi_t$ where $\Pi_t = \int \Pi_{it} di$. Appendix A.1 shows that the firm's investment optimality condition takes exactly the form in (8), where

$$T_{\Lambda t} = \frac{\Phi'(CF_t)}{\Phi'(CF_{t-1})} ,$$

i.e., firms price future payoffs using a distorted version of the household SDF, where the distortion reflects the marginal value of cash flows to the intermediary.[11] Assuming a quadratic cost of equity payouts with parameter θ yields a tractable (approximate) expression for the wedge, $T_{\Lambda t}$, as a function of intermediary cash flows (in logs, henceforth denoted with lowercase):

$$\tau_{\Lambda t} \equiv \log T_{\Lambda t} = -\theta \Delta cf_t, \qquad \Delta cf_t \equiv cf_t - cf_{t-1} .$$

Appendix A.1 shows that intermediary cash flows depend on aggregate output, y_t, lagged output, y_{t-1}, and the real interest rate, r_t, and thus shocks to the wedge are proportional to shocks to output, with the factor of proportionality increasing in the cost of equity payouts, θ, and intermediary leverage (i.e., debt-to-earnings ratio), ξ:

$$\tau_{\Lambda t} - \mathbb{E}_{t-1}[\tau_{\Lambda t}] = -\tau_{\Lambda y}(y_t - \mathbb{E}_{t-1}[y_t]), \quad \tau_{\Lambda y} \equiv \frac{\theta}{1 - \frac{\xi}{1+\xi}\overline{R}} > 0 , \tag{9}$$

where $\overline{R} = \frac{1}{\rho}$ is the steady state (gross) risk free interest rate.[12]

Because $\tau_{\Lambda y} > 0$, the distortion is countercyclical, i.e., it falls in expansions (when output

[9] This setup is equivalent to one where firms, rather than intermediaries, own capital.

[10] See, e.g., Jermann and Quadrini (2012) for a related setup in the firm context. Up to first-order, this formulation is equivalent to one with a risk averse intermediary.

[11] We show in Appendix D.1 that the original GK setup, featuring no equity issuance and an asset-based leverage constraint, also leads to an expression for firm investment exactly of the form in (8), with a closely related, though less tractable form of the wedge, which reflects the marginal value of intermediary net worth (a stock) rather than cash flows. We adopt this latter formulation for our quantitative work in Section 4.

[12] Because y_{t-1} and r_t are known and common across firms, they do not affect risk premia or the allocation. Appendix A.1 shows that $\tau_{\Lambda y} > 0$ is a necessary condition for steady state cash flows to be non-negative.

is unexpectedly high) and rises in downturns (when output is unexpectedly low). As a result, the effective discount factor, $\Lambda_t T_{\Lambda t}$, is excessively countercyclical relative to what preferences and the dynamics of consumption alone would dictate, leading firms to act inefficiently averse to bearing aggregate risk.[13]

Generalizations – Although frictions in financial intermediation are a natural candidate (and connect closely to recent work studying financial frictions in NK environments), the capital wedge can capture a wide range of financial – or other types of – distortions in capital markets. For instance, the wedge clearly picks up any cyclical tax on operating profits. Appendix A.2 provides three additional detailed examples stemming from (1) limited asset market participation on the part of households as in the two agent NK (TANK) model of Debortoli and Galí (2018), (2) externalities in preferences, and (3) expectational biases. More generally, Appendix A.3 shows that the framework can accommodate any distortion to the capital choice that can be expressed (or suitably approximated) as a linear function of exogenous shocks, speaking to the wider applicability of our analysis and results.

Additional distortions – We focus primarily on capital market frictions because they most directly distort the firm's investment choice and hence the capital allocation. However, the main insights of our analysis do not hinge on this choice. For example, in Appendix B.3, we introduce cyclical labor market distortions, i.e., a 'labor wedge' (in addition to the wedge induced by sticky wages), as well as cost-push shocks in the Phillips Curve. Because it leads to inefficient fluctuations in labor supply and hence aggregate output, the labor wedge distorts the SDF and price of risk in the same way as the capital wedge. By affecting wage-setting decisions, cost-push shocks similarly lead to inefficient fluctuations and so distort the SDF, depending on the response of monetary policy to these shocks. Thus, capital market frictions per se are not essential to the analysis; rather, distortions on other margins can lead to similar effects.[14]

[13]The result is distinct from a large literature emphasizing excessive risk taking in the financial sector. Indeed, the model is consistent with high levels of financial sector risk (measured by leverage, i.e., debt-to-earnings ratios, and the resulting excess volatility of financial sector cash flows), and teases out a natural implication for the non-financial sector: when the price of risk is excessively high in financial markets, non-financial firms take steps to mitigate risk. Related findings are in Thesmar and Thoenig (2011), who find that limited stock market participation increases the price of risk and leads to inefficiently low firm risk-taking, and Gormsen and Huber (2025), who document a related (though not isomorphic) 'discount rate wedge' between firms' reported discount rates and financial costs of capital, suggesting higher risk aversion on the part of firms than in financial markets.

[14]The relative importance of labor vs capital (or 'investment') wedges in business cycle dynamics has been the subject of some debate: whereas Chari et al. (2007) find a small role for the investment wedge, under somewhat different assumptions, including aggregate adjustment costs and correlation between wedges – features closer to our setup – Christiano and Davis (2006) find a much larger role. Further, as we show below, the allocational effects of the capital wedge in our heterogeneous firm model show up in the aggregate economy largely via TFP (exclusively so in the simple model in this section) and hence would be picked up in the 'efficiency' wedge in a standard business cycle accounting exercise. The result is reminiscent of Buera and Moll (2015), who show that the mapping between primitives and aggregate wedges can be complicated in economies with heterogeneity.

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 A.4):

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

$$y_t = \phi_\psi \psi_t + \mu_t, \qquad \textit{where} \qquad \psi_t \equiv \log \Psi_t, \quad \mu_t \equiv \phi_l \log \mathcal{M}_t \tag{10}$$

and (ii) aggregate TFP is defined by:

$$Y_t = \Psi_t K_t^{\alpha_1} L_t^{\alpha_2}, \quad \Psi_t = \left(\int A_{it}^{\frac{\nu}{1-\alpha_2 \nu}} \left(\frac{K_{it}}{K_t} \right)^{\alpha} di \right)^{\frac{1-\alpha_2 \nu}{\nu}}, \quad \frac{K_{it}}{K_t} = \frac{\mathbb{E}_{t-1}\left[\Lambda_t T_{\Lambda t} \Pi_{it}\right]}{\int \mathbb{E}_{t-1}\left[\Lambda_t T_{\Lambda t} \Pi_{it}\right] di}, \tag{11}$$

where $\phi_\psi > 0$ and $\phi_l > 0$ are composites of model parameters defined in Appendix A.4.

Expression (10) 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.[15]

The left-hand expression in (11) 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 that 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

[15] 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.

firm expected discounted profits relative to average expected discounted profits.

The remaining equilibrium conditions are the Phillips curve and consumption Euler equation (IS curve). These take the usual form and we provide the expressions in Appendix A.4.

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 analytical expression for TFP.

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

$$a_{it} = \hat{\beta}_i a_t + \mathcal{O}_{it}, \quad a_t = \rho_a a_{t-1} + \varepsilon_t \quad \text{where} \quad \varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right), \quad \hat{\beta}_i \sim \mathcal{N}\left(1, \sigma_{\hat{\beta}}^2\right). \tag{12}$$

Here, a_t is an aggregate shock to technology that follows an AR(1) process with persistence ρ_a 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 $\hat{\beta}_i$.[16] 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 the price 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 A.4.

For simplicity, the remainder of this section focuses on the special case when the aggregate shock is i.i.d., i.e., $\rho_a = 0$. Under this special case, we obtain particularly sharp expressions for our main theoretical results. We study the case with persistent shocks in Appendix B.3 (and allow for persistence in our quantitative work in Section 4) and show that the same intuition from this simpler case goes through largely unchanged (since risk premia depend on conditional covariances, adding terms that are known and common across firms has little effect).[17]

[16] 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).

[17] Specifically, persistent shocks add an exogenous AR(1) component to TFP, but do not affect the capital allocation or endogenous component of TFP.

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[18]

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

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 A.5 we show that common interest rate rules imply an output gap of 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. We study such Taylor rules in our empirical analysis in Section 3 and in our quantitative work in Section 4. 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.

Log-linear solution. Under an output gap of the form in (13) and a linear distortion as in (9), 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 (i.e., inflation and the nominal interest rate) 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. 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. This approach yields analytic tractability of the equilibrium system and ensures that our framework and solution fully nest the standard representative firm NK model.

Proposition 2. *Under (i) the assumed stochastic processes on the exogenous shocks 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 functions of the capital allocation. The log-linear equilibrium is characterized by (10), (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,

[18] Our results do not depend on this precise specification – for example, Appendix A.4 shows that the effects of monetary policy on the allocation and TFP go through more generally for any rule that is log-linear in the shock, i.e., where $\mu_t = \mathbb{E}_{t-1}\left[\mu_t\right] + \mu_a a_t$.

μ_t, with slope $\lambda_w \equiv \frac{\alpha_2 \nu_w}{\theta_w \phi_l}$. The Euler equation relates expected output/consumption growth to the nominal interest rate, i_t, and expected price inflation, π^p_{t+1}. 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. The system is identical to the textbook representative firm NK model (e.g., Galí, 2015), except for the endogeneity of TFP.

Proposition 2 introduces two endogenous objects – $\overline{\psi}$ and ψ_a – that determine (i) the level of TFP and (ii) the loading, or responsiveness, of TFP to exogenous shocks. The average level of TFP is equal to $\overline{\psi}$ and its standard deviation to $\sigma(\psi_t) = \psi_a \sigma_\varepsilon$. We provide exact solutions for these constants in expressions (64) and (65) in Appendix A.4 and show that they both depend on the micro-level capital allocation. Characterizing TFP entails characterizing these two terms. The next subsection lays out 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

Proposition 2 presents an exact solution for aggregate TFP with explicit characterizations of $\overline{\psi}$ and ψ_a in Appendix A.4. It turns out, however, that we can derive particularly sharp expressions using a series of first and second-order approximations, and hence for this subsection only, we use such an approximation simply to provide as transparent intuition as possible. We provide detailed derivations of this approach in Appendix A.6, 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.[19] We return to the exact solution for our welfare results and quantitative work in Sections 2.3 and 4.

First, we can approximate aggregate TFP in (11) to obtain

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

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-normal 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

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

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. Thus, the behavior of TFP – its level and cyclical volatility – depends crucially on the micro-allocation, specifically, moments (i) and (ii).

Risk-taking and capital allocation. The allocation is shaped by the risk-return tradeoff in firm-level capital choices. To gain intuition, approximate the optimality condition (8) to obtain

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

where $\tilde{r}_{t-1}$ is a common term that depends on the rental rate of capital and the properties of the SDF, and $\tilde{\lambda}_t = \lambda_t + \tau_{\Lambda t}$ is the log of the effective SDF ($\lambda_t \equiv \log \Lambda_t$). 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) effective 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 expected return as compensation. The reverse holds for less procyclical firms.

The capital wedge from financial frictions leads to inefficiencies in the pricing of risk and hence distorts the risk-return tradeoff and capital allocation. Because 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 relative to safe ones and an allocation that is overly conservative.

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 (proof in Appendix A.6):

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

$$\kappa = \underbrace{\kappa_\psi \psi_a}_{TFP} + \underbrace{\kappa_l \mu_a}_{output\ gap} + \underbrace{\tau_{\Lambda y}\left(\phi_\psi \psi_a + \mu_a\right)}_{capital\ wedge} \tag{21}$$

and $\tilde{A}_{it} \equiv A_{it}^{\frac{\nu}{1-\alpha_2 \nu}}$ and $\kappa_\psi > 0$ and $\kappa_l > 0$ are composites of model parameters.

The risk adjustment κ is a sufficient statistic capturing the impact of aggregate risk on the cross-sectional allocation. Appendix A.7 characterizes and provides further intuition for the constants κ_ψ and κ_l, and shows that κ measures the (negative of the) elasticity of the discounted profitability of capital to aggregate shocks operating (i) through the discount factor and (ii) via equilibrium effects on the wage, W_t, and aggregate demand, Y_t. Thus, κ captures the aggregate risk facing the firm through the joint movement of all relevant state prices.

Expression (21) makes clear that aggregate risk stems from fluctuations in (i) TFP (e.g., the natural rate of output), (ii) the output gap, determined by monetary policy, and (iii) financial frictions via the capital wedge, which itself is a function of TFP and the output gap. Appendix A.4 shows that the risk adjustment κ takes the precisely the same form in the exact solution, i.e., expression (21) 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 the micro-level resource allocation. First, the right-hand equation shows that the capital choice depends on expected productivity and a firm-specific risk premium, given by $\beta_i \kappa \sigma_\varepsilon^2$: 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, with more aggregate risk, i.e., larger κ, reducing the extent of risk-taking at the micro-level, leading to a reallocation of capital towards less cyclical, low beta firms.

Second, 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, with $\kappa \sigma_\varepsilon^2$ the 'price' of aggregate risk: for each unit increase in β_i, the return on capital must increase by this amount. By introducing a wedge between capital and expected productivity, exposure to aggregate risk induces dispersion in *mrpk*. The effects of aggregate risk on firm-level *mrpk* and capital are two sides of the same coin: higher exposure to aggregate risk leads to a higher expected *mrpk*; due to diminishing 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, in the absence of heterogeneity and/or risk adjustments in the allocation, the elasticity of TFP to the exogenous shock, ψ_a, 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, there is a reallocation of capital from more to less procyclical firms, lowering $\text{cov}(k_i, \beta_i)$ in (17). The larger the risk adjustment, the less is risk-taking at the micro-level, i.e., the lower/more negative is this covariance, reducing the responsiveness of TFP to those shocks. Second, by inducing dispersion in $mrpk$, the risk adjustment lowers the 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 $\text{var}(mrpk_i)$ in (17) and depresses the level of TFP.

Thus, heterogeneity and risk generate a tradeoff between the level and 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 be understood as a form of self-insurance. There are no aggregate savings in the economy, yet agents can partially self-insure against business cycle fluctuations by reallocating capital to less cyclical firms, which 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 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 A.6 we provide an explicit solution for κ under our approximation in terms of model primitives (and the exact analog in Appendix A.4).

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 $\mathcal{M}_t = 1$, implying that $\mu_t = 0$ and monetary policy plays no role; and financial 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:

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

and Proposition 2 implies that the discount factor satisfies

$$\log \Lambda_t - \mathbb{E}_{t-1} \log \Lambda_t = -\gamma y_t = -\gamma \psi_a a_t = -\kappa a_t \ , \tag{22}$$

where $\kappa = \gamma\psi_a$ 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. Proposition 3 characterizes the capital allocation and implies that, evaluated at the ergodic mean, the variance of $mrpk$ is approximately $\text{var}\left(mrpk_i\right) = \left(\kappa\sigma_\varepsilon^2\right)^2 \sigma_\beta^2$ and the covariance of firm-level capital and beta is $\text{cov}\left(k_i, \beta_i\right) = \frac{1}{1-\alpha_1}\kappa\sigma_\varepsilon^2\sigma_\beta^2$.

We can use this simplified example to clearly highlight the equilibrium nature of κ. First, from (22), $\kappa = \gamma\psi_a$, i.e., κ depends on the nature of macro dynamics, specifically, the responsiveness 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:

$$\kappa = \frac{\gamma}{1 + \gamma\frac{\alpha_1}{1-\alpha_1}\sigma_\varepsilon^2\sigma_{\hat{\beta}}^2} \ . \tag{23}$$

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 the larger these effects.

2.2 The Effects of Monetary Policy

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: the monetary policy rule pins down the behavior (i.e., the cyclicality) of the output gap, μ_a, and thus directly affects κ. A more procyclical output gap increases the cyclicality of the economy and hence the amount of aggregate risk; a more countercyclical output gap has the opposite effects. Expressions (19) and (20) show that these effects of policy feed through to the resource allocation and via this channel, to the behavior of TFP, i.e., ψ_a and $\overline{\psi}$. Thus, the monetary policy rule determines (i) the 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. The allocational effects of policy boil down to its impact on a single object, κ.

We summarize these 'positive' effects of monetary policy in the following proposition (proof in Appendix A.8):

Proposition 4. *The more countercyclical is monetary policy, i.e., the smaller is μ_a, (i) the lower is the price of risk, i.e., $\frac{\partial \kappa}{\partial \mu_a} > 0$, and (ii) the higher are the level and volatility of TFP, i.e., $\frac{\partial \psi_a}{\partial \mu_a} < 0$, and, if $\kappa > \underline{\kappa}$ where $\underline{\kappa}$ is a positive constant, then $\frac{\partial \overline{\psi}}{\partial \mu_a} < 0$.*

From (19), more countercyclical policy, i.e., lower μ_a and hence κ, causes a reallocation of capital towards riskier, high beta firms, increasing the covariance of firm-level capital and beta and reducing *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 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.[20]

A first key implication is that, in the presence of nominal rigidities and firm heterogeneity, the monetary policy rule affects the level of TFP: by influencing the capital allocation and dispersion in *mrpk*, the degree of business cycle stabilization 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, and via this channel, the monetary policy rule.

A second implication is that through the endogenous response of TFP, firm heterogeneity and reallocation opportunities dampen the effectiveness of stabilization policies. In particular, assuming $\mu_a > -\phi_\psi \psi_a$ (i.e., the output gap is not so countercyclical as to cause output to move inversely with the shock), the standard deviation of output is equal to

$$\sigma(y_t) = (\phi_\psi \psi_a + \mu_a)\sigma_\varepsilon \quad \Rightarrow \quad \frac{\partial \sigma(y_t)}{\partial \mu_a}\frac{1}{\sigma_\varepsilon} = \phi_\psi \frac{\partial \psi_a}{\partial \mu_a} + 1 < 1 \,.$$

In a representative firm environment, ψ_a is always 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, so that $\frac{\partial \psi_a}{\partial \mu_a} = 0$ and $\frac{\partial \sigma(y_t)}{\partial \mu_a}\frac{1}{\sigma_\varepsilon} = 1$. Here,

[20] These allocational effects are quite different than the usual misallocation from staggered price-setting in New Keynesian models with Calvo frictions; in particular, the risk channel we uncover operates via the anticipated conditionality of the monetary policy rule, rather than the realization of unanticipated shocks. Notably, risk-based allocational effects (1) lead to highly persistent, if not permanent, marginal product deviations at the firm-level, consistent with the evidence from the micro-data (e.g., David and Venkateswaran, 2019), and remain operative in the stochastic steady-state, and (2) affect not only the level of TFP, but also its cyclical volatility.

in contrast, TFP depends on the actions of the central bank – in particular, countercyclical policy induces a reallocation of capital towards more procyclical firms, increasing ψ_a and so the volatility of TFP. Increased risk-taking by the private sector in response to the attempted stabilization mitigates the effects of the policy. The result 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 heterogeneous firm economy, this is no longer the case and the central bank must account for the effects of its actions on TFP and the natural rate of output.[21]

These 'positive' effects of monetary policy qualitatively hold irrespective of financial frictions (though we show in Appendix A.8 they are amplified by these frictions). As evidenced by (21) financial frictions also directly affect (indeed, increase) the degree of aggregate risk and hence impact the risk-return tradeoff, with all the accompanying ramifications for the allocation and TFP. The result bears relevance for the large literature studying the relationship between volatility and growth – whereas $\frac{\partial \sigma(y_t)}{\partial \mu_a} > 0$ and $\frac{\partial \overline{\psi}}{\partial \mu_a} < 0$, i.e., less countercyclical policy leads to higher output volatility and lower TFP/output, we also have $\frac{\partial \sigma(y_t)}{\partial \tau_{\Lambda y}} < 0$ and $\frac{\partial \overline{\psi}}{\partial \tau_{\Lambda y}} < 0$, i.e., more severe financial frictions dampen both. Thus, the two inefficiencies – fluctuations in the output gap and financial frictions – 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.[22]

2.3 Optimal Policy

We use this simple model to analytically characterize the optimal monetary policy in the presence of heterogeneity in risk and financial frictions that distort the allocation. In Section 4, we perform a numerical evaluation of various policy regimes within a quantitative medium-scale NK model. Appendix B.1 derives the following second-order approximation to the welfare loss function, expressed as the equivalent decline as a fraction of steady state consumption:

$$\mathbb{W} = (1-\rho)\,\mathbb{E}_{-1}\sum_{t=0}^{\infty}\rho^t\left[-\overline{\psi} + \frac{1}{2}(\gamma-1)\,\phi_\psi\left(\psi_t - \overline{\psi}\right)^2 + \frac{1}{2}\frac{1}{\phi_l}\mu_t^2 + \frac{1}{2}\frac{\alpha_2\nu_w}{\lambda_w\phi_l}\left(\pi_t^w\right)^2\right], \tag{24}$$

[21] 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. 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.

[22] Empirical evidence has found mixed results on the sign of this relationship, e.g., Kormendi and Meguire (1985), Ramey and Ramey (1995), Barlevy (2004) and Koren and Tenreyro (2007). A summary of the evidence is in Jones and Manuelli (2005). Other theoretical mechanisms can also predict either a positive or negative sign, e.g., Ramey and Ramey (1991), Barlevy (2004), Jones et al. (2005), Aghion et al. (2010), Koren and Tenreyro (2013). Theories featuring versions of the risk-return tradeoff at the micro-level such as the one we focus on include Acemoglu and Zilibotti (1997) and Obstfeld (1994).

which expresses the loss as a function of (i) the 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. The last two terms correspond exactly to the welfare function in the textbook representative firm NK model and capture the standard losses from output gap and inflation fluctuations, conditional on the dynamics of TFP. The effects of heterogeneity enter through the first two terms, which are new to our setting – welfare is increasing in the level of TFP and decreasing in its volatility.[23] 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. The level of TFP enters as a first-order term while the other terms are second-order, suggesting that allocational effects may be a key consideration for optimal policy.

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.[24]

First-best allocation. We study as a benchmark the first-best allocation, i.e., the allocation chosen by a social planner who has access to enough instruments to correct both distortions in the economy, i.e., sticky prices and financial frictions. Appendix B.2 shows that the solution entails complete price and output gap stability, i.e., $\mu_t = \pi_t^w = 0 \; \forall \; t$, and $\kappa^* = \kappa_\psi \psi_a^*$, where ψ_a^* is the value of ψ_a in the undistorted economy. In the first-best, the value of κ is not zero. The economy faces a tradeoff between the level of TFP and its volatility, but this tradeoff is efficient. Reducing the risk adjustment further would increase TFP, but would cause a reallocation of capital towards more procyclical firms and inefficient TFP volatility. Increasing the risk adjustment would have the opposite effects, smoothing TFP but reducing its 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.[25]

Optimal monetary policy. As a single policy tool, monetary policy cannot in general attain the first-best value for κ. Intuitively, monetary policy alone cannot fully correct the inefficiencies

[23] 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 mean of the natural rate. We assume optimal time-invariant subsidies are in place to offset all monopoly distortions in the standard way.

[24] We derive the mapping between (13) and interest rate rules in Appendix A.5. More generally, this is the class in which policy is a function only of current state variables, but cannot condition on lagged states. We show below that the optimal rule studied here approximates well the full commitment optimal policy.

[25] The solution for a constrained planner, who faces, but takes as given, the distortionary effects of sticky prices is the same; even though the economy exhibits inefficiencies on other margins, the constrained planner does not further distort the resource allocation.

in the allocation without further distorting the economy on other margins and must balance these considerations. Appendix B.2 derives the optimal policy, which is fully characterized by the elasticity of the output gap to the realization of the exogenous shock:

Proposition 5. *Optimal monetary policy satisfies*

$$\mu_t = \mu_a^* a_t \qquad \textit{where} \qquad \mu_a^* = -\tau_{\Lambda y}\sigma_\beta^2 \Phi_\Lambda \tag{25}$$

and $\Phi_\Lambda > 0$ *is a composite of model parameters defined in Appendix B.2.*

The proposition yields three main implications: first, when financial markets are undistorted ($\tau_{\Lambda y} = 0$) and nominal rigidities are the only source of inefficiency, $\mu_a^* = 0$: in this case, the only factor distorting the allocation and TFP are inefficient fluctuations in the output gap and optimal policy entails complete price/output gap stability, restoring the efficient flexible price outcome. Thus, an extended version of the 'divine coincidence' holds and the central bank faces no tradeoff in setting policy – complete stabilization of the output gap eliminates distortionary fluctuations in labor supply and inflation in the usual way, and now also ensures the micro-allocation and TFP attain the first-best. The result also holds with a representative firm ($\sigma_\beta^2 = 0$) with or without financial frictions ($\tau_{\Lambda y} > 0$), but for a different reason: when firms are identical, financial frictions affect the rental rate of capital, but have no allocational consequences; hence the flexible price equilibrium remains efficient. In this case, TFP is exogenous and outside the control of policy and the central bank faces the usual objective of minimizing output gap and inflation volatility, leading to the usual policy prescription.[26] Thus, accounting for firm heterogeneity can be crucial to properly assess the impact of financial frictions.

Second, when firms are heterogeneous and the allocation is distorted by financial frictions, complete price/output gap stabilization is no longer optimal and the divine coincidence fails. Financial frictions cause the effective discount factor to be excessively countercyclical, which leads to an inefficiently large risk adjustment and overly conservative allocation, resulting in (i) an inefficiently low cyclicality of TFP and (ii) inefficiently high marginal product dispersion, which depresses the level of TFP. Optimal policy is then more aggressively countercyclical – specifically, a countercyclical output gap ($\mu_a^* < 0$) that leans against cyclical fluctuations to reduce the price of risk and partially offset these effects. Such a policy incentivizes risk-taking and more closely aligns capital and productivity, increasing TFP. Because $\mu_a^* \neq 0$ optimal policy allows for inefficient fluctuations in inflation and the output gap. However, policy does not deviate so far from complete stabilization as to fully correct the allocational distortions from financial frictions – doing so would lead to an overly countercyclical output gap that

[26] The result relies on the assumption of a fixed aggregate capital stock; with endogenous capital accumulation as in our quantitative model in Section 4, the distortion also affects the aggregate investment/savings decision.

would lead to costly volatility in that object and inflation.[27] Rather, optimal policy takes on an intermediate value that balances allocational considerations against volatility ones.[28]

Last, the optimal countercyclicality of the output gap is strictly increasing in the extent of heterogeneity, i.e., $\frac{\partial \mu_a^*}{\partial \sigma_\beta^2} < 0$ – with larger differences across firms, the distortionary effects of financial frictions 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.

Optimal nominal interest rate. Appendix B.3 derives the nominal interest rate that implements the optimal output gap as a function of the exogenous shock:

$$i_t = r_t^n + i_a^* a_t \qquad \text{where} \qquad i_a^* = \tau_{\Lambda y} \sigma_\beta^2 \Phi_\Lambda^i, \quad \Phi_\Lambda^i > 0 \,,$$

where r_t^n denotes the natural rate of interest that implements a zero output gap and restores the flexible price equilibrium. The result aligns closely with Proposition 5: (i) the cyclicality of the optimal nominal rate is an increasing function of the capital distortion, implying that, in conjunction with firm heterogeneity, financial frictions lead the optimal nominal rate to rise (fall) more aggressively in response to expansionary (contractionary) shocks, and (ii) $\frac{\partial i_a^*}{\partial \sigma_\beta^2} > 0$, i.e., the strength of this effect is increasing in the extent of heterogeneity. In the absence of either heterogeneity or financial frictions, the optimal nominal rate tracks the natural rate, i.e., $i_t = r_t^n$; with both, the optimal nominal rate is more procyclical than the natural rate. The cyclicality of the nominal rate thus provides a natural metric to gauge the cyclicality of policy.

Persistent shocks. Appendix B.3 studies the case with persistent shocks, i.e., $\rho_a \neq 0$ in expression (12). We show that (i) our analytic framework naturally extends to this case; (ii) because a_{t-1} is a known common term at the time of the capital choice, it does not affect conditional covariances and hence risk premia; accordingly, (iii) the optimal policy response to shocks and the corresponding welfare losses are similar to the i.i.d. case.

The role of commitment. In our framework, monetary policy affects risk premia and hence the capital allocation and TFP via its anticipated state conditionality, i.e., through the expected monetary policy rule. Without commitment, the policy maker cannot influence private sector expectations and simply acts to maximize welfare on a period-by-period basis, taking as given

[27] Formally, $\mu_a^* > -\frac{\tau_{\Lambda y}\phi_\psi \psi_a}{\kappa_l + \tau_{\Lambda y}}$, which from (21) is the value of μ_a that completely corrects the capital wedge.

[28] In this sense, financial frictions in conjunction with firm heterogeneity generate an endogenous cost-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 discussed below), (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.

the resource allocation and dynamics of TFP.[29] Thus, the ability to commit is crucial. However, Appendix B.3 shows that full commitment is not. We analytically solve for the optimal policy under full commitment and numerically illustrate that the welfare losses are only negligibly smaller than under the optimal rule in (25). The result implies that the lion's share of the gains from optimal policy can be attained by an optimal rule that conditions only on current states; full commitment to a response to the entire sequence of realized shocks adds only little.[30]

Costs of policy mistakes. The combination of firm heterogeneity and financial frictions that distort the resource allocation changes the prescription for optimal policy. A corollary to the result is that deviations from optimal policy are more costly in the presence of heterogeneity (with or without financial frictions). Formally, assume that policy is set according to (13) where $\mu_a \neq \mu_a^*$, i.e., policy is not the optimal one. Appendix B.3 proves that the welfare losses from such a deviation are increasing in the extent of heterogeneity, i.e., $\frac{\partial(\mathbb{W}(\mu_a)-\mathbb{W}(\mu_a^*))}{\partial\sigma_\beta^2} > 0$. To state the result another way, the gains from better policy are larger when there is more heterogeneity. 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 also lead to a sub-optimal amount of aggregate risk that distorts the resource allocation.

Additional distortions/policies. Thus far, we have focused on (i) financial market imperfections that directly distort the capital allocation and (ii) monetary policy as the tool of stabilization. However, the effects of risk and heterogeneity are not limited to these considerations. Appendix B.3 extends our analysis along both of these dimensions:

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.[31] With these distortions, optimal policy satisfies an extended version of (25) 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. If the labor wedge is countercyclical, it further strengthens the incentives for countercyclical policy.[32] Because of the cost-push shock, in the standard

[29] Specifically, without the ability to commit, the policy maker's problem is 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, but completely lose out on the gains from improving allocative efficiency.

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

[31] 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.

[32] A large body of work dating back at least to Chari et al. (2007) documents a countercyclical labor wedge.

way, the central bank faces a tradeoff and cannot simultaneously stabilize inflation and the output gap. We show that heterogeneity affects the optimal inflation-output gap tradeoff, and specifically, leads the central bank to respond less aggressively to the inflationary pressure from the cost-push shock to limit distortions to the allocation. Financial frictions augment the effects of heterogeneity on the optimal policy response to both of these distortions.

Fiscal policy – We study optimal fiscal policy in both sticky and flexible price economies, implemented via a cyclical labor income tax/subsidy. We show that optimal policy takes an analogous form to (25): the fiscal authority uses the tax to reduce the cyclicality of labor supply in order to alleviate the distortions from financial frictions. The strength of the policy response is increasing in the degree of heterogeneity. We also consider the case when fiscal and monetary policy are set in tandem, i.e., there is coordination between the fiscal and monetary authorities. We show (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 fiscal and monetary policy are jointly put to work, a natural ordering emerges: first, monetary policy restores the flexible price outcome. Then, fiscal policy is set as if prices were truly flexible.

3 Empirical Validation

In this section we provide empirical support for the main predictions of the theory, namely, the effects of monetary policy on the degree of aggregate risk and hence the capital allocation and TFP described in Section 2.2. Empirically exploring these predictions faces a key challenge: in the theory, monetary policy affects risk premia and hence the capital allocation and TFP via its anticipated state conditionality, i.e., through the expected monetary policy rule, not through more directly identifiable unexpected monetary policy shocks.

To address this challenge, we build on recent work in, e.g., Bauer and Swanson (2023a,b) and Bauer et al. (2024), which (1) argues that monetary policy surprises measured from high-frequency data around FOMC announcements in part reflect changing perceptions about the monetary policy rule, and (2) demonstrates that under rational learning about the monetary policy rule, an empirical proxy for shocks to these perceptions is given by the interaction of the monetary policy surprises with a measure of the state of the business cycle.[33] Incorporating this approach into our framework, we construct precisely this proxy and estimate a set of local projections (LPs) to show that, as predicted by the theory, an increase in the perceived countercyclicality of the monetary policy rule (i) induces a reallocation of capital towards riskier

[33]Indeed, Bauer and Swanson (2023a) and Bauer and Swanson (2023b) show that imperfect information about the monetary policy rule can rationalize many patterns in the data surrounding FOMC announcements, whereas the data are less consistent with a 'Fed information effect.'

firms, increasing the covariance of capital and beta, which implies (ii) a fall in MRPK dispersion and an increase in both the level and volatility of TFP.

To motivate our empirical approach, we assume that monetary policy is implemented via the following Taylor rule:

$$i_t = \phi_i i_{t-1} + \phi_{yt}\mu_t + \phi_\pi \pi_t^p + \xi_t \ , \tag{26}$$

which is a richer version of the simpler rules we studied in Section 2, and is the rule we work with in our quantitative analysis in Section 4. The rule in (26) features two key elements for our empirical approach: a standard i.i.d. monetary policy shock, $\xi_t \sim \mathcal{N}\left(0, \sigma_\xi^2\right)$, and a time-varying output gap coefficient, ϕ_{yt}, which follows an AR(1) process with persistence ρ_ϕ and innovations $\eta_t \sim \mathcal{N}\left(0, \sigma_\eta^2\right)$.[34]

The private sector does not know the true value of ϕ_{yt}, but forms rational expectations of its value based on all available information. Denote the information set of the private sector at date $t-1$ as $\mathcal{I}_{t-1}$, which includes all economic variables through that date. At date t, the output gap and inflation are realized and known to all agents; the central bank then announces the policy rate. The monetary policy surprise, mps_t, is equal to

$$mps_t = i_t - \mathbb{E}\left[i_t | \mathcal{I}_{t-1}, \mu_t, \pi_t^p\right] = \left(\phi_{yt} - \tilde{\phi}_{yt}\right)\mu_t + \xi_t \ , \tag{27}$$

where $\tilde{\phi}_{yt} \equiv \mathbb{E}\left[\phi_{yt} | \mathcal{I}_{t-1}\right]$ denotes agents' prior perceptions of ϕ_{yt}. The surprise is composed of a term that reflects the difference between agents' perceptions and the true coefficient, and a term capturing the pure monetary policy shock. The surprise does not reveal the true coefficient or monetary policy shock, but serves as a noisy signal of the former. Appendix C.2 derives the following result:

Proposition 6. *(i) Under full information about ϕ_{yt}, monetary policy shocks have the same effects as in a representative firm model, but do not affect the risk adjustment, κ, capital allocation or TFP, i.e., $\frac{\partial \kappa_t}{\partial \xi_t} = \frac{\partial \psi_t}{\partial \xi_t} = 0$.*
(ii) Under imperfect information about ϕ_{yt}, private sector perceptions follow

$$\tilde{\phi}_{yt+1} = \rho_\phi \tilde{\phi}_{yt} + \vartheta_t mps_t \mu_t$$

where

$$\vartheta_t = \frac{\rho_\phi \sigma_t^2}{\sigma_t^2 \mu_t^2 + \sigma_\xi^2}, \qquad \sigma_{t+1}^2 = \rho_\phi \sigma_t^2 \left(\rho_\phi - \vartheta_t \mu_t^2\right) + \sigma_\eta^2 \ ,$$

and (iii) Propositions 2, 3 and 4 continue to hold, but where the perceived responsiveness of the output gap to shocks, $\tilde{\mu}_{at}$, and hence the price of risk, κ_t, and level and volatility of TFP, $\overline{\psi}_t$

[34] Although in principle we can allow the other coefficients to vary as well, Bauer et al. (2024) show that identifying perceptions of the inflation coefficient from available data is difficult. All of our results go through unchanged if the inflation and/or lagged interest rate coefficients are time-varying but perfectly known.

and ψ_{at}, depend on the perceived output gap coefficient, $\tilde{\phi}_{yt}$.

Result (i) shows that on their own, pure monetary policy shocks have the same effects in our model as in the analogous representative firm model, but do not affect risk, the capital allocation or TFP. Thus, monetary policy shocks alone are not informative about the strength of the mechanism. Intuitively, monetary policy affects risk through the systematic component of policy, i.e., ϕ_{yt}, and perceptions thereof, not through unexpected changes in the policy rate.

In contrast, result (ii) shows that with imperfect information about ϕ_{yt}, Bayesian agents use the monetary policy surprise as a noisy signal of the true coefficient. Perceptions of ϕ_{yt} follow an AR(1) process and the interaction of the surprise, mps_t, with the output gap, μ_t, captures shocks to those perceptions. Intuitively, monetary policy surprises cause agents to update their beliefs in a state-contingent manner: a hawkish surprise ($mps_t > 0$) in times of economic expansion ($\mu_t > 0$) causes agents to perceive the monetary policy rule as more countercyclical, i.e., more responsive to economic conditions, as does a dovish surprise ($mps_t < 0$) in downturns ($\mu_t < 0$). In turn, a hawkish (dovish) surprise in a downturn (expansion) causes agents to perceive the rule as less countercyclical. We refer to $pmr = mps \times \mu$ as a 'perceived monetary rule' shock and to a positive value that raises perceptions of ϕ_y as a countercyclical *pmr* shock.

Importantly, result (iii) of the proposition shows that the main results from Section 2.2 continue to hold, but with time variation in the price of risk, capital allocation and TFP induced by the time-varying monetary policy rule and perceptions thereof. In particular, we can derive the following empirical predictions, which follow from Propositions 4 and 6:

Empirical predictions. *If the perceived cyclicality of the output gap is decreasing in the perceived output gap coefficient, i.e., $\frac{\partial \tilde{\mu}_{at}}{\partial \tilde{\phi}_{yt}} < 0$,*[35] *then so is the price of risk, i.e., $\frac{\partial \kappa_t}{\partial \tilde{\phi}_{yt}} < 0$, and*
(i) The covariance of capital with beta is increasing and dispersion in mrpk decreasing in $\tilde{\phi}_{yt}$, i.e., $\frac{\partial cov(k_{it},\beta_i)}{\partial \tilde{\phi}_{yt}} > 0$, $\frac{\partial \sigma^2_{mrpk_t}}{\partial \tilde{\phi}_{yt}} < 0$;
(ii) The level and volatility of TFP are increasing in $\tilde{\phi}_{yt}$, i.e., $\frac{\partial \overline{\psi}_t}{\partial \tilde{\phi}_{yt}} > 0$, $\frac{\partial \psi_{at}}{\partial \tilde{\phi}_{yt}} > 0$.

To investigate these predictions, we estimate local projections to trace out impulse response functions (IRFs) to a *pmr* shock. For mps_t, we use the the high-frequency monetary policy surprises from Bauer and Swanson (2023b), the first principal component of 30-minute changes in Eurodollar futures rates around FOMC announcements, which captures changes in policy rate expectations over a horizon of about one year. To implement our analysis at a quarterly frequency, we sum the announcement surprises within each quarter. Because common measures of the output gap are not generally available in real-time, we use (negative) deviations of the unemployment rate from its mean as a measure of the state of the business cycle, $\widehat{un}_t =$

[35] We show in Section 4 that this condition is satisfied in our quantitative model.

$-(un_t - \overline{un}_t)$, such that when $\widehat{un}_t > 0$ the economy is in an expansion and when $\widehat{un}_t < 0$, in a downturn.[36] In practice, we lag the unemployment rate by one quarter to ensure that it (or a close estimate) is in agents' information sets at the time of the monetary policy announcement. Throughout the analysis, we normalize the monetary policy surprise to have a positive 25 basis point effect when the unemployment rate is one percentage point below its mean, i.e., $pmr_t = mps_t \times \widehat{un}_{t-1} = 0.25$, which is a roughly 2.5 standard deviation shock (the standard deviation of pmr_t is 0.11). The positive value of the shock means that it is countercyclical, i.e., it induces agents to perceive a larger output gap coefficient in the monetary policy rule. Appendix C.1 provides full details on the data used throughout this section.

Following the specification in, e.g., Bauer et al. (2024), we estimate LPs of the form:

$$z_{t+h} = \theta^h + \theta^h_{pmr} mps_t \times \widehat{un}_{t-1} + \theta^h_{mps} mps_t + \theta^h_{\widehat{un}} \widehat{un}_{t-1} + \theta^h_z z_{t-1} + \Theta^h_{t-1} + \zeta_{t+h} \,, \tag{28}$$

where z_{t+h} is the variable of interest at horizon h, Θ^h_{t-1} a vector of controls, and ζ_{t+h} a horizon-specific error term. The IRF is given by the sequence of coefficients $\left\{\theta^h_{pmr}\right\}$. The specification controls for the direct effects of mps_t and $\widehat{un}_{t-1}$, the lag of the variable of interest, z_{t-1}, and the additional variables in the vector Θ_{t-1}, which include the lagged policy rate, i_{t-1}, and a set of quarterly fixed-effects to account for possible seasonality, particularly in the micro-data.

Perceived rule shocks and output gap dynamics. First, we verify that the *pmr* shock (1) moves agents' perceptions of ϕ_{yt} and (2) has real effects on the subsequent dynamics of the output gap. To reach result (1), we estimate the IRF of the direct measure of agents' perceptions of ϕ_{yt} constructed by Bauer et al. (2024).[37] Panel A of Figure 1 plots the IRF: there is a persistent and statistically significant positive effect of a countercyclical *pmr* shock on perceptions of ϕ_{yt}, meaning that agents indeed perceive the monetary policy rule as more countercyclical in response to the shock. To reach result (2), we estimate the IRF of the volatility of the realized output gap, measured as the forward-looking eight quarter rolling standard deviation (so that, e.g., the impact response captures the effect on volatility over the eight quarters following the shock).[38] Panel B of Figure 1 plots the IRF: there is a persistent and

[36] For example, the Congressional Budget Office (CBO) estimates of the output gap are generally updated only once or twice per year, see https://www.cbo.gov/data/budget-economic-data. The connection between the output gap and unemployment rate deviation is implied heuristically by Okun's law, and we provide a formal derivation in the context of our model in Appendix C.2, which follows the setup in Galí (2011a,b).

[37] Bauer et al. (2024) construct this measure using period-by-period cross-sectional regressions of expected future policy rates on expected output gaps, inflation and lagged policy rates from the Blue Chip Economic Indicators survey. They reach a similar result using an indicator variable for the state of the business cycle; our findings extend their analysis to a continuous measure, which allows for the response of perceptions to depend on the strength of an expansion or the severity of a downturn at the time of the monetary policy surprise.

[38] We calculate the output gap as the deviation of real GDP from potential real GDP as estimated by the CBO. Our choice of eight quarters aims to strike a reasonable balance between number of observations and

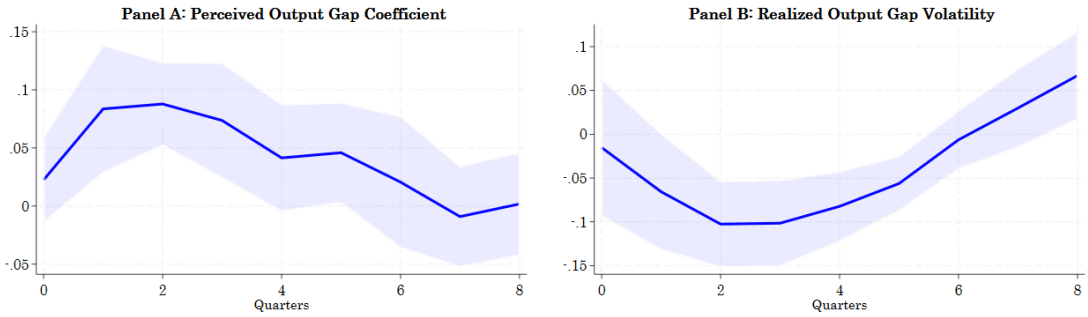


Figure 1: Perceived Rule Shocks, Measured Perceptions and Output Gap Volatility

Notes: Figure displays local projection impulse response functions to a 25 b.p. perceived monetary rule (*pmr*) shock. Panel A displays the response of the perceived output gap coefficient in the monetary policy rule estimated in Bauer et al. (2024). Panel B displays the response of the forward looking eight quarter rolling standard deviation of the realized output gap. Panel B is in log points. Shaded areas report 90% confidence bands based on Newey-West standard errors with the number of lags selected using the nonparametric lag selection procedure of Newey and West (1994).

statistically significant negative effect of the shock on realized output gap volatility, meaning that not only is the rule perceived as more countercyclical, but the actual rule is as well.

Perceived rule shocks and the capital allocation. Turning to our main results, we directly test the key micro-level prediction of the theory by estimating the IRF of the covariance of capital and beta, $\text{cov}(k_{it}, \beta_i)$, to the same *pmr* shock.[39] We use firm-level data from Compustat to compute $\text{cov}(k_{it}, \beta_i)$ on a quarterly basis. We measure k_{it} as (the log of) firm capital. To compute β_i, we calculate firm productivity using data on value-added and capital and then estimate firm-by-firm time-series regressions of firm-level productivity growth on aggregate TFP growth, where the latter is calculated using standard data on aggregate GDP, capital and labor.[40] The coefficients from these regressions yields an observable measure of firm cyclicality, which we denote 'observed beta', β_i^{obs} (relative to the 'true beta', β_i). Because TFP is not equal

meaningfully capturing the short- to medium-term effects of changes in perceptions of the policy rule, but we have verified that similar results hold using other horizons, i.e., six or 10 quarters. In the LP, we control for output gap volatility over the eight quarters preceding the shock and for the lag of its level.

[39]In Appendix C.2, we show that the coefficient from the LP of $\text{cov}(k_{it}, \beta_i)$ is equivalent (up to a constant scaling factor) to that from a panel LP of firm-level capital on the *pmr* shock interacted with firm-level betas, but that the panel specification is more susceptible to issues of measurement and sampling error in the production-side data. In Appendix C.3, we estimate the analogous panel LP using time-varying, high frequency measures of betas computed from stock market data, which are less prone to such errors and allow us to additionally control for both firm and time fixed-effects. We find qualitatively similar results.

[40]Following much of the literature, although we do not have data on firm-level prices, we can infer firm-level productivity (up to a constant) using data on value-added and capital along with the appropriate curvature parameter, α, defined in expression (7), which incorporates the optimal labor choice in (6). In practice, we use sector specific values for α and calculate value-added assuming a common intermediate share of revenues of 0.5, both following David and Venkateswaran (2019). We have verified that our main results are unchanged using sector-specific intermediate shares, calculated as sector-level value-added divided by gross output from the BEA Industry Economic Accounts (the data are available beginning in 1988); for example, the correlation of the betas from the two approaches is 0.91 and their standard deviations are roughly in line (3.1 vs. 3.6, respectively).

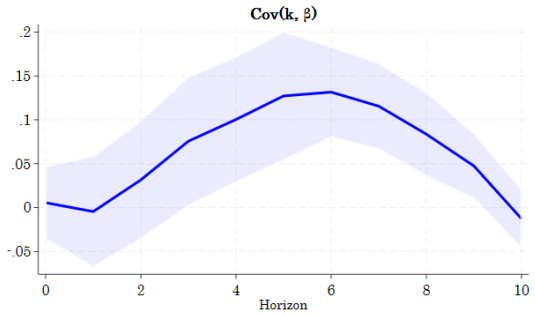


Figure 2: Perceived Rule Shocks and the Allocation of Capital

Notes: Figure displays local projection impulse response function of cov $\left(k_{it}, \beta_i^{obs}\right)$ to a 25 b.p. perceived monetary rule (pmr) shock. Horizon denotes quarter $4+h$ after the shock. Shaded areas report 90% confidence bands based on Newey-West standard errors with the number of lags selected using the nonparametric lag selection procedure of Newey and West (1994).

to the exogenous shock, $\beta_i^{obs} \neq \beta_i$. However, Appendix C.1 shows that they are proportional (up to an additive constant) under a relatively small set of assumptions, namely, the production structure described in expressions (4), (5) and (12). That appendix provides full details of our data sources and method to estimate betas.

Figure 2 plots the IRF of cov $\left(k_{it}, \beta_i^{obs}\right)$ to the pmr shock. To account for the well-documented lags in firm-level capital investment due, for example, to time-to-build/time-to-plan delays or other frictions, we focus on the allocation beginning four quarters following the shock and control for four additional lags of the macro variables (i.e., mps_t, $\widehat{un}_{t-1}$ and i_{t-1}).[41] The figure verifies the key prediction of the theory: a countercyclical pmr shock induces a persistent and statistically significant reallocation of capital towards risky, high beta firms, increasing the covariance between capital and beta.

Perceived rule shocks, marginal product dispersion and TFP. In Appendix C.2, we develop an approximation method to tease out the implications of this reallocation for MRPK dispersion, σ^2_{mrpk}, and for the level and conditional volatility of TFP, $\overline{\psi}$ and ψ_a. The method uses the reallocation implied by the IRF of cov $\left(k_{it}, \beta_i^{obs}\right)$ along with the assumed production structure (in particular, that $\frac{\partial mrpk_{it}}{\partial k_{it}} = -(1-\alpha)$) and two steady state moments that we take directly from the data, namely, the covariance of $mrpk_i$ and β_i^{obs} and the dispersion in β_i^{obs}, but does not use any other model equations; for example, it holds independently of the optimality condition for capital.[42] In brief, there are three main steps: first, the method decomposes the impact of the shock on σ^2_{mrpk} into (1) reallocation across β_i and (2) reallocation

[41] Figure 8 in Appendix D.2 plots the IRF from impact, including the first four quarters after the shock; there is essentially no response of the allocation through the first four quarters, consistent with the timing assumptions in the quantitative analysis in Section 4.

[42] We have also calculated the IRFs of σ^2_{mrpk} and $\overline{\psi}$ using the exact equations of the model and found that the approximation is quite accurate (the response of ψ_a in Figure 3 is exact).

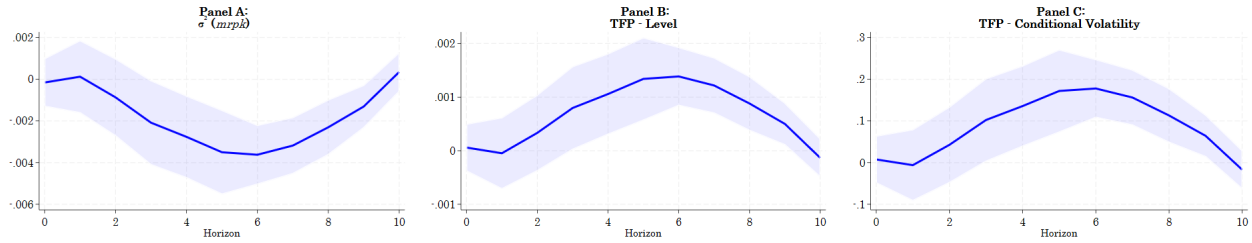


Figure 3: Perceived Rule Shocks, Marginal Product Dispersion and TFP

Notes: Figure displays local projection impulse response functions to a 25 b.p. perceived monetary rule (*pmr*) shock. Horizon denotes quarter $4+h$ after the shock. Panel A displays the response of dispersion in MRPK due to reallocation across β_i. Panel B displays the response of the level of TFP. Panel C displays the response of the conditional volatility of TFP. Panels B and C are in log points. Shaded areas report 90% confidence bands based on Newey-West standard errors with the number of lags selected using the nonparametric lag selection procedure of Newey and West (1994).

that is orthogonal to β_i and hence independent of risk, which, in conjunction with the IRF of cov $\left(k_{it}, \beta_i^{obs}\right)$, enables us to isolate – and thus directly capture – the impact of the shock via risk-based reallocation alone; second, it applies the second-order approximation in (17) to compute the impact on TFP, $\overline{\psi}$; third, it uses the IRF of cov $\left(k_{it}, \beta_i^{obs}\right)$ to compute the impact on the volatility of TFP, ψ_a.[43] We plot the results in Figure 3, which shows that, in line with the model's predictions, the reallocation induced by a countercyclical *pmr* shock implies a decline in σ^2_{mrpk} and an increase in both the level and conditional volatility of TFP.[44]

4 Quantitative Analysis

In this section, we provide a numerical evaluation of the policy effects studied analytically in Section 2. To do so, we extend the simpler analytic model on two dimensions.

First, we allow for aggregate capital investment subject to adjustment costs. A representative firm produces new capital goods and sells these goods to the financial intermediary at the (real) 'cum-dividend' price $\tilde{P}_t^K = P_t^K + R_t^K$ (i.e., price inclusive of the rental rate), where P_t^K is the 'ex-dividend' price. The cost of producing I_t units of new capital is $\left(1 + \frac{\theta_k}{2}\left(\frac{I_t}{I_{t-1}} - 1\right)^2\right) I_t$ units of the final good. Like intermediate good firms, the capital producer is owned by financial intermediaries and acts to maximize expected discounted value. We provide full details of the capital producer's problem in Appendix D.1 and show that optimal investment satisfies

$$\tilde{P}_t^K = 1 + \frac{\theta_k}{2}\left(\frac{I_t}{I_{t-1}} - 1\right)^2 + \theta_k\left(\frac{I_t}{I_{t-1}} - 1\right)\frac{I_t}{I_{t-1}} - \mathbb{E}_t\left[\Lambda_{t+1} T_{\Lambda t+1} \theta_k \left(\frac{I_{t+1}}{I_t} - 1\right)\left(\frac{I_{t+1}}{I_t}\right)^2\right]. \quad (29)$$

[43] Strictly speaking, the approximation yields the IRF of ψ_a up to a proportional constant that depends on the factor mapping observed to true betas. We use the value of this factor from the calibrated model, which puts the IRF into interpretable units, but does not affect its shape or significance.

[44] In Appendix C.3 we estimate the analogous IRFs of aggregate TFP computed directly from the macro-data and show that the same patterns implied by the micro-data show up in the macro-data as well.

The aggregate capital stock evolves as

$$K_{t+1} = I_t + (1-\delta)\, K_t \ . \tag{30}$$

Second, we incorporate the financial intermediation block of Gertler and Karadi (2011) in which intermediary borrowing is limited to a fraction of the stock of net worth, rather than the flow of earnings. The setup largely follows that in GK, modified to allow for heterogeneous firms. Here we provide a brief overview, with full details in Appendix D.1.

As in Section 2, there is a continuum of identical financial intermediaries, owned by households. Intermediaries exit exogenously with probability $1-\sigma$ and are immediately replaced by new entrants, keeping the number of intermediaries fixed. Exiting intermediaries distribute their retained earnings (i.e., net worth, N_t) to households; new entrants are capitalized with a fraction $\frac{n_e}{1-\sigma}$ of the value of exiter assets. As before, intermediaries rent capital to firms for use in production and firms finance this capital by issuing equity claims to intermediaries at price Q_{it}, meaning that intermediaries receive firm operating profits, Π_{it}, as defined in (7). Intermediaries borrow from households at the risk-free rate in order to provide funding to firms.

The end of period t assets of intermediary j are composed of the total market value of its equity claims on firms and the value of its capital holdings, i.e., $A_{jt} = \int Q_{it} S_{ijt} di + P_t^K K_{jt+1}$, where S_{ijt} denotes the quantity of claims held by the intermediary on firm i. Intermediary liabilities, B_{jt}, consist of deposits from households. Intermediary net worth equals $N_{jt} = A_{jt} - B_{jt}$. 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 following incentive constraint:

$$V_{jt} \geq \theta A_{jt} \ , \tag{31}$$

where V_{jt} denotes the intermediary's equity market value. We provide full details of the intermediary's problem in Appendix D.1 and show that the incentive constraint in (31) implies a borrowing constraint of the form

$$A_{jt} \leq \phi_t N_{jt} \quad \Rightarrow \quad B_{jt} \leq (\phi_t - 1)\, N_{jt} \ ,$$

where ϕ_t is an endogenous limit on intermediary leverage (assets-to-net worth ratio), or equivalently, limits intermediary debt as a fraction of net worth, which is the analog of the earnings-based constraint in Section 2.

The firm's investment optimality condition takes exactly the form in (8), where

$$T_{\Lambda t} = \frac{\Theta_t}{\theta + \mathbb{E}_{t-1}\left[\Lambda_t \Theta_t\right] R_t \frac{B_{t-1}}{A_{t-1}}}, \quad \Theta_t \equiv 1 - \sigma + \sigma \frac{\partial V_t}{\partial N_t}$$

and B_t, A_t and V_t denote the debt, assets and equity market value of the intermediary sector, respectively. Thus, the framework leads to a distorted SDF exactly as in Section 2, where the wedge, $T_{\Lambda t}$, reflects the shadow marginal value of net worth to the intermediary sector.

Appendix D.1 lays out the full set of equilibrium conditions for the quantitative model, including the remaining equations characterizing the intermediary sector. Importantly, the exact aggregation results from Section 2 continue to hold, meaning that as before, the equilibrium conditions for macro aggregates are identical to those in the analogous representative firm model with the addition of an equation characterizing the dynamics of TFP as a function of the capital allocation.

4.1 Calibration

The model consists of 19 parameters. We begin by assigning values to the more standard preference and production parameters. Given our focus on capital investment decisions, we assume a period length of one year and set the capital depreciation rate to 0.10 and the annual discount factor, ρ, to 0.98 to match a risk-free rate of 2%.[45] 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, 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.[46] 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 parameters of the intermediary block – the leverage parameter, θ, the survival rate of intermediaries, σ, and the transfer to entering intermediaries, n_e, are set to match the same target moments as GK.[47] We assume that monetary policy in the baseline equilibrium follows the Taylor rule in expression (26) with constant coefficients and no monetary policy shocks.[48]

[45] Calculated as the average nominal yield on one year treasury securities less inflation.

[46] 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 (23) for a simple example and (67) 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 and habits: increasing γ under Epstein-Zin preferences while holding the elasticity of intertemporal substitution fixed also has little effect; adding habits has very similar effects as increasing γ.

[47] The moments are a steady state leverage ratio of four, interest rate spread of 100 basis points, and an average lifespan of intermediaries of 10 years. Because the steady state of the intermediation block here largely coincides with GK, the resulting (annualized) parameter values are the same as well.

[48] These assumptions help keep our welfare exercises transparent (and moreover, conservative), since time variation in the monetary policy rule causes additional inefficient fluctuations and associated welfare losses, but

We incorporate a time-varying coefficient and monetary policy shocks below when comparing the dynamic impulse response to a perceived rule shock in the model to the empirical impulse response estimated in Section 3. The coefficients in the rule are set to standard values of $\phi_i = 0.75$, $\phi_y = 0.5$ and $\phi_\pi = 1.5$. Panel A of Table 1 report the full set of assigned parameters.

The remaining five parameters are calibrated jointly to match five salient moments of the data. First, we return to the more general case of persistent shocks, i.e., we allow ρ_a in (12) to be non-zero. With this modification, Appendix A.4 shows that TFP takes the form

$$\psi_t = \overline{\psi} + \rho_a a_{t-1} + \psi_a \varepsilon_t \ , \tag{32}$$

which is an ARMA(1,1); although the exogenous shock follows an AR(1), endogenous TFP follows an altered dynamic process. We set ρ_a to match an annual persistence of endogenous TFP of 0.81 (i.e., 0.95^4), a standard value in the literature, and σ_ε^2 to match a measured standard deviation of 0.035.[49] The investment adjustment cost, θ_k, is chosen to match the relative volatility of aggregate investment to output. The wage adjustment cost, θ_w, is chosen to match the empirical slope of the wage Phillips curve, i.e., the coefficient from a regression of wage inflation on the output gap.[50] This procedure yields a value of $\theta_w = 191$ to match the flat slope of the empirical wage Phillips curve.[51]

The key micro-level parameter is the degree of heterogeneity across firms, $\sigma_{\hat{\beta}}$. To calculate this value, we use the dispersion of the observed betas estimated in Section 3, $\sigma^2_{\hat{\beta}^{obs}}$, and translate this value into dispersion in the true betas using the mapping described in that section (details in Appendix D.2). An important concern is that these estimates may be affected by issues of sampling and measurement error. Appendix D.2 outlines an approach to adjust the estimated distribution for these types of errors. Implementing this procedure yields a value of $\sigma_{\hat{\beta}} = 3.7$ (the value without correcting for measurement/sampling error would have been roughly double).[52]

Panel B of Table 1 reports the calibrated parameter values and Appendix D.2 provides full details of our data sources and calculations.

ones that are largely distinct from the reallocation mechanism we focus on.

[49]Note that ρ_a and σ_ε are the parameters of the exogenous shock process, not endogenous TFP itself.

[50]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 in the model and data.

[51]See, e.g., Hazell et al. (2022) for related recent evidence on the flat slope of the price Phillips curve. As is common in the literature, the macro-data suggest a flatter Phillips curve than may be implied by micro-data on wage/price setting (see, e.g., Gagliardone et al., 2023 and Galí and Gertler, 1999 for some discussions of the possible disconnect), but what is most relevant for our purposes is the macro-level tradeoff between inflation and output. Hence, we calibrate to match this relationship in the data. Further, under the Rotemberg-style frictions modeled here, firms are continually adjusting wages (though more sluggishly than under flexible wages) and hence the model does not imply an excessively low frequency of wage/price adjustments at the micro-level.

[52]The estimated dispersion is on par with that in David et al. (2022), who find a value of 3.4 using data on stock market returns in conjunction with a structural model similar in spirit to the one here.

Table 1: Parameters and Moments

Panel A: Assigned Parameters

Parameter	*Description*	*Value*	*Parameter*	*Description*	*Value*
	Households			Production	
ρ	Discount factor	0.98	α_1	Capital share	1/3
φ	Inverse Frisch elasticity	1	α_2	Labor share	2/3
γ	Risk aversion	10	ν	Goods elasticity of substitution	0.8
ν_w	Labor elasticity of substitution	6	δ	Capital depreciation rate	0.10
	Financial Intermediaries			Monetary Policy Rule	
θ	Leverage constraint	0.381	ϕ_i	Smoothing parameter	0.75
σ	Intermediary survival rate	0.888	ϕ_y	Output gap coefficient	0.5
n_e	Entrant capitalization	0.008	ϕ_π	Inflation coefficient	1.5

Panel B: Internally Calibrated Parameters and Target Moments

Parameter	*Description*	*Value*	*Moment*	*Data*	*Model*
ρ_a	Persistence of agg. shock	0.67	Persistence of TFP	0.81	0.80
σ_ε	Std. dev. of agg. shock	0.038	Std. dev. of TFP	0.035	0.035
θ_k	Investment adjustment cost	34.9	Rel. vol. of investment, $\frac{\sigma_i}{\sigma_y}$	2.76	2.75
θ_w	Wage adjustment cost	191.4	Slope of wage Phillips curve	0.20	0.20
$\sigma_{\hat{\beta}}$	Std. dev. of risk exposures	3.7	Std. dev. of β_i^{obs}	3.07	3.07

Panel C: Un-Targeted Moments

Moment	*Data*	*Model*
$\text{cov}\left(k_i, \beta_i^{obs}\right)$	-0.99	-1.13
$\text{cov}\left(mrpk_i, \beta_i^{obs}\right)$	0.45	0.44

Model fit and un-targeted moments. Panel B of Table 1 reports the values of the target moments in the data and calibrated model and shows that the model fits the targets closely. Panel C of Table 1 reports the values of two key un-targeted steady state micro-level moments, namely, $\text{cov}\left(mrpk_i, \beta_i^{obs}\right)$ and $\text{cov}\left(k_i, \beta_i^{obs}\right)$. The model-implied values are in line with the data at 0.44 vs. 0.45 for the former and -1.13 vs. -0.99 for the latter, showing that the model quantitatively captures the empirical relationships between firm-level capital, MRPK, and risk. Table 4 in Appendix D.2 reports the model's fit on additional aggregate business cycle moments and shows that it generates patterns of relative volatilities and co-movement between investment, output and consumption in line with the data.

Figure 4 compares the impulse response function of $\text{cov}\left(k_{it}, \beta_i^{obs}\right)$ to a change in the perceived output gap coefficient in the monetary policy rule in the model and data. In the data, this is the IRF plotted in Figure 2. To replicate the empirical exercise in the model, we return to the extended monetary policy rule featuring uncertainty over the output gap coefficient and monetary policy shocks. With these elements, we generate a sequence of perceived coefficients identical to those in the data (i.e., Panel A of Figure 1). Given the annual frequency of the capital decision in the model, each quarterly value of the perceived coefficient determines

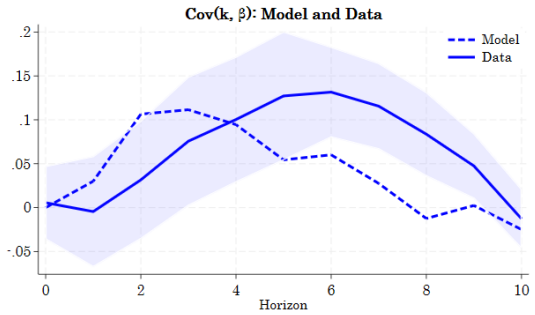


Figure 4: Perceived Rule Shocks and the Allocation of Capital: Model and Data

Notes: Figure displays local projection impulse response functions of cov $\left(k_{it}, \beta_i^{obs}\right)$ to a change in the perceived output gap coefficient in the monetary policy rule in the model and data. Horizon denotes quarter $4+h$ after the shock. The data IRF is as described in Section 3. The model IRF is calculated by generating a sequence of perceived output gap coefficients identical to those in the data. Shaded areas report 90% confidence bands in the data based on Newey-West standard errors with the number of lags selected using the nonparametric lag selection procedure of Newey and West (1994).

the capital allocation four quarters later.[53] Figure 4 shows that the model-generated IRF exhibits dynamics broadly in line with the empirical IRF, namely, a persistent and significant reallocation of capital towards risky, high beta firms. The peak response in the model is of a similar magnitude as the data (0.11 vs. 0.13), which occur at horizons of three and six quarters respectively, suggesting that the reallocation process may be slightly more protracted in the data than in the model given our parsimonious time-to-build specification. The model mean reverts more quickly than the data thereafter, with both model and data essentially at zero at a horizon of 10 quarters. Thus, the model broadly captures the dynamic response of the capital allocation to a surprise change in perceptions of the monetary policy rule.

4.2 Policy Evaluation

Table 2 presents the equilibrium and counterfactual policy exercises. Each column displays welfare losses (as a fraction of steady state consumption; top panel), equilibrium statistics (middle panel), and measures of the cyclicality of monetary policy (bottom panel) under alternative policy regimes. Column (1) corresponds to the equilibrium under the Taylor rule, column (2) the first-best allocation, column (3) optimal policy, and column (4) optimal policy ignoring heterogeneity, i.e., when monetary policy is set to the optimal one in the analogous representative

[53] In the model, the perceived coefficient is sufficient to pin down the price of risk and hence the allocation one year (i.e., four quarters) in advance, allowing us to map the quarterly sequence of perceived coefficients to a quarterly sequence of allocations consistent with the annual nature of the capital choice. Figure 8 in Appendix D.2 plots the IRFs in the model and data from impact, including the first four quarters after the shock; there is essentially no response of the allocation in the data through the first four quarters and, by virtue of the timing assumptions, no response in the model in those quarters.

Table 2: Heterogeneity and Optimal Monetary Policy

	Taylor Rule (1)	First-Best (2)	Optimal Policy (3)	Ignoring Hetero. (4)
Welfare loss (%)				
Total	1.394	0.112	0.324	0.500
Steady State	1.172	0.011	0.028	0.291
Real volatility	0.164	0.123	0.291	0.201
Inflation volatility	0.058	0.000	0.005	0.008
Equilibrium statistics (%)				
$\overline{\psi}$	−0.909	−0.008	−0.015	−0.294
$\sigma(\psi_t)$	3.500	4.906	4.552	4.022
$\sigma(\mu_t)$	0.581	0.000	0.463	0.119
$\sigma(\pi_t^w)$	0.367	0.000	0.108	0.129
Cyclicality of policy				
$\epsilon_{\mu_t,\varepsilon_t}$	0.13	-	−0.12	0.02
ϵ_{i_t,ψ_t}	−1.32	-	0.64	−0.13
$\epsilon_{i_t-r_t^n,\psi_t}$	−0.87	-	1.30	0.34

firm model, without accounting for micro-level heterogeneity and allocational considerations.[54]

We report the total welfare loss, as well as a decomposition of the loss into three components: (i) steady state losses, which stem from a lower average level of TFP and the accompanying effects on labor supply, investment and consumption; (ii) losses from volatility in the 'real' variables, namely labor supply, investment and consumption, which in turn are driven by the exogenous shocks and the response of policy via the output gap; and (iii) losses from inflation volatility.[55] We report four salient statistics of the equilibrium under each policy regime: the reduction in the average level of TFP (relative to the non-stochastic steady state) and the volatilities of TFP, the output gap and inflation. Lastly, we report three measures of the cyclicality of monetary policy: (i) the elasticity of the output gap to the exogenous shock ($\epsilon_{\mu_t,\varepsilon_t}$), which is the most relevant aspect of policy for risk premia, the capital allocation and TFP; (ii) the elasticity of the nominal interest rate to TFP (ϵ_{i_t,ψ_t}); and (iii) the elasticity of the 'interest rate gap' between the nominal rate and the natural interest rate to TFP ($\epsilon_{i_t-r_t^n,\psi_t}$), which captures the additional cyclicality above and beyond movements in the natural rate.

The table yields three main results: first, heterogeneity and risk can have quantitatively sizable effects on TFP and welfare. Column (1) shows that when monetary policy follows a standard Taylor rule, the level of TFP is lower by 0.9% (relative to the non-stochastic steady state), which, including the accompanying effects on labor supply and consumption, leads to

[54] As in Section 2.3, we study optimal policy among the class that loads on current state variables, which are now the current and lagged shocks, ε_t and a_{t-1}, the capital stock, K_t, lagged investment, I_{t-1}, and intermediary net worth, N_t. Our results in Appendix B.3 suggest that optimal rules within this class approximate the full commitment optimal policy well.

[55] In Section 2.3, we were able to decompose 'real' volatility into fluctuations in TFP and the output gap; with investment, these components are no longer separable. With endogenous capital, welfare also includes a term that captures the costs/benefits of the transition from the initial level of capital to its steady state level under the new policy regime. In practice, this term is small and we include it in the steady state loss.

a welfare loss of 1.2%. The welfare costs of real volatility are about 0.16% and of inflation volatility 0.06%. Column (2), the first-best allocation, provides a natural benchmark for these values. In this case, steady state welfare losses are small, 0.1%, and TFP is only 0.008% below the non-stochastic steady state. TFP is more volatile, with a standard deviation of 4.9% vs. 3.5% under the Taylor rule. These findings imply that, as shown in the analytic model in Section 2, the equilibrium allocation is inefficiently conservative – less cyclical firms hold too much capital, which reduces TFP volatility but increases marginal product dispersion, depressing the level of TFP. In contrast, the first-best allocation features a more productive allocation with a higher level of TFP, but also higher TFP volatility. In total, welfare under the Taylor rule is 1.3 p.p. lower than in the first-best. Strikingly, of the total welfare losses under the Taylor rule, the vast majority – roughly 85% – stems from the reduction in TFP due to the distorted allocation. Thus, abstracting from allocational considerations may miss important welfare effects of business cycle fluctuations and, as formally proved in the analytic model, the allocational effects of stabilization policies can be first-order.

Second, the results suggest an important role for policy to improve on equilibrium outcomes by correcting allocative inefficiencies. Column (3) shows that relative to the Taylor rule, optimal policy increases TFP by 0.9 p.p. and reduces the associated welfare losses from lower TFP by 1.14 p.p. 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 – the elasticity of the output gap to the shock is positive under the Taylor rule and becomes negative under the optimal policy. Similarly, the elasticity of the nominal interest rate to TFP is negative under the Taylor rule (due, e.g., to lower marginal costs and inflation) but turns positive under the optimal policy, as does the interest rate gap; under the Taylor rule, the negative elasticity of the interest rate gap to TFP implies that the nominal rate moves less procyclically than the natural rate, while under the optimal policy, the positive elasticity implies that the nominal rate is more procyclical than the natural rate, in line with our analytic findings in Section 2.3.[56] TFP volatility is higher under the optimal policy since the allocation entails more risk-taking, but this offsetting effect turns out to be modest (the welfare costs of real volatility are in total 0.13 p.p. higher under optimal policy). The total welfare gain from implementing the optimal policy relative to the Taylor rule is about 1.1 p.p. Thus, the optimal policy eliminates over 80% of the gap between equilibrium and first-best welfare ($\frac{1.394-0.324}{1.394-0.112}$). Monetary policy alone, however, cannot achieve the first-best, since using cyclical policy to influence the resource allocation also affects the output gap, investment and inflation (optimal policy reduces inflation volatility but increases

[56]The elasticity of the nominal interest rate to TFP of -1.3 under the calibrated Taylor rule is in line with the value in the data of about -1.0.

real volatility relative to the Taylor rule).

Finally, the results imply that ignoring heterogeneity can be costly. Column (4) shows that if the central bank sets policy to the optimal one ignoring heterogeneity, the welfare gain relative to the Taylor rule is 0.9 p.p. Thus, accounting for heterogeneity makes a significant contribution to the potential gains from policy, almost 0.2 p.p. $(0.500 - 0.324)$, or a gain that is 20% larger than when ignoring heterogeneity ($\frac{1.394-0.324}{1.394-0.500}$). The gain comes entirely from an improved allocation and TFP, which is almost 0.3 p.p. higher under optimal policy than when ignoring heterogeneity, but at a cost of higher volatility, particularly in the real variables. Again, the central bank achieves this gain through a more aggressively countercyclical policy – the elasticity of the output gap to the shock (nominal interest rate to TFP) is positive (negative) when ignoring heterogeneity, but negative (positive) under the optimal policy; the elasticity of the interest rate gap to TFP is modestly positive when ignoring heterogeneity, i.e., the nominal rate tracks the natural rate close to one-for-one, but is large and positive under optimal policy.[57] In short, the nominal interest rate is not procyclical enough when not accounting for heterogeneity, nor is the output gap sufficiently countercyclical.

5 Conclusion

In this paper, we have studied the implications of firm heterogeneity – specifically, differences in cyclicality – and financial frictions 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 level and cyclical volatility. We find robust empirical support for these links in the data. Because they inefficiently raise the price of risk, financial frictions lead the allocation to be overly conservative and, to incentivize risk-taking, optimal policy to be more aggressively countercyclical than in the observationally equivalent representative firm model. Thus, taken together, firm heterogeneity and financial frictions strengthen the rationale for such stabilization policies. Evaluating these findings in a quantitative medium-scale New Keynesian model suggests that the welfare gains from implementing policies that account for allocational considerations can be significant.

Because we study this mechanism in an otherwise textbook business cycle model, there is ample room for extensions and further research along several margins. For example, including the richer set of shocks/distortions that have been pointed to in the literature as playing a role in

[57] The interest rate gap is non-zero when ignoring heterogeneity because the financial friction still distorts aggregate investment.

business cycle dynamics (such as those studied qualitatively in Section 2.3) would be valuable. Of particular interest would be shocks that lead to time-variation in risk premia and hence state conditionality in optimal policy. Another fruitful next step would be to study additional dimensions of firm heterogeneity that may interact with (or indeed, lead to) differences in cyclicality, i.e., firm-specific markups or technologies. We believe that the lessons of our study are likely to remain: monetary policy, though a 'blunt instrument,' can have consequential allocational effects, with significant implications for the design of appropriate policy.

References

Abel, A. B. (1990): "Asset Prices under Habit Formation and Catching Up with the Joneses," in *American Economic Review, Papers and Proceedings*, vol. 80, 38–42.

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 (2023): "Optimal monetary policy according to HANK," *American Economic Review*, 113, 1741–1782.

Adrian, T. and H. S. Shin (2010): "Financial intermediaries and monetary economics," in *Handbook of monetary economics*, Elsevier, vol. 3, 601–650.

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*.

Bauer, M. D., C. E. Pflueger, and A. Sunderam (2024): "Perceptions about monetary policy," *The Quarterly Journal of Economics*, 139, 2227–2278.

Bauer, M. D. and E. T. Swanson (2023a): "An alternative explanation for the "fed information effect"," *American Economic Review*, 113, 664–700.

——— (2023b): "A reassessment of monetary policy surprises and high-frequency identification," *NBER Macroeconomics Annual*, 37, 87–155.

Bernanke, B. S., M. Gertler, and S. Gilchrist (1999): "The financial accelerator in a quantitative business cycle framework," *Handbook of macroeconomics*, 1, 1341–1393.

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.

Buera, F. J. and B. Moll (2015): "Aggregate Implications of a Credit Crunch: The Importance of Heterogeneity," *American Economic Journal: Macroeconomics*, 7, 1–42.

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.

Christiano, L. and J. M. Davis (2006): "Two flaws in business cycle accounting," .

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

Cloyne, J., C. Ferreira, M. Froemel, and P. Surico (2023): "Monetary policy, corporate finance, and investment," *Journal of the European Economic Association*, 21, 2586–2634.

David, J. M., L. Schmid, and D. Zeke (2022): "Risk-adjusted capital allocation and misallocation," *Journal of Financial Economics*, 145, 684–705.

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.

Gagliardone, L., M. Gertler, S. Lenzu, and J. Tielens (2023): "Anatomy of the Phillips Curve: Micro Evidence and Macro Implications," Tech. rep., National Bureau of Economic Research.

Galí, J. (2011a): "The return of the wage Phillips curve," *Journal of the European Economic Association*, 9, 436–461.

——— (2011b): *Unemployment fluctuations and stabilization policies: a new Keynesian perspective*, MIT press.

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

Galí, J. and M. Gertler (1999): "Inflation dynamics: A structural econometric analysis," *Journal of monetary Economics*, 44, 195–222.

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.

Gertler, M. and N. Kiyotaki (2010): "Financial intermediation and credit policy in business cycle analysis," in *Handbook of monetary economics*, Elsevier, vol. 3, 547–599.

Gomme, P., B. Ravikumar, and P. Rupert (2011): "The return to capital and the business cycle," *Review of Economic Dynamics*, 14, 262–278.

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

Gormsen, N. J. and K. Huber (2025): "Corporate discount rates," *American Economic Review*, 115, 2001–2049.

Hazell, J., J. Herreno, E. Nakamura, and J. Steinsson (2022): "The slope of the Phillips Curve: evidence from US states," *The Quarterly Journal of Economics*, 137, 1299–1344.

He, Z. and A. Krishnamurthy (2013): "Intermediary asset pricing," *American Economic Review*, 103, 732–770.

——— (2018): "Intermediary asset pricing and the financial crisis," *Annual Review of Financial Economics*, 10, 173–197.

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

Jeenas, P. (2023): "Firm balance sheet liquidity, monetary policy shocks, and investment

dynamics," .

Jermann, U. and V. Quadrini (2012): "Macroeconomic effects of financial shocks," *American Economic Review*, 102, 238–271.

Jones, L. E. and R. E. Manuelli (2005): "Neoclassical models of endogenous growth: The effects of fiscal policy, innovation and fluctuations," *Handbook of Economic Growth*, 1, 13–65.

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 (2022): "Monetary policy, redistribution, and risk premia," *Econometrica*, 90, 2249–2282.

Kendrick, J. W. (1961): *Productivity Trends in the United States*, National Bureau of Economic Research, Inc.

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 (2022): "Optimal monetary policy in production networks," *Econometrica*, 90, 1295–1336.

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.

Newey, W. K. and K. D. West (1994): "Automatic lag selection in covariance matrix

estimation," *The review of economic studies*, 61, 631–653.

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.

Piazzesi, M., M. Papoutsi, and M. Schneider (2021): "How unconventional is green monetary policy?" Tech. rep., Mimeo, Stanford University.

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.

Rubbo, E. (2023): "Networks, Phillips curves, and monetary policy," *Econometrica*, 91, 1417–1455.

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.

Thesmar, D. and M. Thoenig (2011): "Contrasting trends in firm volatility," *American Economic Journal: Macroeconomics*, 3, 143–180.

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 Derivations and Proofs

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

A.1 Frictional Financial Intermediation

The setup is a modified version of the canonical model of Gertler and Karadi (2011) (GK) that allows for sharp analytic tractability. There is a representative financial intermediary, owned by the household. The intermediary owns the capital stock, which it rents to firms for use in production. Firms finance their capital by issuing equity claims to the intermediary. Thus, the intermediary receives both the return on capital and any economic profits that stem from firm monopoly power, which sum to firm operating profits Π_{it} as defined in (7). The intermediary issues equity and debt to households in order to provide funding to firms. There is a tax advantage of debt, which gives the intermediary an incentive to take on leverage. Taxes are rebated to the intermediary lump-sum.

Intermediary cash flows are equal to

$$\begin{aligned} CF_t &= (1-\tau)\left(\int_i \left(S_{it-1}\tilde{\Pi}_{it} + (S_{it-1}-S_{it})\,Q_{it}\right)di + R_t^K K_{t+1} - \tilde{P}_t^K\left(K_{t+1}-K_t\right)\right) \\ &+ B_t - B_{t-1}R_t + \tau_d B_{t-1}\left(R_t - 1\right) + \mathcal{T}_t \end{aligned} \tag{33}$$

where S_{it} and Q_{it} denote the quantity and price of claims held on firm i, $\tilde{\Pi}_{it} = \Pi_{it} - R_t^K K_{it+1}$ firm i operating profits less capital costs, $\tilde{P}_t^K$ the cum-dividend price of capital (inclusive of the rental rate) as defined in Section 4, B_t intermediary debt, R_t the (real) risk-free rate, τ the tax rate, τ_d the fraction of interest payments that are deductible and $\mathcal{T}_t$ lump-sum rebates.

In equilibrium, the intermediary owns all claims on firms and the entire capital stock ($S_{it} = 1, K_t = \overline{K}$), which together with the fact that all taxes are rebated lump-sum, implies $CF_t = \Pi_t + B_t - B_{t-1}R_t$, where $\Pi_t = \int \Pi_{it} di$ equals intermediary operating (pre-interest income/expense) earnings.

The intermediary faces two frictions. First, there is a convex cost of issuing dividends such that the value of cash flows net of this cost is equal to $\Phi\left(CF_t\right)$ where $\Phi'\left(CF_t\right) > 0$ and $\Phi''\left(CF_t\right) < 0$. In practice, we assume this cost is quadratic, i.e.,

$$\Phi\left(CF_t\right) = CF_t - \frac{\theta}{2}\left(\frac{CF_t}{\overline{CF}} - 1\right)^2 \overline{CF} \tag{34}$$

where $\overline{CF}$ denotes cash flows in the steady state. The cost is purely financial and is paid to households.

Second, the intermediary faces a constraint limiting its debt as a proportion of its operating earnings, given by:

$$B_t \leq \xi \Pi_t \tag{35}$$

Intermediary problem. The intermediary value function satisfies

$$V_t = \max_{S_{it}, B_t} \; \Phi\left(CF_t\right) + \mathbb{E}_t\left[\Lambda_{t+1} V_{t+1}\right]$$

subject to (33), (34) and (35).

The optimality condition on asset holdings (S_{it}) is given by

$$Q_{it} = \mathbb{E}_t\left[\Lambda_{T+1}\frac{\Phi'\left(CF_{t+1}\right)}{\Phi'\left(CF_t\right)}\left(\tilde{\Pi}_{it+1} + Q_{it+1}\right)\right] \tag{36}$$

where $\Phi'(CF_t) = 1 - \theta\left(\frac{CF_t}{\overline{CF}} - 1\right)$.

Similarly, the optimality condition on debt (B_t) is given by

$$\Phi'(CF_t) = \mathbb{E}_t\left[\Lambda_{t+1}\Phi'(CF_{t+1})\left(R_{t+1} - \tau_d\left(R_{t+1} - 1\right)\right)\right] + \zeta_t \tag{37}$$

where ζ_t is the Lagrange multiplier on the constraint (35).

Firm problem. The firm solves

$$\max_{K_{it}} \; -R^K_{t-1}K_{it} + Q_{it-1}$$

where Q_{it} satisfies (36). The optimality condition yields

$$\mathbb{E}_{t-1}\left[\Lambda_t \frac{\Phi'(CF_t)}{\Phi'(CF_{t-1})} MRPK_{it}\right] = R^K_{t-1}$$

which is expression (8) with $T_{\Lambda t} = \frac{\Phi'(CF_t)}{\Phi'(CF_{t-1})}$.

Linear distortion. Log-linearizing $T_{\Lambda t}$,

$$\tau_{\Lambda t} \approx \frac{\Phi''\left(\overline{CF}\right)\overline{CF}}{\Phi'\left(\overline{CF}\right)}\left(cf_t - cf_{t-1}\right) = -\theta\Delta cf_t \tag{38}$$

where the second equality follows from (34).

To solve for the behavior of intermediary cash flows, first note that in the steady state, (37) satisfies $\overline{\zeta} = \Phi'\left(\overline{CF}\right)\tau_d\left(1 - \frac{1}{\overline{R}}\right) > 0$, which shows that the constraint binds in the steady state. Further, under mild regularity conditions, the constraint will bind in a neighborhood around the steady state, i.e., $B_t = \xi\Pi_t$. Using this, intermediary cash flows satisfy

$$CF_t = (1+\xi)\,\Pi_t - \xi\Pi_{t-1}R_t$$

Log-linearizing,

$$cf_t \approx \frac{1+\xi}{1+\xi-\xi\overline{R}}\pi_t - \frac{\xi\overline{R}}{1+\xi-\xi\overline{R}}\pi_{t-1} - \frac{\xi\overline{R}}{1+\xi-\xi\overline{R}}r_t$$

so that combining with (38) and using the fact that in log deviations $\pi_t = y_t$,

$$\tau_{\Lambda t} \;\; = \;\; -\tau_{\Lambda y}\Delta y_t + (\theta - \tau_{\Lambda y})\left(\Delta y_{t-1} + \Delta r_t\right), \quad \tau_{\Lambda y} = \frac{\theta}{1 - \frac{\xi}{1+\xi}\overline{R}} \tag{39}$$

Taking expectations yields expression (9). Note that $1 + \xi - \xi\overline{R} > 0 \;\; \Rightarrow \;\; \tau_{\Lambda y} > 0$ is a necessary condition for steady state cash flows to be non-negative.

A.2 Capital distortions – Additional Detailed Examples

Limited asset market participation. 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). 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. Aggregate consumption is the sum of consumption across constrained and unconstrained:

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

where $G_t = \frac{C_t^U - C_t^K}{C_t^U}$ is defined as the gap between the consumption of unconstrained and constrained households. A first-order approximation yields

$$c_t^U = c_t + \frac{\theta}{1-\theta\overline{G}} g_t \tag{40}$$

where $g_t = G_t - \overline{G}$ is the deviation of the consumption gap from its steady-state value, $\overline{G}$.

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 (common) price markup shocks denoted T_t^p, which leads to variation in the share of output paid to labor vs. capital and profits.[58] They 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 approximating and substituting into (40) gives

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

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

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

Using this SDF to price assets yields (8), where the distortion, $T_{\Lambda t}$, reflects the proportion of constrained households and time-variation in their share of aggregate income.[59]

[58] 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.

[59] 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.

Externalities in preferences. Assume the period utility function over consumption is given by

$$\frac{\left(C_{jt}/C_{t-1}^{\lambda}\right)^{1-\gamma}}{1-\gamma}$$

a la Abel (1990), where C_{jt} is own consumption of individual j and C_t aggregate consumption.

The SDF for individual j is

$$\tilde{\Lambda}_{jt} = \rho\left(\frac{C_{jt}}{C_{jt-1}}\right)^{-\gamma}\left(\frac{C_{t-1}}{C_{t-2}}\right)^{-\lambda(1-\gamma)}$$

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

$$\tilde{\Lambda}_{t} = \rho\left(\frac{C_{t}}{C_{t-1}}\right)^{-\gamma}\left(\frac{C_{t-1}}{C_{t-2}}\right)^{-\lambda(1-\gamma)}$$

Instead, for a planner who internalizes the consumption externality, the SDF can be derived as

$$\Lambda_t = \rho\frac{1-\lambda\rho\mathbb{E}_t\left[\left(\frac{C_{t+1}}{C_t}\right)^{1-\gamma}\right]\left(\frac{C_t}{C_{t-1}}\right)^{-\lambda(1-\gamma)}}{1-\lambda\rho\left(\frac{C_t}{C_{t-1}}\right)^{1-\gamma}\left(\frac{C_{t-1}}{C_{t-2}}\right)^{-\lambda(1-\gamma)}}\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma}\left(\frac{C_{t-1}}{C_{t-2}}\right)^{-\lambda(1-\gamma)}$$

and hence

$$\tilde{\Lambda}_t = T_{\Lambda t}\Lambda_t \quad \text{where} \quad T_{\Lambda t} = \frac{1-\lambda\rho\left(\frac{C_t}{C_{t-1}}\right)^{1-\gamma}\left(\frac{C_{t-1}}{C_{t-2}}\right)^{-\lambda(1-\gamma)}}{1-\lambda\rho\mathbb{E}_t\left[\left(\frac{C_{t+1}}{C_t}\right)^{1-\gamma}\right]\left(\frac{C_t}{C_{t-1}}\right)^{-\lambda(1-\gamma)}}$$

Here, the wedge depends on the dynamics of consumption and the extent of the externality, λ.

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

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

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 (log) bias is positive for ε_t small and negative for ε_t large, 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

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

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

A.3 General Distortion

Take any distortion of the form

$$\tau_{\Lambda t} = \mathbb{E}_{t-1}\left[\tau_{\Lambda t}\right] - \tau_{\Lambda a}\varepsilon_t \tag{41}$$

The first term is fully general, i.e., the realized wedge depends on its conditional expectation. The only restriction is the second term, which requires that the wedge can be expressed (or suitably approximated) as a log-linear function of the exogenous shock, ε_t.[60] The loading on the shock, however, is fully general, i.e., $\tau_{\Lambda a}$ can be any function of exogenous parameters and/or endogenous model objects, including the policy rule.

In this more general case, all the equilibrium conditions from Section 2 go through unchanged with

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

and we can derive the following proposition (proof below), which generalizes Propositions 4 and 5:

Proposition 7. *(i) If* $\frac{\partial \tau_{\Lambda a}}{\partial \mu_a} > -\kappa_l$*, then the more countercyclical is monetary policy, i.e., the smaller is* μ_a*, the lower is the price of risk, i.e.,* $\frac{\partial \kappa}{\partial \mu_a} > 0$*, and the higher are the level and volatility of TFP, i.e.,* $\frac{\partial \psi_a}{\partial \mu_a} < 0$*, and, if* $\kappa > \underline{\kappa}$ *where* $\underline{\kappa}$ *is a positive constant, then* $\frac{\partial \overline{\psi}}{\partial \mu_a} < 0$*.*

(ii) Optimal monetary policy satisfies

$$\mu_a^* = -\tau_{\Lambda a}^* \Omega \frac{\kappa_l + \frac{\partial \tau_{\Lambda a}^*}{\partial \mu_a^*}}{\left(\kappa_l + \frac{\partial \tau_{\Lambda a}^*}{\partial \mu_a^*}\right)\kappa_l \Omega + \frac{1}{\phi_l} + \frac{\alpha_2 \nu_w \lambda_w}{\phi_l}}, \tag{42}$$

where asterisks denote equilibrium values under the optimal policy and $\Omega > 0$ *is a composite of model parameters defined in Appendix B.2.*

Part (i) of the proposition extends the results in Proposition 4 and shows that the equilibrium effects of policy hold more generally, beyond our particular micro-foundation for the financial friction.[61] Part (ii) of the proposition similarly generalizes Proposition 5 to the case of an arbitrary (log-linear) distortion and shows that the cyclicality of optimal policy is increasing (in absolute value) in the sensitivity of the distortion to exogenous shocks, $\tau_{\Lambda a}$, and the sensitivity of the distortion to the monetary policy rule itself, $\frac{\partial \tau_{\Lambda a}}{\partial \mu_a}$ (both evaluated at their equilibrium values under the optimal policy).

The distortion in Section 2 is nested in this more general formulation as follows: using (9), (10) and (64), the distortion satisfies (41) with

$$\begin{aligned} \tau_{\Lambda a} &= \tau_{\Lambda y}\left(\phi_\psi \psi_a + \mu_a\right) \\ &= \tau_{\Lambda y} \frac{\phi_\psi + \left(1 + \left(\kappa_\psi - \phi_\psi \kappa_l\right)\omega \Upsilon\right)\mu_a}{1 + \left(\kappa_\psi + \phi_\psi \tau_{\Lambda y}\right)\omega \Upsilon} \end{aligned}$$

Since $\tau_{\Lambda y} > 0$ and $\kappa_\psi - \phi_\psi \kappa_l = \frac{\nu}{1-\alpha_2 \nu} > 0$, we clearly have $\frac{\partial \tau_{\Lambda a}}{\partial \mu_a} > 0$ and $\frac{\partial \kappa}{\partial \mu_a} > 0$. Note that the distortion can be cleanly decomposed into two components: one that is independent of policy and depends only on the direct effect of exogenous shocks (terms that are independent of μ_a), and one that is endogenous to policy (terms that involve μ_a).

[60] If the wedge is any arbitrary non-linear function of the shock, our exact aggregation results, e.g., for TFP, may not hold. However, approximate aggregation may.

[61] The condition that $\frac{\partial \tau_{\Lambda a}}{\partial \mu_a} > -\kappa_l$ is mild and simply requires that a more *countercyclical* monetary policy rule does not increase the *procyclicality* of the distortion and hence the price of risk to too great an extent; the condition is easily satisfied in both our analytical model in Section 2 and quantitative model in Section 4, where $\frac{\partial \tau_{\Lambda a}}{\partial \mu_a}$ and $\frac{\partial \kappa}{\partial \mu_a}$ are both positive.

Proof of Proposition 7. (i) Assuming $\frac{\partial \tau_{\Lambda a}}{\partial \mu_a} + \kappa_l > 0$, generalizing the proof of Proposition 4 in Appendix A.8 yields:

$$\begin{aligned}
\frac{\partial \kappa}{\partial \mu_a} &= \frac{\kappa_l + \frac{\partial \tau_{\Lambda a}}{\partial \mu_a}}{1 + \kappa_\psi \omega \Upsilon} > 0 \\
\frac{\partial \psi_a}{\partial \mu_a} &= -\frac{\left(\kappa_l + \frac{\partial \tau_{\Lambda a}}{\partial \mu_a}\right) \omega \Upsilon}{1 + \kappa_\psi \omega \Upsilon} < 0 \\
\frac{\partial \overline{\psi}}{\partial \mu_a} &= \frac{\kappa_l + \frac{\partial \tau_{\Lambda a}}{\partial \mu_a}}{1 + \kappa_\psi \omega \Upsilon} \left(\omega \frac{\nu}{1 - \alpha_2 \nu} \sigma_\varepsilon^2 - \alpha_1 \kappa \sigma_\beta^2 \left(\frac{1}{1-\alpha} \frac{\sigma_\varepsilon^2}{1 + \sigma_\varepsilon^2 \sigma_\beta^2} \right)^2 \left(1 - \alpha + \sigma_\varepsilon^2 \sigma_\beta^2\right) \right)
\end{aligned}$$

where the last expression is negative if $\kappa > \underline{\kappa}$ where $\underline{\kappa}$ is a positive constant.

(ii) Following similar steps as the proof of Proposition 5 in Appendix B.2, we obtain the following generalization of (76):

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

and solving for μ_a yields (42).

□

A.4 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 (12) and derives exact expressions for the terms κ, ψ_a and $\overline{\psi}$.

Proof of Proposition 1. To derive expression (11), aggregate the labor demand condition (6) to obtain

$$\alpha_2 \nu Y_t = W_t L_t \tag{43}$$

which implies $\frac{L_{it}}{L_t} = \left(\frac{Y_{it}}{Y_t}\right)^\nu$ and substituting into the firm-level production function and then into (4) and rearranging yields the first two equations in (11).

To derive the third expression in (11), integrate (8) and impose capital market clearing.

To derive (10), combine (43) with the labor supply condition (2) and resource constraint to obtain

$$L_t = \left(\frac{\alpha_2 \nu}{\chi}\right)^{\frac{1}{1+\varphi}} \mathcal{M}_t^{\frac{1}{1+\varphi}} Y_t^{\frac{1-\gamma}{1+\varphi}} \tag{44}$$

Substituting into the aggregate production function in (11) gives

$$Y_t = \left(\frac{\alpha_2 \nu}{\chi}\right)^{\phi_l} \Psi_t^{\phi_\psi} \mathcal{M}_t^{\phi_l} K_t^{\alpha_1 \phi_\psi} \tag{45}$$

where

$$\phi_\psi = \frac{1}{1 - \alpha_2 \frac{1-\gamma}{1+\varphi}}, \qquad \phi_l = \frac{\alpha_2}{1+\varphi} \frac{1}{1 - \alpha_2 \frac{1-\gamma}{1+\varphi}}$$

Taking logs (and suppressing constants) yields (10).

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

hold problem is standard. The household solves:

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

$$s.t.$$

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

$$P_tC_{jt}+B^n_{jt}+B_{jt}=(1-\tau_{nt})\,W^n_{jt}L_{jt}+(1+i_t)\,B^n_{jt-1}+R_tB_{jt-1}+\Pi_{jt}+\mathcal{T}_{jt}$$

where P_t denotes the price of the final good, W^n_{jt} the nominal wage, B^n_{jt} holdings of nominally risk-free bonds (in zero net supply), B_{jt} holdings of real risk-free bonds issued by financial intermediaries, and Π_{jt} the household's share of equity payouts from financial intermediaries, which are composed of the return on rented capital plus economic profits (which sum to operating profits Π_t) less the cost of investment in new capital and (net) debt payments, i.e., $\int\Pi_{jt}dj=\Pi_t-P_t^KI_t-(R_tB_{t-1}-B_t)$ where $I_t=K_{t+1}-(1-\delta)\,K_t$. In the analytic model of Section 2, $\delta=0$, $K_t=\overline{K}$ and hence $I_t=0$. $\mathcal{T}_{jt}$ denotes lump-sum taxes/transfers. 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:

$$0=\nu_w\left(\frac{MRS_t}{W_t}-1\right)-\theta_w\left(\Pi^w_t-1\right)\frac{Y_t}{L_t}\frac{P_t}{P_{t-1}}\frac{1}{W_{t-1}}+\mathbb{E}_t\left[\Lambda_{t+1}\theta_w\left(\Pi^w_{t+1}-1\right)\Pi^w_{t+1}\frac{Y_{t+1}}{L_tW_t}\right] \tag{46}$$

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 $MRS_t\equiv\chi L_t^{\varphi}C_t^{\gamma}$. 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 for nominal bonds is given by:

$$1=(1+i_t)\,\mathbb{E}_t\left[\Lambda_{t+1}\frac{1}{\Pi^p_{t+1}}\right] \tag{47}$$

where i_t denotes the nominal interest rate and $\Pi^p_{t+1}=\Pi^w_t\frac{W_{t-1}}{W_t}$ gross price inflation, and for real bonds

$$1=R_t\mathbb{E}_t\left[\Lambda_{t+1}\right] \tag{48}$$

□

Proof of Proposition 2. We prove that TFP takes the form in expression (32), which nests (14) in the case of i.i.d. shocks, i.e., when $\rho_a=0$.

We assume firm-level productivity takes the form in (12), where $\mathcal{O}_{it}$ is defined as

$$\mathcal{O}_{it}=\xi_0+\xi_\beta\hat{\beta}_i^2+\xi_{\varepsilon^2}\varepsilon_t^2+\xi_{a_{-1}^2}\mathbb{E}_{t-1}\left[a_t\right]^2+\xi_{a_{-1},\varepsilon}\mathbb{E}_{t-1}\left[a_t\right]\varepsilon_t+\xi_{a_{-1}}\mathbb{E}_{t-1}\left[a_t\right] \tag{49}$$

with

$$\xi_0=\frac{\alpha_1}{2}\log\left(1+\sigma_\beta^2\sigma_\varepsilon^2\right)+\alpha_1\left(\frac{\nu}{1-\alpha_2\nu}\right)^2\frac{\sigma_\varepsilon^2}{2},\qquad \xi_\beta=-\alpha_1\frac{\left(\frac{\nu}{1-\alpha_2\nu}\right)^2\frac{\sigma_\varepsilon^2}{2}}{1+\sigma_\beta^2\sigma_\varepsilon^2} \tag{50}$$

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

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

We assume the following form for the output gap:

$$\mu_t = \mu_a \varepsilon_t + \mathbb{E}_{t-1}\left[\mu_t\right] \tag{51}$$

which is a more general version of (13) that encompasses all classes of monetary policy studied in the paper, i.e., the simple rule in (13) (extended to (78) with persistent shocks), the more complicated rule in (26) that we work with in Section 4 (with constant coefficients and no monetary policy shocks), and the optimal policy we solve for in that section.

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

$$\psi_t = \mathbb{E}_{t-1}\left[\psi_t\right] + \psi_a \varepsilon_t \tag{52}$$

Substituting the form of the profit function in (7) into (8) and rearranging yields the following expression for the capital choice

$$K_{it} = \left(\alpha G \left(R^K_{t-1}\right)^{-1} \mathbb{E}_{t-1}\left[\Lambda_t T_{\Lambda t} A_{it}^{\frac{\nu}{1-\alpha_2\nu}} Y_t^{\frac{1-\nu}{1-\alpha_2\nu}} W_t^{-\frac{\alpha_2\nu}{1-\alpha_2\nu}}\right]\right)^{\frac{1}{1-\alpha}} \tag{53}$$

and relative capital stocks:

$$\frac{K_{it}}{K_t} = \frac{\left(\mathbb{E}_{t-1}\left[\Lambda_t T_{\Lambda t} A_{it}^{\frac{\nu}{1-\alpha_2\nu}} Y_t^{\frac{1-\nu}{1-\alpha_2\nu}} W_t^{-\frac{\alpha_2\nu}{1-\alpha_2\nu}}\right]\right)^{\frac{1}{1-\alpha}}}{\int \left(\mathbb{E}_{t-1}\left[\Lambda_t T_{\Lambda t} A_{it}^{\frac{\nu}{1-\alpha_2\nu}} Y_t^{\frac{1-\nu}{1-\alpha_2\nu}} W_t^{-\frac{\alpha_2\nu}{1-\alpha_2\nu}}\right]\right)^{\frac{1}{1-\alpha}} di} \tag{54}$$

Next, combine (43) with the labor supply condition (2) and resource constraint to obtain

$$W_t = (\alpha_2\nu)^{1-\frac{1}{1+\varphi}} \chi^{\frac{1}{1+\varphi}} Y_t^{1-\frac{1-\gamma}{1+\varphi}} \mathcal{M}_t^{-\frac{1}{1+\varphi}} \tag{55}$$

Substituting for Y_t and W_t using (45) and (55) and rearranging, write the expectation in (53) as

$$\mathbb{E}_{t-1}\left[\Lambda_t T_{\Lambda t} A_{it}^{\frac{\nu}{1-\alpha_2\nu}} Y_t^{\frac{1-\nu}{1-\alpha_2\nu}} W_t^{-\frac{\alpha_2\nu}{1-\alpha_2\nu}}\right] = \text{Const.} \times \mathbb{E}_{t-1}\left[T_{\Lambda t} A_{it}^{\frac{\nu}{1-\alpha_2\nu}} \mathcal{M}_t^{-\kappa_l\phi_l} \Psi_t^{-\kappa_\psi}\right] \tag{56}$$

where Const. captures time-varying terms that are common across firms and known at time $t-1$ and

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

We provide further intuition for these terms in Appendix A.7.

Using (39), (51), (52) and the definition of μ_t in (10), we can substitute to write

$$\mathbb{E}_{t-1}\left[T_{\Lambda t} A_{it}^{\frac{\nu}{1-\alpha_2\nu}} \mathcal{M}_t^{-\kappa_l\phi_l} \Psi_t^{-\kappa_\psi}\right] = \text{Const.} \times \mathbb{E}_{t-1}\left[A_{it}^{\frac{\nu}{1-\alpha_2\nu}} e^{-\kappa\varepsilon_t}\right] \tag{57}$$

where κ is as defined in equation (21). Substituting into (54),

$$\frac{K_{it}}{K_t} = \frac{\left(\mathbb{E}_{t-1}\left[A_{it}^{\frac{\nu}{1-\alpha_2\nu}} e^{-\kappa\varepsilon_t}\right]\right)^{\frac{1}{1-\alpha}}}{\int \left(\mathbb{E}_{t-1}\left[A_{it}^{\frac{\nu}{1-\alpha_2\nu}} e^{-\kappa\varepsilon_t}\right]\right)^{\frac{1}{1-\alpha}} di} \tag{58}$$

and into the expression for TFP in (11), we obtain:

$$\Psi_t = \frac{\left(\int \left(A_{it}\left(\mathbb{E}_{t-1}\left[A_{it}^{\frac{\nu}{1-\alpha_2\nu}}e^{-\kappa\varepsilon_t}\right]\right)^{\alpha_1\frac{1-\alpha_2\nu}{1-(\alpha_1+\alpha_2)\nu}}\right)^{\frac{\nu}{1-\alpha_2\nu}}di\right)^{\frac{1-\alpha_2\nu}{\nu}}}{\left(\int\left(\mathbb{E}_{t-1}\left[A_{it}^{\frac{\nu}{1-\alpha_2\nu}}e^{-\kappa\varepsilon_t}\right]\right)^{\frac{1-\alpha_2\nu}{1-(\alpha_1+\alpha_2)\nu}}di\right)^{\alpha_1}} \tag{59}$$

Substituting for a_{it} from (12) and (49), we can write

$$\begin{aligned}\mathbb{E}_{t-1}\left[A_{it}^{\frac{\nu}{1-\alpha_2\nu}}e^{-\kappa\varepsilon_t}\right] &= e^{\frac{\nu}{1-\alpha_2\nu}\left(\hat{\beta}_i\mathbb{E}_{t-1}[a_t]+\xi_0+\xi_\beta\beta_i^2+\xi_{a_{-1}^2}\mathbb{E}_{t-1}[a_t]^2+\xi_{a_{-1}}\mathbb{E}_{t-1}[a_t]\right)}\\ &\times \mathbb{E}_{t-1}\left[e^{\varepsilon_t\left(\frac{\nu}{1-\alpha_2\nu}\left(\hat{\beta}_i+\xi_{a_{-1},\varepsilon}\mathbb{E}_{t-1}[a_t]\right)-\kappa\right)+\frac{\nu}{1-\alpha_2\nu}\xi_{\varepsilon^2}\varepsilon_t^2}\right]\end{aligned} \tag{60}$$

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

$$\mathbb{E}\left[e^{ax+bx^2}\right] = \frac{1}{\sqrt{1-2b\sigma_x^2}}e^{\frac{a\bar{x}+\frac{a^2}{2}\sigma_x^2+b\bar{x}^2}{1-2b\sigma_x^2}} \tag{61}$$

Therefore, the expectational term in (60) can be written as:

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

and substituting into (59), together with (60) and the definitions of A_{it}, ξ_β, ξ_{ε^2} and $\xi_{a_{-1},\varepsilon}$, some lengthy algebra yields:

$$\log\Psi_t = \log A_t^h + \mathcal{I}_n - \mathcal{I}_d \tag{62}$$

where

$$\begin{aligned}A_t^h &= e^{\xi_0+\xi_{\varepsilon^2}\varepsilon_t^2+\xi_{a_{-1}^2}\mathbb{E}_{t-1}[a_t]^2+\xi_{a_{-1}\varepsilon}\mathbb{E}_{t-1}[a_t]\varepsilon_t+\xi_{a_{-1}}\mathbb{E}_{t-1}[a_t]}\\ \mathcal{I}_n &= \frac{1-\alpha_2\nu}{\nu}\log\left(\int e^{\hat{\beta}_i\frac{\nu}{1-\alpha_2\nu}\left(\varepsilon_t+\mathbb{E}_{t-1}[a_t]\left(\frac{1-\alpha_2\nu}{1-(\alpha_1+\alpha_2)\nu}\right)-\frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\kappa\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}-\frac{\alpha_1\nu}{1-(\alpha_1+\alpha_2)\nu}\frac{\mathbb{E}_{t-1}[a_t]\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\right)}di\right)\\ \mathcal{I}_d &= \alpha_1\log\left(\int e^{\hat{\beta}_i\frac{\nu}{1-(\alpha_1+\alpha_2)\nu}\left(\mathbb{E}_{t-1}[a_t]-\frac{\kappa\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}-\frac{\mathbb{E}_{t-1}[a_t]\sigma_\varepsilon^2\sigma_\beta^2}{\frac{(1-\nu(\alpha_1+\alpha_2))}{(1-\alpha_2\nu)}+\sigma_\varepsilon^2\sigma_\beta^2}\right)+\frac{1}{2}\frac{\left(\frac{\nu}{1-\alpha_2\nu}\right)^2\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}\hat{\beta}_i^2}di\right)\end{aligned}$$

A_t^h consists of terms that are common across firms. Both $\mathcal{I}_n$ and $\mathcal{I}_d$ are Gaussian integrals over $\hat{\beta}_i$, which is normally distributed, and so we can apply result (61) to obtain an expression 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

$$\psi_t = \overline{\psi} + \psi_a\varepsilon_t + \mathbb{E}_{t-1}\left[a_t\right] \tag{63}$$

where

$$\psi_a = 1 - \frac{\alpha_1 (1 - \alpha_2 \nu)}{1 - (\alpha_1 + \alpha_2) \nu} \frac{\kappa \sigma_\varepsilon^2 \sigma_\beta^2}{1 + \sigma_\varepsilon^2 \sigma_\beta^2} \tag{64}$$

$$\overline{\psi} = \kappa \sigma_\varepsilon^2 \frac{\alpha_1 \nu}{1 - (\alpha_1 + \alpha_2) \nu} \frac{\sigma_\varepsilon^2 \sigma_\beta^2}{1 + \sigma_\varepsilon^2 \sigma_\beta^2} - \frac{\sigma_\beta^2}{2} \kappa^2 \left(\frac{1}{1 - \alpha} \frac{\sigma_\varepsilon^2}{1 + \sigma_\varepsilon^2 \sigma_\beta^2} \right)^2 \alpha_1 \left(1 - \alpha + \sigma_\varepsilon^2 \sigma_\beta^2 \right) \tag{65}$$

which verifies our conjecture in (52) and yields (14) in the i.i.d. case and (32) when a_t follows an AR(1) process with persistence ρ_a. From here, we can substitute for κ from (21) to solve for

$$\psi_a = \frac{1 - \mu_a (\kappa_l + \tau_{\Lambda y}) \omega \Upsilon}{1 + (\kappa_\psi + \phi_\psi \tau_{\Lambda y}) \omega \Upsilon} \tag{66}$$

and

$$\kappa = \frac{\kappa_\psi + \phi_\psi \tau_{\Lambda y} + (\kappa_l + \tau_{\Lambda y}) \mu_a}{1 + (\kappa_\psi + \phi_\psi \tau_{\Lambda y}) \omega \Upsilon} \tag{67}$$

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 $\Upsilon \equiv \frac{\sigma_\varepsilon^2 \sigma_\beta^2}{1 + \sigma_\varepsilon^2 \sigma_\beta^2}$.

Using (62) and (50), it is straightforward to verify that $\psi_{a_{-1},\varepsilon} = \psi_{\varepsilon^2} = \psi_{a_{-1}^2} = 0$ and $\psi_{a_{-1}} = 1$ so that (63) holds and to derive (64) and (65).

The proof of Proposition 2 is completed with the linearized Phillips curve and Euler equation. Log-linearizing (46) yields (15), 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 (47) and combining with the resource constraint, $c_t = y_t$, yields (16).

□

A.5 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.

Using the log-linearized Phillips curve (15) along with (13), we can guess and verify that

$$\pi_t^w = \zeta_a^w \varepsilon_t, \quad \zeta_a^w = \lambda_w \mu_a \tag{68}$$

and combining (10), (14) and (55), in deviations,

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

By definition

$$\pi_{t+1}^p = w_t - w_{t+1} + \pi_{t+1}^w = w_a (\varepsilon_t - \varepsilon_{t+1}) + \zeta_a^w \varepsilon_{t+1} \quad \Rightarrow \quad \mathbb{E}_t \left[\pi_{t+1}^p \right] = w_a \varepsilon_t$$

and substituting into the Euler equation (16):

$$i_t = \left(\frac{\varphi (1 - \gamma)}{1 + \varphi} \phi_\psi \psi_a - \frac{1 - \alpha_2 (1 - \gamma)}{\alpha_2} \mu_a \right) \varepsilon_t = i_a \varepsilon_t$$

Last, substitute for ψ_a from (66) to obtain

$$i_a = \frac{\varphi (1 - \gamma)}{1 + \varphi} \frac{\phi_\psi}{1 + (\kappa_\psi + \phi_\psi \tau_{\Lambda y}) \omega \Upsilon} - \left(\frac{\varphi (1 - \gamma)}{1 + \varphi} \frac{\phi_\psi (\kappa_l + \tau_{\Lambda y}) \omega \Upsilon}{1 + (\kappa_\psi + \phi_\psi \tau_{\Lambda y}) \omega \Upsilon} + \frac{1 - \alpha_2 (1 - \gamma)}{\alpha_2} \right) \mu_a \tag{69}$$

$$= \Phi^i - \Phi_\mu^i \mu_a \tag{70}$$

which shows that any μ_a in (13) can be implemented by an interest rate rule, i_a, that satisfies (69).

It is straightforward to verify that $\Phi^i_\mu > 0$ and therefore 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, e.g., rules in the output gap and expected price inflation or in 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, (69) 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

$$\begin{aligned} i_t &= i_a^{TR} \varepsilon_t \qquad \text{where} \\ i_a^{TR} &= \phi_\pi \left(1 - \frac{1-\gamma}{1+\varphi}\right) \frac{\phi_\psi}{1 + (\kappa_\psi + \phi_\psi \tau_{\Lambda y})\,\omega \Upsilon} \\ &+ \left(\phi_y - \phi_\pi \left(\left(1 - \frac{1-\gamma}{1+\varphi}\right) \frac{\phi_\psi\,(\kappa_l + \tau_{\Lambda y})\,\omega \Upsilon}{1 + (\kappa_\psi + \phi_\psi \tau_{\Lambda y})\,\omega \Upsilon} + \frac{1-\alpha_2}{\alpha_2}\right)\right) \mu_a \end{aligned}$$

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

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

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. Next, consider a rule in 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

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

A.6 Approximate Solution

Here we derive the approximate expressions in Section 2.1. For purposes of the approximation, assume $\mathcal{O}_{it} = 0$ in (12) so that $a_{it} = \hat{\beta}_i a_t$, which is log-normally distributed. To derive (17), we can approximate (11) around the ergodic mean where $K_{it} = K_i$ and $a_t = 0$ to obtain

$$\psi_t = \frac{1-\alpha_2\nu}{\nu} \log\left(\int \left(\frac{K_{it}}{\overline{K}}\right)^\alpha di\right) + \int \left(\frac{K_i}{\overline{K}}\right)^\alpha \hat{\beta}_i di a_t = \overline{\psi} + \psi_a a_t \tag{71}$$

At the ergodic mean, relative capital satisfies

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

and substituting,

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

Taking a second-order approximation to the integral terms gives the expression for $\overline{\psi}$ in (17).

Turning to the second term in (71), we have

$$\psi_a = \int\left(\frac{K_i}{\overline{K}}\right)^{\alpha}\hat{\beta}_i di \approx \int\left(1+\alpha\left(k_i-\overline{k}\right)\right)\hat{\beta}_i di = 1+\alpha\text{cov}\left(k_i,\hat{\beta}_i\right) = 1+\alpha_1\text{cov}\left(k_i,\beta_i\right)$$

To derive expression (18), write (8) as

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

Taking a second-order approximation and suppressing constants yields the result.

Proof of Proposition 3. To derive (19), use expression (58) to write

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

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 (7) and (8),

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

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,

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

and substituting into (17) yields (20).

The derivation of (21) is the same as in expression (57), which shows that that expression is exact. □

A.7 Interpretation of κ

To gain intuition for κ, consider a second-order approximation to the optimality condition in (53):

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

where $\tilde{\lambda}_t = \lambda_t + \tau_{\Lambda t}$ and we have suppressed constants. Comparing to (19), we can see

$$\kappa = -\frac{\partial}{\partial \varepsilon_t}\left(\tilde{\lambda}_t + \frac{1-\nu}{1-\alpha_2\nu} y_t - \frac{\alpha_2\nu}{1-\alpha_2\nu} w_t\right) \tag{72}$$

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 (56) and (57) show that using the equilibrium conditions to substitute, we obtain (21):

$$\kappa = \kappa_\psi \psi_a + \kappa_l \mu_a + \tau_{\Lambda y}\left(\phi_\psi \psi_a + \mu_a\right)$$

where

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

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, 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 (negative of 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 output gap on profitability through changes in output. Multiplying through by μ_a translates this term into an elasticity with respect to ε_t.

Third, through its effects on TFP and the output gap and hence aggregate output, ε_t affects the capital wedge; $\tau_{\Lambda y}$ captures the elasticity of the wedge to output and the term in parentheses, $\phi_\psi \psi_a + \mu_a$, the elasticity of output to the shock.

A.8 The Effects of Monetary Policy

Proof of Proposition 4. From (67), (66) and (65),

$$\begin{aligned}
\frac{\partial \kappa}{\partial \mu_a} &= \frac{\kappa_l + \tau_{\Lambda y}}{1 + \left(\kappa_\psi + \phi_\psi \tau_{\Lambda y}\right)\omega\Upsilon} > 0 \\
\frac{\partial \psi_a}{\partial \mu_a} &= -\frac{\left(\kappa_l + \tau_{\Lambda y}\right)\omega\Upsilon}{1 + \left(\kappa_\psi + \phi_\psi \tau_{\Lambda y}\right)\omega\Upsilon} < 0 \\
\frac{\partial \overline{\psi}}{\partial \mu_a} &= \frac{\kappa_l + \tau_{\Lambda y}}{1 + \left(\kappa_\psi + \phi_\psi \tau_{\Lambda y}\right)\omega\Upsilon}\left(\frac{\nu}{1-\alpha_2\nu}\sigma_\varepsilon^2 \omega\Upsilon - \alpha_1 \kappa \sigma_\beta^2\left(\frac{1}{1-\alpha}\frac{\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}\right)^2\left(1-\alpha+\sigma_\varepsilon^2\sigma_\beta^2\right)\right)
\end{aligned}$$

where the last expression is negative if $\kappa > \underline{\kappa}$ where $\underline{\kappa}$ is a positive constant.

□

It is straightforward to verify that $\frac{\partial \frac{\partial \kappa}{\partial \mu_a}}{\partial \tau_{\Lambda y}} > 0$, i.e., the effects of monetary policy are larger the more severe are financial frictions.

The effects of financial frictions. To prove the signs reported in the text, note that

$$\frac{\partial \sigma\left(y_t\right)}{\partial \tau_{\Lambda y}}=\phi_\psi \frac{\partial \psi_a}{\partial \tau_{\Lambda y}}=-\phi_\psi \omega \Upsilon \frac{\partial \kappa}{\partial \tau_{\Lambda y}}$$

and

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

Using (67), we can show $\frac{\partial \kappa}{\partial \tau_{\Lambda y}}>0$ iff $\mu_a>-\frac{\phi_\psi}{1+\frac{\nu}{1-\alpha_2 \nu} \omega \Upsilon}$, which can be shown to be equivalent to our assumption that $\frac{\partial \mu_t}{\partial a_t}>-\phi_\psi \frac{\partial \psi_t}{\partial a_t}$. Thus, $\frac{\partial \sigma\left(y_t\right)}{\partial \tau_{\Lambda y}}<0$ and $\frac{\partial \overline{\psi}}{\partial \mu_a}<0$ when $\kappa>\underline{\kappa}$ where $\underline{\kappa}$ is defined in the proof of Proposition 4.

B Welfare and Optimal Policy

This appendix derives the results in Section 2.3.

B.1 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 using the fact that the optimal time-invariant production subsidy, $\frac{1}{\nu}$, is in place, a second-order approximation around the ergodic mean (denoted subscript m) yields

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

Next, from (45), in the non-stochastic steady state (denoted with bars) and ergodic mean,

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

so that

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

and combining, we can write

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

Using (44) and (45) (in log deviations) along with the definition of λ_w, we can write

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

and since $\mathbb{E}_{-1}\left[\psi_t\right]=\overline{\psi}$, the negative of the expression is the period loss function in (24). To see that the last two terms correspond exactly to the welfare function in the textbook representative firm NK model, compare to Galí (2015), equation (26) of Chapter 6 and note that $\frac{1}{\phi_l}=\gamma+\frac{\varphi+1-\alpha_2}{\alpha_2}$.

B.2 Optimal Policy

In the i.i.d. case we have

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

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

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

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

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

Expression (68) gives π_t^w in terms of μ_t and using that expression,

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

Finally, substituting into (24),

$$\mathbb{W} = -\overline{\psi} + (\gamma-1)\,\phi_\psi\psi_a^2\frac{\sigma_\varepsilon^2}{2} + \frac{1}{\phi_l}\mu_a^2\frac{\sigma_\varepsilon^2}{2} + \frac{\alpha_2\nu_w\lambda_w}{\phi_l}\mu_a^2\frac{\sigma_\varepsilon^2}{2} \tag{73}$$

First-best allocation. The first-best (and constrained optimal) risk adjustment solves the problem of a planner who chooses κ to minimize (73), 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 (64) and (65), the derivatives of ψ_t with respect to κ are

$$\frac{\partial\psi_a}{\partial\kappa} = -\omega\Upsilon, \quad \frac{\partial\overline{\psi}}{\partial\kappa} = \omega\frac{\nu}{1-\alpha_2\nu}\sigma_\varepsilon^2\Upsilon - \alpha_1\kappa\left(\frac{1}{1-\alpha}\frac{\sigma_\varepsilon^2}{1+\sigma_\beta^2\sigma_\varepsilon^2}\right)^2\left(1-\alpha+\sigma_\beta^2\sigma_\varepsilon^2\right)\sigma_\beta^2 \tag{74}$$

Using these expressions (along with the definitions of ω, Υ and α) and equation (64), some lengthy but straightforward algebra yields

$$-\frac{\partial\overline{\psi}}{\partial\kappa} + (\gamma-1)\,\phi_\psi\psi_a\sigma_\varepsilon^2\frac{\partial\psi_a}{\partial\kappa} = -\omega\Upsilon\sigma_\varepsilon^2\left(\kappa_\psi - \kappa\left(1+\omega\Upsilon\kappa_\psi\right)\right) \tag{75}$$

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 5. Optimal policy within the class (13) is characterized by a value μ_a^* that minimizes (73), accounting for the effects on $\overline{\psi}$ and ψ_a. Taking the first order condition and substituting from (67), (74) and (75) yields

$$0 = \frac{(\kappa_l+\tau_{\Lambda y})\left(\phi_\psi\tau_{\Lambda y} + (\kappa_l+\tau_{\Lambda y})\left(1+\kappa_\psi\omega\Upsilon\right)\mu_a\right)\omega\Upsilon\sigma_\varepsilon^2}{\left(1+\left(\kappa_\psi+\phi_\psi\tau_{\Lambda y}\right)\omega\Upsilon\right)^2} + \frac{1}{\phi_l}\sigma_\varepsilon^2\mu_a + \frac{\alpha_2\nu_w\lambda_w}{\phi_l}\sigma_\varepsilon^2\mu_a \tag{76}$$

and rearranging,

$$\mu_a^* = -\tau_{\Lambda y}\sigma_\beta^2\Phi_\Lambda, \qquad \Phi_\Lambda = \frac{\frac{\phi_\psi\omega(\kappa_l+\tau_{\Lambda y})}{(1+(\kappa_\psi+\tau_{\Lambda y}\phi_\psi)\omega\Upsilon)^2}\frac{\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2\sigma_\beta^2}}{\frac{1}{\phi_l}+\frac{\alpha_2\nu_w\lambda_w}{\phi_l}+\left(\kappa_l+\tau_{\Lambda y}\right)^2\Omega} \tag{77}$$

where $\Omega = \frac{\left(1+\kappa_\psi\omega\Upsilon\right)\omega\Upsilon}{\left(1+\left(\kappa_\psi+\tau_{\Lambda y}\phi_\psi\right)\omega\Upsilon\right)^2}$. 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.

□

B.3 Optimal Policy – Extensions and Additional Results

Optimal nominal interest rate. Combining (25) and (70), we obtain an equation for the optimal nominal rate of the form $i_t = i_a a_t$, where

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

Substituting for Φ_Λ and Φ_μ^i from those expressions, we can show $\Phi_\Lambda^i > 0$ and $\frac{\partial i_a}{\partial\sigma_\beta^2} > 0$. From (70), Φ^i is independent of μ_a, and hence $\Phi^i a_t$ is the natural rate of interest that holds when the output gap is closed. Thus, we can write the optimal nominal rate as

$$i_t = r_t^n + i_a^* a_t, \quad i_a^* = \tau_{\Lambda y}\sigma_\beta^2\Phi_\Lambda^i$$

Persistent shocks. In the case of persistent shocks, i.e., $\rho_a \neq 0$, the equilibrium conditions in Proposition 2 remain unchanged with the exception of TFP, which takes the form in (32) and is a function of both a_{t-1} and ε_t. Accordingly, we assume that monetary policy follows

$$\mu_t = \mu_{a_{-1}} a_{t-1} + \mu_a\varepsilon_t \tag{78}$$

i.e., the central bank sets the output gap as a function of the two natural state variables in the economy. 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 $\pi_t^w = \frac{\lambda_w\mu_{a_{-1}}}{1-\rho\rho_a}a_{t-1} + \left(\frac{\rho\lambda_w\mu_{a_{-1}}}{1-\rho\rho_a}+\lambda_w\mu_a\right)\varepsilon_t$. Substituting into (24) and evaluating the summations gives the welfare loss as

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

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

For purposes of illustration, columns (1) and (2) of Table 3 numerically compare the optimal policy rule under i.i.d. and persistent shocks. To calibrate these versions of the model, we set the distortion, $\tau_{\Lambda y}$, to match the same risk adjustment, κ, as in the calibrated quantitative model in Section 4. This holds the level and volatility of TFP in the baseline equilibrium the same across the two cases (and equal to their values in the baseline equilibrium of the full quantitative model). We assume a simple Taylor rule in the output gap and expected price inflation, which leads to an output gap that satisfies (78). In the i.i.d. case we set $\rho_a = 0$. All other parameters are set to their calibrated parameters from the quantitative model. The table shows that adding persistent shocks has little effect on the optimal policy rule or welfare under the optimal rule: the welfare losses from a lower level of TFP

Table 3: Optimal Monetary Policy – Persistence and Commitment

	I.I.D. Shocks Optimal Rule (1)	Persistent Shocks Optimal Rule (2)	Persistent Shocks Full Commitment (3)
Welfare loss (%)			
Total	0.196	0.179	0.178
TFP level	0.016	0.017	0.016
TFP volatility	0.107	0.106	0.106
Output gap volatility	0.036	0.034	0.036
Inflation volatility	0.037	0.022	0.019
Cyclicality of policy			
μ_a	-2.462	-2.369	-2.372

are 0.016% in the i.i.d. case vs. 0.017% in the persistent case, and from TFP volatility 0.107% vs. 0.106% respectively. Total welfare losses are also similar, 0.196% vs. 0.179%, as is the nature of the optimal rule itself: the elasticity of the output gap to the exogenous shock, μ_a, is sizable and negative in both cases, -2.462 in the i.i.d. case vs. -2.369 with persistence. Intuitively, because the persistent component of the shock is known and common across firms, it does not affect conditional covariances and hence risk premia, and thus plays little role in changing the prescriptions for optimal policy.

Across the scenarios in Table 3, the gains from optimal policy are significant. For example, in results not reported in the table, in the empirically relevant case with persistent shocks, total welfare losses under the Taylor rule are about 1.6% (on par with, though somewhat larger than, the analogous value in the full quantitative model in Table 2 of about 1.4%), meaning that the optimal policies shown in the Table 3 increase welfare by roughly 1.4 p.p. (1.6-0.18). The losses in TFP under the Taylor rule are about 0.9% (by construction, equal to the value in the full quantitative model since the simple exercises here are calibrated to match the same κ), which falls to about 0.02% under the policies in Table 3, meaning that TFP increases by roughly 0.9 p.p. under these policies, which accounts for the lion's share –almost two-thirds – of the welfare gains from optimal policy.

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.,

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

The optimal policy consists of the full sequence of ζ_s^{μ}'s, which we analytically characterize below. Columns (2) and (3) of Table 3 numerically compare the optimal simple rule of the form (78) to the full commitment optimal policy. The calibration is the same as described immediately above for the case of persistent shocks and we take the case with persistence for purposes of illustration. The table shows that the optimal rule entails a very similar elasticity of the output gap to the exogenous shock, i.e., μ_a, as under full commitment, and welfare outcomes are also very similar across the two policies. In short, the optimal simple rule approximates the full commitment optimal policy well.

Characterization of full commitment optimal policy – Through the Phillips curve, the loadings of the output gap on shocks have a one-to-one mapping into loadings of (wage) inflation on the shocks, i.e., we can write $\pi_t^w = \sum_{s=0}^{\infty} \zeta_s^{\pi} \varepsilon_{t-s}$. Substituting this policy function and the Phillips curve into the welfare loss function (24) and evaluating the summations yields

$$\mathbb{W} = -\overline{\psi} + (\gamma - 1)\, \phi_\psi \psi_a^2 \frac{\sigma_\varepsilon^2}{2} + \frac{1}{\phi_l} \frac{\sigma_\varepsilon^2}{2} \sum_{s=0}^{\infty} \rho^s \left(\frac{\zeta_s^{\pi} - \rho \zeta_{s+1}^{\pi}}{\lambda_w} \right)^2 + \frac{\alpha_2 \nu_w}{\lambda_w \phi_l} \frac{\sigma_\varepsilon^2}{2} \sum_{s=0}^{\infty} \rho^s \left(\zeta_s^{\pi} \right)^2 \tag{79}$$

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 (75) and (67), we can write the first order conditions from (79) as

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

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

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

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

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

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

Costs of policy mistakes. From (76), the first derivative of the welfare function w.r.t. μ_a is:

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

so that the second derivative is

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

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

$$\mathbb{W}\left(\mu_a\right) = \int_{x=\mu_a^*}^{\mu_a}\mathbb{W}'\left(x\right)dx + \mathbb{W}\left(\mu_a^*\right), \qquad \mathbb{W}'\left(\mu_a\right) = \int_{x=\mu_a^*}^{\mu_a}\mathbb{W}''\left(x\right)dx$$

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

Combining, we can derive

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

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

Additional distortions. With a labor wedge and cost-push shocks, the Phillips curve and aggregate labor supply condition 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).

With labor market distortions, the output gap is given by $(\mu_a + \phi_l \tau_{la}) \varepsilon_t$, and with this substitution the remaining equilibrium conditions, including the form of TFP and κ are unchanged. Using the Phillips curve and the method of undetermined coefficients, wage inflation satisfies $\pi_t^w = (\lambda_w \mu_a + \eta_a) \varepsilon_t$ and the welfare function is given by

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

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

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

where

$$\Phi_l = \frac{1 + (\kappa_l + \tau_{\Lambda y})^2 \phi_l \Omega}{\frac{1}{\phi_l} + \frac{\alpha_2 \nu_w \lambda_w}{\phi_l} + (\kappa_l + \tau_{\Lambda y})^2 \Omega} > 0, \qquad \Phi_\eta = \frac{\frac{\alpha_2 \nu_w}{\phi_l}}{\frac{1}{\phi_l} + \frac{\alpha_2 \nu_w \lambda_w}{\phi_l} + (\kappa_l + \tau_{\Lambda y})^2 \Omega} > 0$$

and $\frac{\partial \Phi_\eta}{\partial \sigma_\beta^2} < 0$, $\frac{\partial \Phi_l}{\partial \sigma_\beta^2} > 0$, and Φ_Λ is as defined in (77).

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. For example, consider the case with no other distortions and flexible prices, i.e., $\tau_{\Lambda y} = \eta_a = \lambda_w = 0$. Optimal policy entails completely neutralizing the labor distortion and hence keeping the output gap always at zero, i.e., $\mu_a^* = -\phi_l \tau_{la}$.[62] With nominal rigidities, the optimal response is less than one-for-one, i.e., $\Phi_l < \phi_l$: 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, even if there are no financial frictions; with heterogeneity, the wedge not only has a direct effect on 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. Financial frictions further augment this response, i.e., $\frac{\partial \Phi_l}{\partial \tau_{\Lambda y}} > 0$.

Because of the cost-push shock, in the standard way, the central bank faces a tradeoff and cannot

[62] Of course, with flexible prices, monetary policy cannot achieve this outcome. However, fiscal policy can.

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 allows for fluctuations in the output gap. The slope of the optimal tradeoff between inflation and the output gap is given by Φ_η, which is strictly decreasing in σ_β^2 – with 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 – choosing μ_a to lean against the shock distorts the capital allocation.[63] Thus, heterogeneity reduces the optimal response to the cost-push shock and affects the tradeoff between inflation/output gap volatility, even if there are no financial frictions. Financial frictions further augment the dampening effects of heterogeneity, i.e., $\frac{\partial \Phi_\eta}{\partial \tau_{\Lambda y}} < 0$.

Optimal fiscal policy. In a flexible price environment, the equilibrium and policy objective function are special cases of the environment laid out in Section 2 where (i) the nominal side of the economy is completely disentangled from real quantities and the macro dynamics are fully characterized by (10) and (14) with the output gap driven by the labor tax chosen by the fiscal authority, and (ii) the welfare function takes the form in (24) with no costs of inflation. Specifically, the output gap is given by $\phi_l \tau_{la} a_t$ where τ_{la} captures the cyclicality of the labor income tax chosen by the fiscal authority, and following similar steps as above, we can derive the welfare loss function as

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

which is the analog to (73). Again following similar steps as with optimal monetary policy, we can take the first order condition to obtain the optimal fiscal policy:

$$\tau_{la}^* = -\tau_{\Lambda y} \sigma_\beta^2 \Phi_\Lambda^f \quad \text{where} \quad \Phi_\Lambda^f \equiv \frac{\frac{\phi_\psi \omega (\kappa_l + \tau_{\Lambda y})}{(1+\omega \Upsilon (\kappa_\psi + \tau_{\Lambda y} \phi_\psi))^2} \frac{\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2 \sigma_\beta^2}}{1 + \phi_l \left(\kappa_l + \tau_{\Lambda y}\right)^2 \Omega} > 0, \qquad \frac{\partial \tau_{la}^*}{\partial \sigma_\beta^2} < 0 \tag{80}$$

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

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

Taking first order conditions with respect to μ_a and τ_{la}, we can show that $\mu_a^* = 0$ and τ_{la}^* satisfies (80), 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.

C Empirical Analysis

This appendix provides details of the empirical analysis in Section 3.

C.1 Data

We compute firm-level betas using quarterly data from Compustat over the period 1963-2019. We first calculate firm-level TFP as $z_{it} = \log\left(P_{it} Y_{it}\right) - \alpha k_{it}$, which, from expressions (7) and (12), equals $\beta_i a_t$ plus a term that is common across firms that captures the effects of movements in aggregate output

[63] For example, when $\tau_{\Lambda y} = 0$, the allocation and dynamics of TFP are efficient when $\mu_a = 0$, and thus the impact of the policy response to the cost-push shock on these margins is unambiguously distortionary.

and wages. Note that using α as defined in (7) as the appropriate curvature parameter incorporates the optimal labor choice in (6) and ensures that we calculate physical productivity (i.e., 'TFPQ') rather than revenue productivity ('TFPR'). We measure k_{it} as (the log of) property, plant and equipment (Compustat variable PPENTQ). We measure $P_{it}Y_{it}$ as value-added, which we compute as a constant fraction of revenues (Compustat variable SALEQ) using a share of intermediates of 0.5. We use sector-specific α's calculated from sectoral labor shares for the nine major sectors of the industrial classification following the method in David and Venkateswaran (2019). In brief, expression (6) implies that the labor share is equal to $LS = \alpha_2\nu$. Substituting into the definition of α yields $\alpha = \frac{\alpha_1\nu}{1-\alpha_2\nu} = \frac{\nu - LS}{1-LS}$. Sector-specific labor shares are calculated using data on value-added and labor compensation from the Bureau of Economic Analysis (BEA) Industry Economic Accounts.

Next, for each firm we estimate a time-series regression of firm-level TFP growth, Δz_{it}, on aggregate TFP growth, $\Delta\psi_t$. We use rolling four-quarter growth rates to control for issues of seasonality in the firm-level data and include firms with at least 40 (possibly non-consecutive) observations. Using the production structure of the economy, i.e., expressions (4), (5) and (12), which together with the second-order approximation in Appendix A.6 imply (32), the coefficient from this regression is equal to

$$\beta_i^{obs} = \beta_i \frac{\rho_a^2 - \rho_a + 2\psi_a + \psi_a\rho_a - \psi_a\rho_a^2}{2\psi_a^2 + 2\rho_a^2 - 2\psi_a\rho_a + 2\psi_a^2\rho_a - 2\psi_a\rho_a^2} + \text{const.} \tag{81}$$

which shows that observed betas, β_i^{obs} are proportional to true betas, β_i, up to an additive constant.

From expression (11), we calculate aggregate TFP as $\psi_t = y_t - \alpha_1 k_t - \alpha_2 l_t$ using data on real GDP, aggregate labor and capital. We measure labor as total hours worked in the non-farm business sector as reported by the Bureau of Labor Statistics (BLS). We compute quarterly aggregate capital stocks following the method in Gomme et al. (2011) using data on annual capital stocks and quarterly investment flows (both deflated by the GDP deflator) from the BEA. In brief, the procedure requires solving for the depreciation rate that reconciles the reported annual stocks with the quarterly investment flows. We then construct the quarterly capital stocks using the perpetual inventory method. Data on real GDP and hours are downloaded from the St. Louis FRED database (series GDPC1 and HOANBS, respectively). The annual capital stock is measured as the net stock of private non-residential fixed assets, obtained from the BEA Fixed Assets Tables. Quarterly investment is similarly measured as private non-residential fixed investment, obtained from the BEA.

For the local projections in Figure 2, we calculate cov $\left(k_{it}, \beta_i^{obs}\right)$ on a quarterly basis. For Figure 3, we compute $mrpk_{it} = \log P_{it}Y_{it} - k_{it}$ for each firm-quarter, calculate cov $\left(mrpk_{it}, \beta_i^{obs}\right)$ and $\sigma^2_{\beta^{obs}}$ on a quarterly basis, and use their quarterly averages as initial values. We trim the 1% tails of k_{it}, $mrpk_{it}$ and β_i^{obs}. We use the monetary policy surprise measure of Bauer and Swanson (2023b) over the period 1988-2019, available on the Federal Reserve Bank of San Francisco website at `https://www.frbsf.org/research-and-insights/data-and-indicators/monetary-policy-surprises/`. Data on the unemployment rate are downloaded from the St. Louis FRED database (series UNRATE). The Bauer et al. (2024) measure of the perceived monetary policy rule is available at `https://www.michaeldbauer.com/files/bps_bluechip_rules.xlsx`, series 'gamma_inertial.' The realized output gap is calculated as the percent deviation of real GDP from real potential GDP as estimated by the CBO (downloaded from the St. Louis FRED database, series GDPPOT) and is one-sided HP-filtered. For the plots in Figure 5, we trim firms with observed betas more than 10 in absolute value as a rough adjustment for measurement/sampling error.

Stylized facts. Figure 5 illustrates the two main features of the micro-data that motivate our analysis. Panel A plots the cross-sectional distribution of β_i^{obs} and shows the large dispersion in observable measures of firm cyclicality. Panel B displays a bin-scatter of $mrpk_{it}$ against β_i^{obs} and shows the strong positive relationship between MRPK and beta.

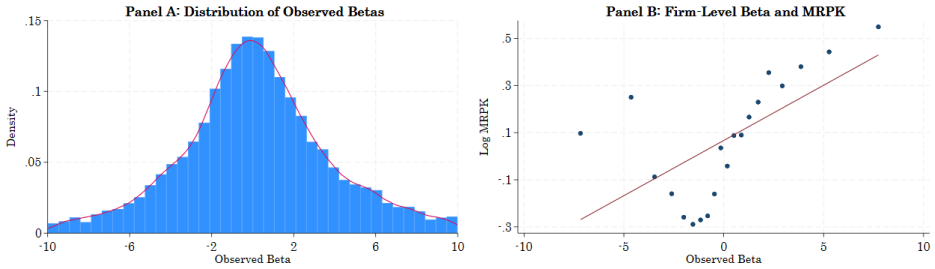


Figure 5: Firm Cyclicalities and MRPK

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 line the kernel density estimate. Panel B displays a bin-scatter of $mrpk_{it}$ against β_i^{obs}.

C.2 Empirical Motivation

Proof of Proposition 6. (i) Under full information about ϕ_{yt}, $\tilde{\phi}_{yt} = \phi_{yt}$ and hence $mps_t = \xi_t$, i.e., the monetary policy surprise reflects only the monetary policy shock and thus contains no information about the monetary policy rule. Following similar steps as in Appendix A.4, it is straightforward to verify that κ, and hence the capital allocation and TFP are unaffected by the monetary policy shocks. Because the remainder of the model is identical to the analogous representative firm model, monetary policy shocks have the same effects as in that model.

(ii) Rearranging expression (27) shows that mps_t is a noisy signal of the true coefficient ϕ_{yt}, i.e.,

$$\frac{mps_t}{\mu_t} = \left(\phi_{yt} - \tilde{\phi}_{yt}\right) + \frac{\xi_t}{\mu_t}$$

Applying the equations of the Kalman filter yields the result.

(iii) Under the monetary policy rule in (26), the output gap satisfies $\mu_t = \mathbb{E}_{t-1}[\mu_t] + \mu_{at}\varepsilon_t + \mu_{\xi t}\xi_t$, which is a version of (51) extended to the case of a time-varying output gap coefficient and monetary policy shocks. Denote agents' perceptions of μ_{at} at date $t-1$ as $\tilde{\mu}_{at} = \mathbb{E}_{t-1}[\mu_{at}|\mathcal{I}_{t-1}]$. The proof of Proposition 2 is unaffected by the monetary policy shock, ξ_t, since it is independent of ε_t and substituting $\tilde{\mu}_{at}$ for μ_t is otherwise the same as in Appendix A.4 but where $\tilde{\mu}_{at}$ and hence κ_t are time-varying. The proofs of Propositions 3 and 4 are unchanged save for the time-varying κ_t.

□

Output gap and unemployment deviations. We formally derive the link between the output gap and unemployment rate deviation in the context of the model from Section 2. As there, households are divided into differentiated labor types $j \in [0,1]$ but, as in Galí (2011a,b), each household has a continuum of members indexed by i that vary in their disutility from work, which is given by χi^{φ} if the member is employed and zero otherwise. There is full risk sharing within and across households.

Household utility is

$$U(C_t, \{L_t(j)\}) = \frac{C_t^{1-\gamma}}{1-\gamma} - \chi \int_0^1 \int_0^{L_t(j)} i^{\varphi} di dj = \frac{C_t^{1-\gamma}}{1-\gamma} - \chi \int_0^1 \frac{L_t(j)^{1+\varphi}}{1+\varphi} dj$$

where $L_t(j)$ is the fraction (or total number) of type j workers who are employed.

A worker of type i desires to work if $W_t \geq \chi C_t^{\gamma} i^{\varphi}$ and thus, the marginal worker of type j, which we denote by $N_t(j)$, satisfies

$$W_t = \chi C_t^{\gamma} N_t(j)^{\varphi}$$

Note that $N_t(j)$ is also the total number of type j workers who desire to work.

The aggregate labor force (or participation rate) is $N_t = \int_0^1 N_t(j)\,dj$. Taking logs, integrating, and approximating, we have

$$w_t = \gamma c_t + \varphi n_t$$

In deviations, the unemployment rate is defined as

$$u_t = \frac{N_t - L_t}{N_t} = 1 - \frac{L_t}{N_t} \approx n_t - l_t$$

Combining the last two equations with the definition of the output gap, $\mu_t = \phi_l(\varphi l_t + \gamma y_t - w_t)$, yields

$$-u_t = \frac{1}{\varphi\phi_l}\mu_t$$

i.e., the negative unemployment rate deviation is proportional to the output gap.

Panel local projection. We show that, in theory, the IRF from the covariance LP is equal to that from a panel LP, up to a constant scaling factor. Consider the following panel LP of firm-level capital on the interaction of β_i with $pmr_t = mps_t \times \widehat{un}_{t-1}$:

$$k_{it+h} = \gamma_0^h + \gamma_{\beta,pmr}^h \beta_i mps_t \widehat{un}_{t-1} + \gamma_{\beta,mps}^h \beta_i mps_t + \gamma_{\beta,\hat{\mu}}^h \beta_i \widehat{un}_{t-1} + \gamma_k^h k_{it-1} + \gamma_i^h + \gamma_t^h + \zeta_{it+h} \tag{82}$$

where by assumption, the error term ζ_{it+h} is uncorrelated with β_i. The specification controls for lagged capital, firm-horizon and time-horizon fixed-effects, and all elements of the triple interaction of β_i with mps_t and $\widehat{un}_{t-1}$ that are not picked up by the fixed-effects, i.e., there is no need to additionally control for β_i, mps_t, $\widehat{un}_{t-1}$, or the interaction of the latter two.[64] The coefficients of interest are $\left\{\gamma_{\beta,pmr}^h\right\}$, which capture the differential response of firms with different betas to the *pmr* shock.

This specification implies

$$\text{cov}(k_{it+h}, \beta_i) = \gamma_0^h + \gamma_{\beta,pmr}^h \sigma_\beta^2 mps_t \widehat{un}_{t-1} + \gamma_{\beta,mps}^h \sigma_\beta^2 mps_t + \gamma_{\beta,\hat{\mu}}^h \sigma_\beta^2 \widehat{un}_{t-1} + \gamma_k^h \text{cov}(k_{it-1}, \beta_i)$$

where $\gamma_0 = \text{cov}\left(\gamma_i^h, \beta_i\right)$, which is equivalent to the covariance LP in (28) where $\theta_{pmr}^h = \gamma_{\beta,pmr}^h \sigma_\beta^2$, i.e., the coefficient from the covariance LP equals that from the panel LP scaled by a constant, σ_β^2.

Note, however, that if β_i is observed with error, the coefficient estimates from the panel LP will be biased toward zero. In contrast, $\text{cov}(k_{it+h}, \beta_i)$ is not affected by such errors so long as the errors in capital and beta are uncorrelated. Moreover, the coefficient estimates from the covariance LP are not affected even by correlated errors so long as the degree of error does not respond to the shock.

Approximate effects of *pmr* shocks on σ_{mrpk}^2 and TFP. Here we describe our method of approximating the impact of the *pmr* shock on σ_{mrpk}^2 and TFP using only the production structure of the economy, namely, expressions (4), (5), which together imply (7) and hence $\frac{\partial mrpk_{it}}{\partial k_{it}} = -(1-\alpha)$, and (12), but no other conditions of the model.

To approximate the IRF of σ_{mrpk}^2, note that by definition $\sigma_{mrpk_t}^2 = \int \left(mrpk_{it} - \int mrpk_{it}\right)^2 di$ and therefore a first-order Taylor expansion of the change in σ_{mrpk}^2 in response to the shock yields:

$$\begin{aligned}\Delta\sigma_{mrpk_t}^2 &\approx 2\int \left(mrpk_i - \int mrpk_i di\right)\left(\frac{\partial mrpk_i}{\partial k_{it}}\Delta k_{it} - \int \frac{\partial mrpk_i}{\partial k_{it}}\Delta k_{it} di\right) di \\ &= -2(1-\alpha)\,\text{cov}(mrpk_i, \Delta k_{it})\end{aligned}$$

[64] The equivalence result extends to the case of time-varying betas.

where $mrpk_i$ denote values at the point of approximation, e.g., the stochastic steady state.

Next, by the OLS formula, we can decompose Δk_{it} as

$$\Delta k_{it} = \frac{\text{cov}\left(\Delta k_{it}, \beta_i\right)}{\sigma_\beta^2}\beta_i + e_{it} \Rightarrow \text{cov}\left(mrpk_i, \Delta k_{it}\right) = \frac{\text{cov}\left(\Delta k_{it}, \beta_i\right)\text{cov}\left(mrpk_i, \beta_i\right)}{\sigma_\beta^2} + \text{cov}\left(mrpk_i, e_{it}\right)$$

where e_{it} is orthogonal to β_i and we use the fact that $\text{cov}\left(\Delta k_{it}, \beta_i\right) = \Delta\text{cov}\left(k_{it}, \beta_i\right)$. Then,

$$\Delta\sigma^2_{mrpk_t} = -2\left(1-\alpha\right)\left(\frac{\text{cov}\left(mrpk_i, \beta_i\right)\Delta\text{cov}\left(k_{it}, \beta_i\right)}{\sigma_\beta^2} + \text{cov}\left(mrpk_i, e_{it}\right)\right)$$

The impact of the reallocation induced by the shock across β_i on σ^2_{mrpk} is equal to

$$-2\left(1-\alpha\right)\frac{\text{cov}\left(mrpk_i, \beta_i\right)\Delta\text{cov}\left(k_{it}, \beta_i\right)}{\sigma_\beta^2} \tag{83}$$

whereas $\text{cov}\left(mrpk_i, e_{it}\right)$ captures reallocation that is orthogonal to β_i and hence independent of risk. Expression (83) thus isolates – and directly captures – the change in $mrpk$ dispersion induced by the shock via risk-based reallocation alone, and is the object we plot in Panel A of Figure 3. Note that (83) is invariant to whether we use β_i or β_i^{obs} and with this substitution, we calculate $\text{cov}\left(mrpk_i, \beta_i^{obs}\right)$ and $\sigma^2_{\beta^{obs}}$ directly from the data and $\Delta\text{cov}\left(k_{it}, \beta_i^{obs}\right)$ is the empirical IRF of $\text{cov}\left(k_{it}, \beta_i^{obs}\right)$.

Following the derivations in Appendix A.6, a second-order approximation yields (log) TFP as

$$\psi_t = \rho_a a_{t-1} - \frac{1}{2}\omega\sigma^2_{mrpk_t} + \left(1 + \alpha_1\text{cov}\left(k_{it}, \beta\right)\right)\varepsilon_t$$

Thus, the percentage changes in the level and conditional volatility of TFP are

$$\begin{aligned}\Delta\psi_t &= -\frac{1}{2}\omega\Delta\sigma^2_{mrpk_t}\\ \frac{\Delta\psi_{at}}{\psi_a} &= \frac{\alpha_1\Delta\text{cov}\left(k_{it}, \beta_i\right)}{1+\alpha_1\text{cov}\left(k_i, \beta_i\right)} = \frac{\alpha_1\lambda}{1+\alpha_1\lambda\text{cov}\left(k_i, \beta_i^{obs}\right)}\Delta\text{cov}\left(k_{it}, \beta_i^{obs}\right)\end{aligned}$$

The former can be calculated from (83) and the latter from the empirical IRF of $\text{cov}\left(k_{it}, \beta_i^{obs}\right)$ up to a scaling factor that depends on the initial value of $\text{cov}\left(k_i, \beta_i^{obs}\right)$, which we compute as the average quarterly value in the data, and the factor of proportionality mapping observed to true betas in (81), which we denote λ. In practice, we use the value of this factor from the calibrated model, which puts the IRF into interpretable units, but does not affect its shape or significance.

C.3 Additional Empirical Results

TFP from macro-data. Figure 6 estimates the IRFs of TFP computed from macro-data on aggregate output, labor and capital as described in Appendix C.1, rather than as implied by the micro-level IRF as in Figure 3. We calculate the volatility of TFP in the same way we calculated the volatility of the output gap in Figure 1, i.e., as the forward-looking eight quarter rolling standard deviation (analogous to Panel B of Figure 1, we control for TFP volatility over the eight quarters preceding the shock and for the lag of its level). The figure shows that the patterns implied by our results in the micro-data show up in the macro-data as well, namely, the *pmr* shock leads to persistent and statistically significant increases in both the level and volatility of aggregate TFP.[65]

[65] Qualitatively similar results hold using the TFP series from the Federal Reserve Bank of San Francisco.

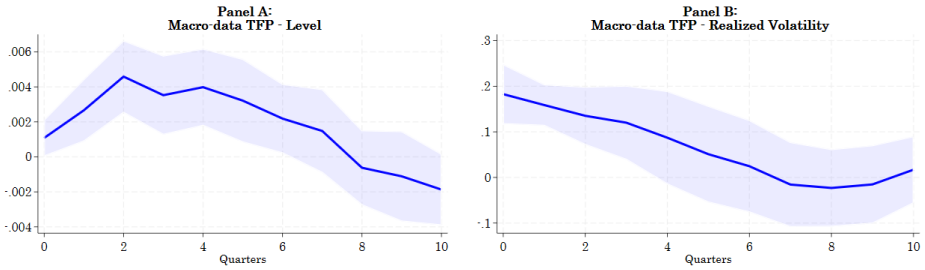


Figure 6: Perceived Rule Shocks and TFP from Macro-Data

Notes: Figure displays local projection impulse response functions of aggregate TFP computed from macro-data to a 25 b.p. perceived monetary rule (*pmr*) shock. Panel A displays the response of the level of TFP. Panel B displays the response of the forward looking eight quarter rolling standard deviation of TFP. Both panels are in log points. Shaded areas report 90% confidence bands based on Newey-West standard errors with the number of lags selected using the nonparametric lag selection procedure of Newey and West (1994).

Panel local projection. Figure 7 estimates a panel local projection of firm-level capital on the *pmr* shock interacted with a high-frequency, time-varying measure of firm cyclicality computed from stock market data. We estimate market betas, β^s_{it}, using a standard market model of the form $r_{it} = \text{const} + \beta^s_{it} r_{mt} + \zeta_{it}$, which relates firm-level stock returns, r_{it}, to the aggregate market return, r_{mt}, with a firm-specific, time-varying exposure β^s_{it}, and an idiosyncratic component, ζ_{it}. We estimate this model quarter-by-quarter using daily stock market data such that β^s_{it} varies at a quarterly frequency and use a backward-looking four-quarter moving average. We then estimate a panel LP of the form in (82), extended in a straightforward way to accommodate time-varying betas. The figure plots the coefficient of firm-level capital, k_{it+h} at each horizon h, on the interaction of β^s_{it-1} with the *pmr* shock and thus captures the heterogeneous effects of the shock on the capital choice across firms with different betas. The figure shows positive, persistent, and statistically significant coefficients in response to a countercyclical *pmr* shock, meaning that the shock leads to a reallocation of capital toward high beta firms as predicted by the theory and confirming our baseline findings in Figure 2.

D Quantitative Analysis

This appendix provides details of the quantitative analysis in Section 4.

D.1 Quantitative Model

The quantitative model extends the model from Section 2 on two dimensions: first, we allow for aggregate capital investment subject to adjustment costs. Second, we adopt the financial intermediation block of Gertler and Karadi (2011). We describe these two elements in turn, then lay out the full set of equilibrium conditions.

Capital producing sector. A representative competitive firm produces new capital goods and sells these goods to the financial intermediary at (real, cum-dividend) price $\tilde{P}_{kt}$. The production function for capital goods features quadratic investment adjustment costs, so that the cost of producing I_t units of new capital is $\Phi(I_t, I_{t-1}) = \left(1 + \frac{\theta_k}{2}\left(\frac{I_t}{I_{t-1}} - 1\right)^2\right) I_t$ units of the final good. Similar to intermediate good firms, the capital producer is owned by the financial intermediary and acts to

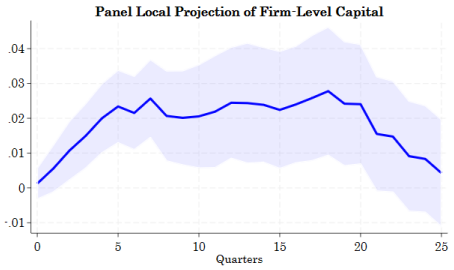


Figure 7: Perceived Rule Shocks and Firm-Level Capital

Notes: Figure displays local projection impulse response function of (log) firm-level capital to a 25 b.p. perceived monetary rule (*pmr*) shock interacted with time-varying firm-level stock market betas. Shaded areas report 90% confidence bands based on standard errors two-way clustered by firm and date.

maximize expected discounted value, i.e.,

$$V_t^K = \max_{I_t} \ \tilde{P}_{kt} I_t - \Phi\left(I_t, I_{t-1}\right) + \mathbb{E}_t\left[\Lambda_{t+1} T_{\Lambda t+1} V_{t+1}^K\right]$$

The first order and envelope conditions yield (29) and the aggregate capital stock evolves as in (30).

For simplicity, we assume that the cost of adjusting investment is a purely financial cost paid to households and thus the aggregate resource constraint takes the standard form

$$Y_t = C_t + I_t \tag{84}$$

Financial intermediation sector. The setup is largely the same as in GK, modified to allow for heterogeneous firms. 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. Intermediaries exit exogenously with probability $1-\sigma$ and are immediately replaced by new entrants, keeping the relative proportion of intermediaries fixed. The average lifespan of an intermediary 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 (7). Intermediaries borrow from households at the risk-free rate in order to provide funding to firms.

The balance sheet of intermediary j at the end of date t is given by

$$\int Q_{it} S_{ijt} di + P_t^K K_{jt+1} = N_{jt} + B_{jt} \tag{85}$$

Intermediary assets consist of the total market value of its ownership claims on firms and the value of its capital holdings, where S_{ijt} denotes the quantity of claims held by intermediary j on firm i. The liability side of the balance sheet consist of deposits, B_{jt}, and net worth, N_{jt}. Net worth is equal to the gross return on assets less costs of borrowing:

$$N_{jt} = \int R_{it}^e Q_{it-1} S_{ijt-1} di + R_{kt} P_{t-1}^K K_{jt} - R_t B_{jt-1} \tag{86}$$

Here $R_{it}^e = \frac{D_{it}+Q_{it}}{Q_{it-1}}$ denotes the return on the equity claim on firm i, where D_{it} is the equity payout of the firm, $R_{kt} = \frac{(1-\delta)\tilde{P}_t^K}{P_{t-1}^K}$ the return on capital, which is common across across firms, and R_t the

risk-free rate.[66] Combining (85) and (86), net worth evolves as

$$N_{jt+1} = \int R^e_{it+1} Q_{it} S_{ijt} di + R_{kt+1} P^K_t K_{jt+1} + R_{t+1} \left(N_{jt} - \int Q_{it} S_{ijt} di - P^K_t K_{jt+1} \right) \tag{87}$$

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

$$V_{jt} = \mathbb{E}_t \left[\Lambda_{t+1} \left((1-\sigma) N_{jt+1} + \sigma V_{jt+1} \right) \right] \tag{88}$$

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:

$$V_{jt} \geq \theta \left(\int Q_{it} S_{ijt} di + P^K_t K_{jt+1} \right) \tag{89}$$

The intermediary chooses its holdings of firms' securities, S_{ijt}, and capital, K_{jt+1}, to maximize (88) subject to (87) and (89). Taking first order conditions and rearranging yields

$$\mathbb{E}_t \left[\Lambda_{t+1} \Theta_{jt+1} \left(R^e_{it+1} - R_{t+1} \right) \right] = \mathbb{E}_t \left[\Lambda_{t+1} \Theta_{jt+1} \left(R_{kt+1} - R_{t+1} \right) \right] = \theta \frac{\zeta_t}{1+\zeta_t} \tag{90}$$

where ζ_t is the Lagrange multiplier on (89) and

$$\Theta_{jt} \equiv 1 - \sigma + \sigma \frac{\partial V_{jt}}{\partial N_{jt}} \tag{91}$$

Denote the value of the total portfolio of intermediary j assets $A_{jt} = \int Q_{it} S_{ijt} + P^K_t K_{jt+1}$ and the return on this portfolio $R^A_{jt+1} = \frac{\int R^e_{it+1} Q_{it} S_{ijt} di + R_{kt+1} P^K_t K_{jt+1}}{\int Q_{it} S_{ijt} di + P^K_t K_{jt+1}}$. As in GK, we can guess and verify that

$$V_{jt} = \nu_t A_{jt} + \eta_t N_{jt} \tag{92}$$

where $\nu_t = \mathbb{E}_t \left[\Lambda_{t+1} \Theta_{jt+1} \left(R^A_{t+1} - R_{t+1} \right) \right]$ and $\eta_t = \mathbb{E}_t \left[\Lambda_{t+1} \Theta_{jt+1} \right] R_{t+1}$.

Expression (92) shows that as in GK, the intermediary problem is homogeneous of degree one in N_{jt}, which implies $\frac{\partial V_{jt}}{\partial N_{jt}} = \frac{\partial V_t}{\partial N_t} = \frac{V_t}{N_t}$, where variables without j subscripts denote aggregates of the entire intermediary sector. Thus, we can work with a representative intermediary with equity market value $V_t = \int V_{jt} dj$, net worth $N_t = \int N_{jt} dj$, value of assets $A_t = \int A_{jt} dj$, and return on assets $R^A_{t+1} = \frac{\int R^e_{it+1} Q_{it} S_{it} di + R_{kt+1} P^K_t K_{t+1}}{\int Q_{it} S_{it} di + P^K_t K_{t+1}}$. Substituting $\Theta_{jt} = \Theta_t$ into (90) we obtain that for any asset x,

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

and because this also holds for the intermediary portfolio (i.e., $x = A$), substituting yields

$$\mathbb{E}_t \left[\Lambda_{t+1} T_{\Lambda t+1} R^x_{t+1} \right] = 1 \quad \text{where} \quad T_{\Lambda t+1} = \frac{\Theta_{t+1}}{\mathbb{E}_t \left[\Lambda_{t+1} \Theta_{t+1} R^A_{t+1} \right]} \tag{93}$$

Because (93) holds for any asset, $\Lambda_{t+1} T_{t+1}$ represents a valid SDF to price any asset, which shows that the GK framework leads to a distorted SDF as in the analytic model in Section 2, where the wedge,

[66] The capital return reflects our assumption that the rental rate for period t capital is paid in period $t-1$ so that the effective price is $P^K_{t-1} = \tilde{P}^K_{t-1} - R^K_{t-1}$. Equilibrium quantities are unchanged if we assume these payments are made in period t, in which case the return would be $\frac{R^K_{t-1} + (1-\delta) P^K_t}{P^K_{t-1}}$.

$T_{\Lambda t}$, reflects 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 equation (8).

Note that the collateral constraint in (89), $V_t \geq \theta A_t$, implies

$$\nu_t A_t + \eta_t N_t \geq \theta A_t \;\Rightarrow\; A_t \leq \phi_t N_t \text{ or } B_t \leq (\phi_t - 1) N_t$$

where $\phi_t = \frac{\eta_t}{\theta - \nu_t}$, i.e., the constraint is equivalent to an endogenous, time-varying limit on intermediary leverage, $\frac{N_t}{A_t}$, or alternatively, on intermediary debt as a fraction of net worth.

Aggregate intermediary equations – New entrants are capitalized with a fraction $\frac{n_e}{1-\sigma}$ of the value of exiter assets, so that the total net worth of entrants is $n_e A_t$. Summing across (87) and accounting for entry/exit, net worth of the intermediary sector evolves as

$$N_{t+1} = \sigma \left(R^A_{t+1} A_t + R_{t+1} (N_t - A_t) \right) + n_e A_t \tag{94}$$

The collateral constraint always binds in a neighborhood of the steady state, which implies $V_t = \theta A_t$. Substituting into (92),

$$\frac{A_t}{N_t} = \frac{\mathbb{E}_t \left[\Lambda_{t+1} \Theta_{t+1} \right] R_{t+1}}{\theta - \mathbb{E}_t \left[\Lambda_{t+1} \Theta_{t+1} \left(R^A_{t+1} - R_{t+1} \right) \right]} \tag{95}$$

The return on all intermediary sector assets is given by

$$R^A_{t+1} = \frac{(1 - \alpha_2 \nu) Y_{t+1} - P^K_{t+1} \left(K_{t+2} - (1 - \delta) K_{t+1} \right) + A_{t+1}}{A_t} \tag{96}$$

where we have used the fact that aggregate firm equity payouts are equal to $D_{t+1} = \int \Pi_{it+1} di - R^K_t K_{t+1} = (1 - \alpha_2 \nu) Y_{t+1} - R^K_t K_{t+1}$.

Similarly, substituting $\frac{\partial V_t}{\partial N_t} = \theta \frac{A_t}{N_t}$ into (91), we can write

$$\Theta_t = 1 - \sigma + \sigma \theta \frac{A_t}{N_t} \tag{97}$$

and using (95), we can write the wedge as

$$T_{\Lambda t+1} = \frac{\Theta_{t+1}}{\theta + \mathbb{E}_t \left[\Lambda_{t+1} \Theta_{t+1} \right] R_{t+1} \left(1 - \frac{N_t}{A_t} \right)} \tag{98}$$

Because the intermediary's choice of capital must satisfy (93), we have $1 = \mathbb{E}_t \left[\Lambda_{t+1} T_{\Lambda t+1} R_{kt+1} \right]$ and substituting for the definition of the capital return, R_{kt+1}, and combining with (8) yields

$$P^K_t = \alpha_1 \nu \mathbb{E}_t \left[\Lambda_{t+1} T_{\Lambda t+1} \frac{Y_{t+1}}{K_{t+1}} \right] + (1 - \delta) \mathbb{E}_t \left[\Lambda_{t+1} T_{\Lambda t+1} P^K_{t+1} \right] \tag{99}$$

The intermediary sector is characterized by expressions (94)-(99), i.e., the law of motion of net worth, the leverage ratio, the financial wedge and aggregate capital demand.

Full set of equilibrium conditions. With the addition of these two elements, the remaining equilibrium conditions from Section 2 continue to hold. The full set of equilibrium conditions consists of 24 equations in 24 unknowns: (1) the household utility function, SDF, Philips curve and Euler equations for nominal and real bonds, i.e., expressions (1), (3), (46), (47) and (48) which constitute the household block; (2) the aggregate production function, labor demand, TFP, process on exogenous shocks, aggregate capital supply and capital accumulation, i.e., expressions (11), (43), (32), (12), (29) and (30), which along with the generic definition of κ in (72) and the definitions of ψ_a and $\overline{\psi}$

Table 4: Business Cycle Moments – Model vs. Data

	Data		Model	
	Rel. Std. Dev.	Corr. w/ y_t	Rel. Std. Dev.	Corr. w/ y_t
Output	1.00	1.00	1.00	1.00
Consumption	0.76	0.87	0.71	0.90
Investment	2.76	0.67	2.75	0.89

in (64) and (65) constitute the production block; (3) the law of motion of intermediary net worth, the return on intermediary assets, intermediary leverage ratio, capital wedge and aggregate capital demand, i.e., expressions (94)-(99), which constitute the intermediary block; (4) the monetary policy rule in (26) with constant coefficients and no monetary policy shocks, where $Y_t^{\text{flex}} = \Psi_t K_t^{\alpha_1} \left(L_t^{\text{flex}}\right)^{\alpha_2}$ and $\alpha_2 \frac{Y_t^{\text{flex}}}{L_t^{\text{flex}}} = \chi \left(Y_t^{\text{flex}} - I_t\right)^{\gamma} \left(L_t^{\text{flex}}\right)^{\varphi}$ characterize the flexible price equilibrium, which constitutes the policy block; and (5) the aggregate resource constraint in expression (84).

D.2 Calibration

Data and calculations. We compute the relative standard deviation of investment using annual data on private non-residential fixed investment and real GDP from the BEA National Income and Product Accounts. Investment is deflated by the GDP deflator. Aggregate TFP is calculated as in Section 3 using the annual series for real GDP, the capital stock, and hours worked. We use the reported series for hours worked from the BLS for 1951-2019 and the data reported in Kendrick (1961) for 1929-1950 (the series overlap for four years and have almost identical growth rates in each year). BLS data on average hourly earnings of production and nonsupervisory employees and business sector compensation per hour are downloaded from the St. Louis FRED database (series AHETPI and HCOMPBS, respectively). The output gap is calculated as in Section 3 using the annual series for real GDP and real potential GDP. All series are HP-filtered. Calculating wage inflation from the two measures of wages and regressing them on the output gap 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.

To compute $\sigma^2_{\hat{\beta}}$, we use the same betas as calculated in Section 3 and described above. From (81), the cross-sectional standard deviation of $\hat{\beta}_i$ is related to the standard deviation of observed betas as

$$\sigma\left(\hat{\beta}_i\right) = \sigma\left(\beta_i^{obs}\right) \frac{1-\alpha_2\nu}{\nu} \frac{2\psi_a^2 + 2\rho_a^2 - 2\psi_a\rho_a + 2\psi_a^2\rho_a - 2\psi_a\rho_a^2}{\rho_a^2 - \rho_a + 2\psi_a + \psi_a\rho_a - \psi_a\rho_a^2} \tag{100}$$

We 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 $\mathbb{E}\left[e_i\right] = 0$, we have

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

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

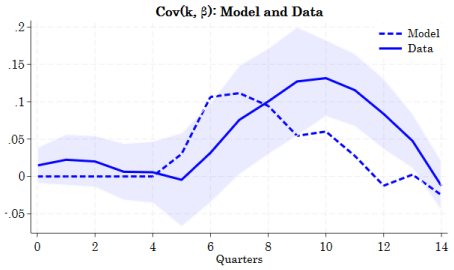


Figure 8: Perceived Rule Shocks and the Allocation of Capital: Model and Data

Notes: Figure displays local projection impulse response functions of cov $\left(k_{it}, \beta_i^{obs}\right)$ to a change in the perceived output gap coefficient in the monetary policy rule in the model and data. The data IRF is as described in Section 3. The model IRF is calculated by generating a sequence of perceived output gap coefficients identical to those in the data. Shaded areas report 90% confidence bands in the data based on Newey-West standard errors with the number of lags selected using the nonparametric lag selection procedure of Newey and West (1994).

Finally, using the fact that the mean squared standard error of the beta estimates represents just such an unbiased estimator, i.e., $\frac{1}{N}\sum_i \widehat{\sigma^2\left(e_i\right)} = \frac{1}{N}\sum_i \left(s.e.\left(\beta_i^{obs}\right)\right)^2$, 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 (100).

For Table 1, we calculate cov $\left(mrpk_i, \beta_i^{obs}\right)$ as the quarterly average. We calculate cov $\left(k_i, \beta_i^{obs}\right)$ as the quarterly average after conditioning out firm-level TFP, z_{it}, i.e., as cov $\left(k_{it}, \beta_i^{obs} | z_{it}\right) = \text{cov}\left(k_{it}, \beta_i^{obs}\right) - \frac{\text{cov}(k_{it}, z_{it})\text{cov}\left(z_{it}, \beta_i^{obs}\right)}{\text{var}(z_{it})}$.

To line up model and data, we estimate the model-generated IRF in Figure 4 in elasticity units, i.e., $\frac{\text{cov}\left(k_{it}, \beta_i^{obs}\right)}{\text{var}\left(\beta_i^{obs}\right)}$ and re-scale by the quarterly average of var $\left(\beta_i^{obs}\right)$ observed in the data.

Business cycle moments. Table 4 reports the fit of the model across a broader set of business cycle moments, namely, the relative volatilities of investment and consumption and their co-movement with output. We use post-war (annual, HP-filtered) data on GDP, Personal Consumption Expenditures (PCE) and private non-residential fixed investment from the BEA to measure output, consumption and investment respectively, all deflated by the GDP deflator. The table shows that the model is largely in line with the broader set of post-war business cycle moments, specifically, that it generates patterns of relative volatilities and co-movement consistent with the data.[67]

Impact effect of perceived rule shock. Figure 8 plots the IRF of cov $\left(k_{it}, \beta_i^{obs}\right)$ to a change in the perceived output gap coefficient in the monetary policy rule in the model and data, beginning at the quarter of impact (rather than with a four quarter lag as in Figures 2 and 4). The figure shows that there is essentially no response of the allocation in the data until about six quarters after the shock. By virtue of the annual frequency of the capital decision in the model, there is no response in the model until five quarters after the shock. Thus, both the model and data exhibit no response of the allocation from impact through the first four to five quarters following the shock.

[67] To complete the correlation matrix, the correlation of investment and consumption is 0.55 in the data vs. 0.60 in the model.