HBANK: Monetary Policy with Heterogeneous Banks[†]

Marco Bellifemine **Rustam Jamilov** **Tommaso Monacelli**

May 2025

Abstract

We revisit the bank lending channel of monetary policy transmission in a Heterogeneous-Bank New Keynesian (HBANK) model with endogenous bank default risk. Using a sufficient-statistic approach, we show that the combination of banks' heterogeneity in the marginal propensity to lend and costly insolvency amplifies the real effects of monetary policy shocks. The central bank faces a trade-off between macroeconomic and financial stability: contractionary monetary policy shrinks bank net worth, raising the aggregate probability and resource cost of default. Addressing persistent inflationary pressure comes at the price of exacerbating financial fragility. Automatic *micro*-prudential regulation—targeting only the top quartile of banks—effectively mitigates the macroeconomic-financial stabilization trade-off. We provide empirical evidence supporting the heterogeneous effects of monetary policy on bank assets and insolvency probabilities, reinforcing the necessity of incorporating bank-level heterogeneity in monetary policy design.

Keywords: Monetary policy, prudential policy, financial stability, bank heterogeneity, New Keynesian.

JEL Classification E32, E44, E52, G21.

[†]We thank Javier Bianchi, Saki Bigio, Benjamin Born, Dean Corbae, Pablo D'Erasmo, Jeppe Druedahl, Andrea Fabiani, Ester Faia, Sebastian Fanelli (discussant), Fadi Hassan, Peter Karadi, Luc Laeven, Ben Moll, Dima Mukhin, Pablo Ottonello, Vincenzo Quadrini, Hélène Rey, Farzad Saidi, Ludwig Straub, Alejandro Van der Ghote, Ia Vardishvili (discussant), Olivier Wang, Ivan Werning, Russell Wong and seminar participants at numerous venues for valuable comments and suggestions. First version: July 2022.

Bellifemine: London School of Economics. Email: m.bellifemine@lse.ac.uk.

Jamilov: University of Oxford. Email: rustam.jamilov@all-souls.ox.ac.uk.

Monacelli: Bocconi University, CEPR, and IGIER. Email: tommaso.monacelli@unibocconi.it.

1 Introduction

The emphasis on the role of banks in the transmission of monetary policy has strengthened after the 2007-08 financial crisis and recession. In that context, disruptions in financial intermediation have been shown to significantly affect real economic activity. Following the 2023 U.S. regional banking crisis, there is a further renewed interest in the interactions between monetary policy and financial stability.

Most standard macro-finance frameworks study the link between monetary policy and financial stability without explicitly accounting for the *distribution* of financial intermediaries. However, empirical evidence has shown that the effects of monetary policy are not homogeneous across individual financial institutions.[1] Additionally, measures of financial instability, such as insolvency risk, are not distributed uniformly but, instead, covary systematically with bank size. As a result, the aggregate effects of monetary policy may depend explicitly on the cross-section of banks' balance sheets and insolvency risk.

The study of monetary policy transmission in models with *heterogeneous* banks is an emerging area of research in monetary macroeconomics. There is, however, a missing link between the workhorse model for the analysis of monetary policy—i.e., the New Keynesian (NK) framework (Woodford, 2003; Galí, 2008)—and the extensive macro-finance literature developed over the last decade (Gertler and Kiyotaki, 2010; Brunnermeier and Sannikov, 2014). In this paper, we aim at providing that link and study monetary policy transmission in a Heterogeneous-Bank New Keynesian (HBANK) model with four key features: (i) a realistic distribution of bank size, (ii) endogenous bank-specific insolvency risk, (iii) costly default, and (iv) nominal rigidities.

Within our HBANK environment, we document three main results. First, the transmission of monetary policy shocks is characterized by a rich heterogeneity in the response of bank lending and default probabilities. Due to a higher *marginal propensity to lend* (MPL), small banks' balance sheets are significantly more responsive than the ones of large banks. Furthermore, their probability of default rises above the one of the average intermediary. At the same time, even a marginal increase in the probability of default by the large banks suffices to increase default costs substantially, akin to a "too big to fail" phenomenon. These patterns could have direct implications for the optimal coordination of monetary and prudential policy.

Second, bank heterogeneity combined with costly insolvency can *amplify* the transmission of *non-systematic* monetary policy shocks. The mechanism operates through the model's endogenous financial stability channel. In HBANK, a monetary tightening shifts

[1] See, for example, the seminal empirical contributions by Kashyap and Stein (1995) and Kashyap and Stein (2000) which document the unequal incidence of monetary policy shocks across bank balance sheets.

the distribution of bank net worth leftward, reducing the average distance to default and increasing realized default costs. By contrast, in a representative-bank (RBANK) framework with complete markets and no default risk, this amplification channel is absent. Crucially, HBANK features an endogenous prudential policy response: in reaction to the rise in financial fragility, the regulator tightens leverage constraints on large banks. This policy feedback induces a second-round contraction in credit supply and output, reinforcing the original shock.

Third, a *systematic* monetary policy cannot simultaneously contain persistent inflationary pressures and enhance financial stability. While a contractionary policy helps reduce inflation, it also compresses banks' balance sheets and depresses asset valuations, shifting the distribution of net worth leftward. This raises both the incidence and cost of bank defaults. In the RBANK benchmark, such a policy has negligible effects on default probabilities. In HBANK, however, the monetary contraction disproportionately affects banks with low initial net worth—those closer to default and with greater sensitivity of lending to funding costs. This asymmetry increases the average default probability and amplifies real default costs, which rise with bank size. As a result, a monetary policy tightening aimed at curbing inflationary pressures raises *both* the aggregate probability and the realized cost of bank insolvency, generating a trade-off between macroeconomic and financial stability.

Setup. Our general framework features incomplete financial markets and uninsured idiosyncratic bank return risk. Idiosyncratic bank shocks and scale variance deliver endogenous, right-skewed distributions of bank assets, deposits, and net worth. Drawing a particularly large negative idiosyncratic draw can push a bank into bankruptcy. Insolvency risk is endogenous and, in equilibrium, concentrated in the left tail of the bank size distribution. Furthermore, default is costly. The cost of default is convex in bank size and is thus concentrated in the *right* tail of the distribution, capturing the systemic-risk features of modern banking sectors (Adrian and Brunnermeier, 2016). Bank deposits are not insured, and insolvency risk is priced competitively into the cross-section of retail deposit rates. On the other hand, the household and firm sectors are standard. Our framework nests a heterogeneous-bank model *without* nominal rigidities and costly default (Jamilov and Monacelli, 2025), the Gertler and Kiyotaki (2010); Gertler and Karadi (2011) representative-bank macro-banking model, and the canonical New Keynesian model (Woodford, 2003; Galí, 2008) all as special cases.

Distribution of banks. In the steady state, our model delivers realistic cross-sectional distributions of bank balance sheets: assets, deposits, and net worth. Insolvency risk and the cost of default are systematically decreasing and increasing with bank size, respectively. Because small banks are riskier ex ante, they pay a premium in the market for time deposits via higher rates of interest. Thus, the equilibrium marginal cost of operating the banking franchise is not uniformly distributed. As a result, the sensitivity to cost-of-funds shocks is heterogeneous across banks.

In fact, we show that in response to an unexpected interest rate shock, the bank-specific response is strongly heterogeneous. First, in line with the existing and our own empirical evidence, smaller banks cut lending relatively more and experience a significantly larger decline in net worth. As a result, while insolvency risk increases in the aggregate, small banks are especially closer to default. Interestingly, the response of the *cost* of default across the distribution is markedly different. Larger banks suffer a greater increase in the cost of default. A monetary contraction, therefore, causes a rise in *systemic riskiness* of the financial sector as captured by the increase in financial fragility of the large, systematically important intermediaries.

Sequence-space methods. We study aggregate transition dynamics in the sequence-space domain and characterize the general-equilibrium solution in terms of measurable sufficient statistics, following a burgeoning methodological literature (Mankiw and Reis, 2007; Boppart et al., 2018; Auclert et al., 2021a). To first order, the bank lending channel of monetary policy transmission (Bernanke and Blinder, 1988; Bernanke and Gertler, 1995) can be conveniently summarized by just *two sufficient statistics*: (i) a policy Jacobian and (ii) the general-equilibrium adjustment. The former collects the direct, partial-equilibrium responses of bank lending to an interest rate shock, holding the return on aggregate capital constant. The latter determines the indirect, general-equilibrium feedback effect on capital, which can include automatic prudential policy reactions to the initial monetary impulse. Importantly, the sequence of bank lending is sufficient to recover every other endogenous object in the model, such as inflation, output or bank deposits. This simple structure summarizes the bank lending channel in the model with high transparency.

Amplification of non-systematic monetary policy shocks. Our first key quantitative result is to demonstrate that HBANK delivers amplification of non-systematic monetary policy shocks over the representative-bank counterfactual. Quantitatively, the macroeconomic response is 30% greater in HBANK than in RBANK, and especially on impact. This finding is reminiscent of the financial accelerator mechanism (Bernanke and Gertler, 1989;

Bernanke et al., 1999), whereas costly financial constraints act as a powerful amplifier of exogenous aggregate shocks. Decomposing the total macroeconomic response to a monetary policy shock into direct and indirect effects reveals that the amplification result is due to the presence of stronger *direct* effects in HBANK, i.e., the interaction of heterogeneity with bank-level costly insolvency. This pattern has two simultaneous dimensions. First, a contrationary monetary policy shock redistributes bank net worth towards the banks that are small, closer to default, and have a high lending elasticity, thereby increasing the aggregate, economy-wide response. Second, the automatic micro-prudential policy response tightens leverage regulation on the largest banks, reducing financial fragility but also leading to second-round declines in credit supply.

Macroeconomic-financial stability trade-off. Our second main quantitative result is to show that, in HBANK, systematic monetary policy faces a trade-off between conventional macroeconomic stabilization and financial stability. Consider a monetary policy tightening that comes as a response to an inflationary supply or demand shock. For one, with nominal rigidities, this generates a contraction in output and inflation. However, the ensuing compression of bank balance sheets and negative asset valuation effects bring more banks closer to insolvency. In other words, the elasticity of credit supply to changes in the real interest rate is negative while the elasticity of aggregate bank default risk is positive. The contraction in bank size increases the economy-wide probability of default but is especially stronger for the ex-ante small banks that are closer to the default threshold initially. Simultaneously, the policy tightening leads to an increase in real resource costs of default, which are an increasing function of bank size. As a result, a monetary policy contraction raises *both* the probability of default and the aggregate default resource cost.

Endogenous micro-prudential policy. In the Heterogeneous-Agent New Keynesian (HANK) class of models, the indirect effect of unexpected interest rate cuts, such as the fiscal reaction function, is a key driver of the total macroeconomic response (Kaplan et al., 2018). In HBANK, our analysis puts at center stage the endogenous *prudential* policy response to a monetary policy shock. Systematic *micro*-prudential policy—which targets only the largest 25% of banks—automatically limits the leverage multiple of the banks whose collapse is especially costly for the economy. We show that the trade-off between macroeconomic stabilization and financial stability can be addressed with this *targeted* systematic prudential policy. Noticeably, this type of *micro*-prudential policy is only feasible in an environment with realistic cross-sectional distributions of bank-level characteristics.

The 2021-2023 U.S. inflation and banking crisis. A central bank can face sharp constraints in addressing persistent inflationary pressures even if it conducts monetary policy systematically. In an exercise that mimics the 2023 banking crisis in the United States, we show that a policy authority trying to tackle, via rising interest rates, a rise in inflation of the size and persistence experienced in the U.S. between 2021 and 2023, can generate a significant degree of delayed financial instability in the form of hightened probability and costs of bank default. In general, suppose the economy is hit by a persistent inflationary shock, requiring a higher real interest rate in response. With some delay, the higher real interest rate raises the ex-ante likelihood and the ex-post realized cost of bank default. Therefore a central bank that caters to financial stability concerns must raise interest rates by relatively *less*, thus taming inflation less aggressively than otherwise.

Empirical evidence. Finally, we validate our model using micro-data on U.S. commercial banks. We show—both in the model and in the data—that heterogeneity in bank size is crucial for understanding the responsiveness to monetary policy shocks. In particular, in the data, the balance sheets of smaller banks are significantly more responsive than those of larger banks, a finding consistent with earlier evidence from Kashyap and Stein (1995). This result underscores that a comprehensive theoretical and quantitative account of the bank lending channel requires modeling realistic bank size heterogeneity.

Next, we document a robust empirical relationship between bank size and insolvency-driven default risk. We proxy default risk using two complementary measures: the widely used "Z-score" (Laeven and Levine, 2009) and distance to default (Nagel and Purnanandam, 2019). Our findings show that smaller banks are systematically closer to insolvency, both in the cross-sectional distribution and over time. Hence, our model successfully reproduces the observed correlation between bank size and insolvency risk. Moreover, we find strong empirical evidence that default risk increases in response to identified positive monetary policy shocks, with the effect being especially pronounced for smaller banks. This conditional moment is in line with the model's predictions, serving as an important validation of the mechanism.

Literature review. Our paper is contributing to the burgeoning theoretical literature on heterogeneous financial intermediaries. Coimbra and Rey (2023) develop a general equilibrium framework with endogenous entry and financial intermediaries that are heterogeneous in Value-at-Risk constraints. Corbae and D'Erasmo (2021) build a quantitative model of banking industry dynamics with uninsured idiosyncratic return risk and imperfect credit-market competition. Begenau and Landvoigt (2021) develop a quantitative model with two banking sectors that approximate the divide between standard commercial and "shadow" banks. Bianchi and Bigio (2022) study the credit channel of monetary policy in an environment where banks face deposit withdrawal shocks. Goldstein et al. (2024) build a model of the financial system in which heterogeneous and interconnected intermediaries are prone to runs and fire sales.[2] Our contribution relative to this stream of papers is to embed bank heterogeneity and costly bank insolvency risk into the canonical NK framework and to study the interplay between monetary policy and financial stability.

We are contributing to the macro-finance literature that studies the general impact of the financial sector on the real economy. Contributions to this strand include, among many others, Brunnermeier and Pedersen (2009), Gertler and Kiyotaki (2010), Gertler and Karadi (2011), Jermann and Quadrini (2012), He and Krishnamurthy (2013), and Brunnermeier and Sannikov (2014).[3] The monetary transmission mechanism in our framework builds on the canonical bank lending channel (Bernanke and Blinder, 1988, Bernanke and Gertler, 1995, Stein, 1998). This channel is closely related to the financial accelerator mechanism (Bernanke and Gertler, 1989; Kiyotaki and Moore, 1997; Bernanke et al., 1999). The bank lending channel has received overwhelming empirical support over the years (Bernanke and Blinder, 1992; Gertler and Gilchrist, 1994). In particular, the heterogeneous incidence of monetary policy shocks has been repeatedly highlighted in the empirical literature (Kashyap and Stein, 1995, 2000; Kishan and Opiela, 2000). Using the latest U.S. bank-level data we revisit and reconfirm the findings of this literature, thereby reinforcing the motivation of incorporating realistic bank heterogeneity into benchmark macro-finance models.

Our HBANK framework relates to the influential HANK class of models that study fiscal and monetary policy in non-Ricardian environments with nominal rigidities (McKay and Reis, 2016; Kaplan et al., 2018; Auclert et al., 2024b).[4] The standard HANK framework for monetary policy analysis features heterogeneity in the household sector (Werning, 2015; McKay et al., 2016; Auclert, 2019; Luetticke, 2021; Kekre and Lenel, 2022; Wolf, 2025). Some HANK models allow for financial frictions and a banking sector (Faccini et al., 2024). Many papers also study monetary policy and heterogeneity on the side of non-financial firms. For example, Ottonello and Winberry (2020) quantify the investment channel of monetary policy with firms that differ in risk and distance to default. Jeenas (2024) studies the role of firm liquidity for monetary policy and investment. Baqaee et

[2]See also Gerali et al. (2010), Ríos-Rull et al. (2020), Abadi et al. (2023), Abad et al. (2024), and Dempsey (2024).

[3]See also Bocola (2016), Nuño and Thomas (2017), Gertler et al. (2016, 2019), Boissay et al. (2016), Bigio and Sannikov (2021), Elenev et al. (2021), Mendicino et al. (2024), and Begenau et al. (2025).

[4]See Auclert et al. (2024a) and Auclert (2025) for the recent literature review as well as practical applications.

al. (2024) study the supply side of monetary policy in a model with heterogeneous firms and endogenous product market power. González et al. (2023) study—both positively and normatively—a NK model with heterogeneous firms and financial frictions. Our contribution is to zoom in both empirically and quantitatively on the impacts of *bank* balance sheet heterogeneity and default risk channels of monetary transmission in a NK model with frictional and risky financial intermediation.

The aforementioned HANK literature places a lot of emphasis on the careful decomposition of aggregate effects of monetary policy into direct (intertemporal substitution) and indirect (multiplier, fiscal reaction) channels. In the standard HANK model, the fiscal reaction to monetary policy shocks is a critical component of the indirect effect that determines the total response to a monetary shock (Kaplan et al., 2018). In HBANK, we characterize a novel indirect channel of monetary policy—the prudential policy response. The automatic prudential reaction to a monetary policy contraction tightens bank leverage regulation, thereby impacting real economic activity through the bank lending and default risk channels. We emphasize prudential policy as a potentially significant indirect channel of monetary transmission in the HANK class of models.

Finally, methodologically, our study belongs to the rapidly growing set of papers that solve complex general equilibrium macroeconomic models with sequence-space methods (Mankiw and Reis, 2007; Boppart et al., 2018; Auclert et al., 2021a). The sequence-space approach has been successfully applied to the case of household heterogeneity (Auclert et al., 2020, 2024b), firm heterogeneity (González et al., 2023), input-output frameworks (Schaab and Tan, 2023), open-economy environments (Auclert et al., 2021b; Aggarwal et al., 2023), and optimal policy (Davila and Schaab, 2023). To the best of our knowledge, ours is the first study that solves in sequence space a NK macro-banking framework with heterogeneous intermediaries. In doing so, we summarize the bank lending channel of monetary policy in a compact set of just two sufficient statistics: the partial-equilibrium response to a policy shock and the general-equilibrium adjustment inclusive of the systematic prudential reaction.

2 A New-Keynesian Model with Heterogeneous Banks

This section presents our baseline HBANK framework. Time $t \geq 0$ is discrete. There are five agents in the economy: a capital good producer, a final good producer, a continuum of differentiated retailers, a representative household, and a continuum of measure unity of heterogeneous banks that are indexed by j. There is no aggregate uncertainty. In Section 3, we cast our model in sequence space to study transition dynamics following unexpected aggregate demand and supply shocks.[5]

2.1 Banks

Balance sheet. The balance sheet of each bank j consists of three items. First, on the asset side, banks provide funding to capital good producers in the form of claims on the capital stock, $l_{j,t}$, that are priced at Q_t. Second, on the liability side, banks can acquire uninsured time deposits from households, $b_{j,t}$, that pay the gross interest rate $R^b_{j,t}$. Finally, banks accumulate net worth, $n_{j,t}$. The balance sheet constraint of bank j reads:

$$b_{j,t} + n_{j,t} = Q_t l_{j,t}. \tag{1}$$

In order to operate the franchise, each bank must also incur non-interest expenses that are governed by $\{\zeta_1, \zeta_2\}$.

Idiosyncratic risk. In exchange for purchasing claims on capital, banks receive a bank-specific return, $R^T_{j,t}$, which is given by the realized aggregate return on the capital stock, R^k_{t+1}, perturbed by an idiosyncratic component $\xi_{j,t}$. Financial markets are incomplete, and $\xi_{j,t}$ represents uninsured return risk in the spirit of Benhabib and Bisin (2018) and Benhabib et al. (2019). The bank-specific return, $R^T_{j,t}$, is determined as follows:

$$R^T_{j,t} = \xi_{j,t} R^k_t \tag{2}$$

where $\xi_{j,t}$ follows an AR(1) process: $\xi_{j,t} = \rho_\xi \bar{\xi} + (1 - \rho_\xi)\xi_{j,t-1} + \epsilon_{j,t}$, with $\epsilon_{j,t} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2_\xi)$. Thus, banks are *ex-ante* identical but heterogeneous *ex-post* due to persistent idiosyncratic shocks. The law of motion of bank net worth can now be written as follows:

$$n_{j,t+1} = R^T_{j,t+1} Q_t l_{j,t} - R^b_{j,t+1} b_{j,t} - \zeta_1 l_{j,t}^{\zeta_2} \tag{3}$$

Breaking scale invariance. As long as $\zeta_2 > 1$, and non-interest expenditures are convex, the scale invariance property that is generally inherent to this class of models (Gertler and Kiyotaki, 2010; Gertler and Karadi, 2011) is eliminated. *Scale variance*, instead, makes bank-level net worth, $n_{j,t}$, a relevant state variable. In equilibrium, this will generate an

[5] As shown by Boppart et al. (2018) and Auclert et al. (2021a), as long as so-called "MIT shocks" are not too large, this approach is equivalent to computing impulse responses by first-order perturbation in the model with aggregate uncertainty.

endogenous distribution of net worth along with other bank-specific characteristics such as assets $l_{j,t}$ or deposits $b_{j,t}$.

Leverage constraint. In order to motivate prudential regulation and to introduce a hard limit on the leverage multiple, we impose the following constraint on risk-taking:[6]

$$\lambda_{j,t} Q_t l_{j,t} \leq V_{j,t} \tag{4}$$

where $V_{j,t}$ is the franchise value of bank j, and $0 < \lambda_{j,t} < 1$ is the *prudential policy* instrument that targets the leverage of bank j. Importantly, observe that $\lambda_{j,t}$ varies both across time *and* in the cross-section. In the standard model, λ is a parameter. In HBANK, $\lambda_{j,t}$ is a policy choice. It will be set systematically as a function of the aggregate state as well as bank-specific characteristics following a specified rule. Section 2.4 describes this rule in detail.

Finally, in every period, with an exogenous probability $(1-\sigma)$ banks exit the economy and their franchise value, $V_{j,t}$, gets transferred to the household in the form of lump-sum dividends (Gertler and Kiyotaki, 2010). Upon exit, any remaining accumulated net worth gets transferred to the household. All non-interest expenses are also rebated back to the household in the form of lump-sum payments. Banks are risk-neutral and cannot operate with negative net worth.

Costly default. Because markets are incomplete and banks face uninsured idiosyncratic return risk, upon drawing a sufficiently large negative $\xi_{j,t}$, a bank can be forced to default. Insolvency risk constitutes the endogenous component of bank exit. Since banks cannot operate with negative equity, the *ex-ante* probability of hitting the zero-net-worth bound can be defined as:

$$\varphi_{j,t} = \mathbb{E}_t\Big(\Pr\big(n_{j,t+1} < 0\big)\Big) \tag{5}$$

Furthermore, we assume that each individual bank default is costly in terms of real resource units. In particular, the expected bank-specific cost of default in terms of lost assets is given by:

$$s_{j,t} = \omega_1 \varphi_{j,t} l_{j,t}^{\omega_2} \tag{6}$$

where ω_1 is the fraction of bank assets that cannot be recovered, on average, conditional on default, and $\omega_2 > 1$.[7]

[6] This constraint is also often motivated with moral hazard frictions in the deposit market (Gertler and Kiyotaki, 2010; Gertler and Karadi, 2011).

[7] Empirical studies estimate that the average FDIC loss from bank failure in the U.S. is around 28%,

Importantly, because $\omega_2 > 1$, the cost of default is convex in bank size. This parsimonious relationship captures systemic-risk features of any modern banking sector. The macroeconomic consequence of a collapse of a small, regional credit union is insignificant. However, failure of a systematically important financial institution can have large, persistent real-economy implications. In Section 7, we will present empirical evidence that directly supports the convex relationship between $l_{j,t}$ and $s_{j,t}$.

The total, economy-wide expected default cost, in terms of units of bank assets, is:

$$S_t = \int s_{j,t} dj \equiv \int \omega_1 \varphi_{j,t} l_{j,t}^{\omega_2} dj. \tag{7}$$

In terms of units of the final good, the loss of resources due to insolvency is $\underline{Y}_t = \psi S_t$. ψ is a free parameter that will be used to target the ratio $\tilde{S}_t \equiv \frac{\underline{Y}_t}{Y_t}$ of default costs to total output with empirical estimates on the long-run macroeconomic costs of financial crises and panics. In our baseline calibration, the targeted $\tilde{S}_t$ will be 10% per year in the steady state.

The realized return on aggregate capital, which all banks take as given, can then be written as:

$$R_{t+1}^k = \frac{(1-\tilde{S}_t)\alpha A_{t+1} K_{t+1}^{\alpha-1} H_{t+1}^{1-\alpha}}{Q_t} \tag{8}$$

where, as explained in detail further below, K_t and H_t are aggregate capital and labor supply, respectively, A_t is aggregate productivity, and Q_t is the price of capital.

Equation 8 is a key condition that describes the feedback between aggregate financial instability and the banking sector. In response to an exogenous disturbance, banks adjust their balance sheet size and the leverage multiple. This impacts the probability of default, $\varphi_{j,t}$, and its realized cost, $s_{j,t}$. The cross-section of bank-specific choices aggregates into S_t which, through the return on capital, affects the bank lending problem again. In equilibrium, the distributions of both $s_{j,t}$ and $\varphi_{j,t}$ will be systematically related to the state variable, $n_{j,t}$. The size-risk relationship will also define key empirical tests of the model in Section 7.

Pricing uninsured deposits. Deposits are not insured.[8] The distribution of bank-specific insolvency risk is priced competitively into the cross-section of deposit rates according to an asset pricing condition:

$$1 = \left[(1-\varphi_{j,t})\mathbb{E}_t(\Lambda_{t+1}|\textit{no default}) + \varphi_{j,t}(1-\omega_1)\mathbb{E}_t(\Lambda_{t+1}|\textit{default})\right] \times R_{j,t+1}^b \tag{9}$$

which is how we will parameterize ω_1 (Granja et al., 2017).

[8] In practice, uninsured depositors comprise around half of all deposits in the U.S. (Egan et al., 2017).

where Λ_{t+1} is the stochastic discount factor of the household and $\mathbb{E}_t(.|.)$ is the expectation conditional on either default or no default in period $t+1$. The deposit rate is essentially a weighted average of the risk-less return in the state of non-default and the partial recovery rate, $1-\omega_1$, in the state of default. Notice that in the absence of insolvency risk, one recovers the standard Euler equation for the risk-free interest rate. In our framework, however, there is generally a spread between the risk-free rate and the interest rate on deposits, $R^b_{j,t}$, that precisely captures insolvency risk.[9]

Dynamic bank lending problem. We now adopt a recursive notation in order to summarize the dynamic optimization problem of an individual bank. The idiosyncratic state vector includes bank-specific net worth, n, and the idiosyncratic return draw, ξ. The problem is as follows:

$$V_t(n,\xi) = \max_{\{l,b,n'\}\geq 0} \mathbb{E}_t\Big\{\beta\Big[\Big(1-\varphi_t(n,\xi)\Big)\Big((1-\sigma)n' + \sigma V_{t+1}(n',\xi'|\xi)\Big)\Big]\Big\} \tag{10}$$

subject to:

$$\begin{aligned}
n' &= \mathbb{E}_t\left[R^k_{t+1}\xi'\right]Q_t l - R^b_{t+1}(n,\xi)b - \zeta_1 l^{\zeta_2} \\
b + n &= Q_t l \\
\lambda_t(n,\xi)Q_t l &\leq V_t(n,\xi) \\
1 &= \Big[(1-\varphi_t(n,\xi))\mathbb{E}_t(\Lambda_{t+1}) + \varphi_t(n,\xi)(1-\omega_1)\mathbb{E}_t(\Lambda_{t+1})\Big]R^b_{t+1}(n,\xi) \\
\xi' &= \rho_\xi\bar{\xi} + (1-\rho_\xi)\xi + \epsilon'
\end{aligned}$$

Marginal propensity to lend. To provide more intuition on the workings of the bank lending decision, we now consider a partial-equilibrium solution to the above problem. For now, the bank takes all aggregate quantities and prices as given. After substituting out the balance sheet constraint and the deposit pricing condition, the problem becomes:

$$V_t(n,\xi) = \max_{l\geq 0}\Big\{\mathbb{E}_t\Omega_{t+1}\Big[\Big(R^k_{t+1}\xi' - R^b_{t+1}(n,\xi)\Big)Q_t l - \zeta_1 l^{\zeta_2} + R^b_{t+1}(n,\xi)n\Big]\Big\}$$

[9] Note that, while banks understand the asset-pricing link between $R^b_{j,t}$ and $\varphi_{j,t}$, they do not internalize the impact of their individual choices either on aggregate riskiness of the economy, S_t, or on profitability, R^k_t. This generates an externality akin to the one emphasized in Di Tella (2019).

subject to:

$$
\begin{aligned}
&\lambda_t(n,\xi)Q_t \le V_t(n,\xi)\\
&1 = \Big[(1-\varphi_t(n,\xi))\mathbb{E}_t(\Lambda_{t+1}) + \varphi_t(n,\xi)(1-\omega_1)\mathbb{E}_t(\Lambda_{t+1})\Big]R^b_{t+1}(n,\xi)\\
&\xi' = \rho_\xi\bar{\xi} + (1-\rho_\xi)\xi + \epsilon'
\end{aligned}
$$

where $\Omega_{t+1} \equiv \Big(1-\varphi_t(n,\xi)\Big)\beta\Big(1-\sigma+\sigma\frac{V_{t+1}(n',\xi'|\xi)}{n'}\Big)$ is the bankers' augmented stochastic discount factor. Because banks are risk-neutral, they always lever up until the leverage constraint binds. Thus, the lending policy function can be computed, as an implicit function of the choice variable, as follows:

$$
l^*_t(n,\xi) = \frac{\mathbb{E}_t\left\{\Omega_{t+1}\left(R^b_{t+1}(n,\xi)n - \zeta_1 l^{\zeta_2}\right)\right\}}{Q_t\Big(\lambda_t(n,\xi) - \mathbb{E}_t\Big\{\Omega_{t+1}\Big(R^k_{t+1}\xi' - R^b_{t+1}(n,\xi)\Big)\Big\}\Big)} \tag{11}
$$

The policy function for lending succinctly summarizes the partial-equilibrium solution of the model. First, conditional on the aggregate state of the world, the bank-level lending choice is increasing in the risk premium, which is in the denominator of (11). Limits to arbitrage make private leverage an increasing function of excess returns. Second, the prudential constraint parameter, λ, lowers credit supply. This important relationship is behind the workings of automatic prudential policies that we discuss later. Third and finally, high bank-specific default risk, φ, through the discount factor channel, also reduces loan supply.

Now, it is clear from (11) that banks with different size-risk profiles will choose heterogeneous *levels* of claims. However, it is not yet obvious if the *elasticity* of the lending decision varies systematically with bank-level characteristics. Following Jamilov and Monacelli (2025), we define the bank-specific Marginal Propensity to Lend (MPL) as the marginal change in lending in response to a change in net worth. The MPL can be calculated as follows:

$$
MPL_t(n,\xi) = \frac{\mathbb{E}_t\left\{\Omega_{t+1}R^b_{t+1}(n,\xi)\right\}}{Q_t\Big(\lambda_t(n,\xi) - \mathbb{E}_t\Big\{\Omega_{t+1}\Big(R^k_{t+1}\xi' - R^b_{t+1}(n,\xi)\Big) + \zeta_1\zeta_2 l^{\zeta_2-1}\Big\}\Big)} \tag{12}
$$

The MPL object captures the elasticity of bank-level lending choices to changes in bank-level net worth. These changes can be induced by any underlying aggregate shock, such as an interest rate change. From (12), we see that the MPL is heterogeneous and is a function of four bank-specific characteristics: size, n, the idiosyncratic return draw, ξ, the probability of default, φ (through the SDF), and the retail deposit rate, R^b. Thus, our model can feature rich heterogeneity in the *responsiveness* to aggregate shocks. Recall, that it is precisely this heterogeneity that the empirical literature on the credit channel of monetary policy emphasizes.

The partial-equilibrium banking problem can be readily solved in isolation. However, the aggregate return on capital, R^k, deposit and capital prices, as well as the entire cross-sectional distribution, must be determined in general equilibrium. We now discuss the problems of firms and households, whose behavior is purposely kept simple in order to keep the spotlight on the banks.

2.2 Firms

Capital good producers. Capital is required for the production of the final good, which is in turn consumed by the household. There is a continuum of measure one of competitive capital good producers that are indexed by z. These firms are cash-strapped and require external financing to finance new investment. Firms borrow exclusively from banks in the form of state-contingent, equity-type claims on the end-of-period return on aggregate capital.[10] Let L_t be the total demand for bank credit. Credit is fully intermediated by the banking sector, such that $L_t = \int_0^1 l_{j,t} dj$ and $K_{t+1} = L_t$ is the evolution of capital in the economy.[11]

On the supply side of the economy, new capital is formed with $K_{t+1} = \Phi(I_t)$, where I_t is investment, with $\Phi(\cdot)' > 0$ and $\Phi(\cdot)'' < 0$. Each firm z solves the following problem:

$$\max_{I_{z,t}} Q_t \Phi(I_{z,t}) - I_{z,t} \tag{13}$$

The above problem is symmetric and the price of capital, Q_t, is determined following the standard Tobin's Q optimality condition: $Q_t = [\Phi'(I_t)]^{-1}$. In equilibrium, the cross-section of bank-level loans, $\int l_{j,t}$, pins down the aggregate demand for capital and its price. Capital depreciates fully every period, and new capital always equates the total amount of claims intermediated by the banks.

New Keynesian block. Non-financial firms consist of a final good producer and of a continuum of differentiated retailers, indexed by $i \in [0,1]$, that produce intermediate

[10] We are interchangeably referring to these claims as loans or credit.

[11] It is also possible to allow households to intermediate a fraction of the market for claims, as is done, for example, in Gertler et al. (2019). As long as banks, on the margin, are more effective at managing risky investments on behalf of the household, this extension would not materially affect any of our conclusions.

goods. Differentiated goods produced by retailers are aggregated into the final good by the final good producer:

$$Y_t = \left(\int_0^1 y_{i,t}^{\frac{\gamma-1}{\gamma}} di\right)^{\frac{\gamma}{\gamma-1}} \tag{14}$$

where $\gamma > 1$ is the elasticity of substitution between differentiated varieties. Each retailer i rents labor $H_{i,t}$ and capital $K_{i,t}$ to produce intermediate goods using a constant returns to scale production technology:

$$y_{i,t} = A_t K_{i,t}^{\alpha} H_{i,t}^{1-\alpha} \tag{15}$$

where $0 < \alpha < 1$. Retailers set a relative price for their variety, $p_{i,t}$, and pay quadratic price adjustment costs $\frac{\theta}{2}\left(\frac{p_{i,t}}{p_{i,t-1}} - 1\right)^2 Y_t$. The demand function for each retailer is: $y_{i,t} = \left(\frac{p_{i,t}}{P_t}\right)^{-\gamma} Y_t$, where $P_t = \left(\int_0^1 p_{i,t}^{1-\gamma} di\right)^{\frac{1}{1-\gamma}}$ is the aggregate price index. Cost minimization yields the following expression for the (common) nominal marginal cost: $MC_t = \frac{1}{A_t}\left(\frac{w_t}{1-\alpha}\right)^{1-\alpha}\left(\frac{Z_t}{\alpha}\right)^{\alpha}$, where Z_t is the rental cost of capital and w_t is the real wage rate.

Retailers' symmetrical problem yields the familiar New Keynesian Phillips Curve relationship that links current inflation to the future expected inflation and current deviations of the real marginal cost from its desired steady-state value:

$$\log \Pi_t = \frac{\gamma - 1}{\theta}(\log MC_t - \log MC_{ss}) + E_t\left[\Lambda_{t+1} \log \Pi_{t+1}\right] \tag{16}$$

where x_{ss} denotes variables in the steady state. The Phillips Curve connects the NK block with the cross-section of banks via the relative price p_t. In response to interest rate changes, banks' cost of capital shifts. Banks react by adjusting their supply of credit to the capital good producer. This, in turn, alters the rental cost of capital and, via the marginal cost and due to nominal rigidities, inflation. This is, in a nutshell, the bank lending channel of monetary policy in our environment. The complication is that we must keep track of (i) the full distribution of bank-level decisions as well as (ii) endogenous financial stability considerations.

2.3 Representative Household

The household discounts the future with $\beta \in (0, 1)$, and derives utility from consumption, C_t, as well as disutility from labor, H_t. Preferences are given by:

$$\mathbb{E}_t \sum_{s=0}^{\infty} \beta^s \mathcal{U}\left(C_{t+s}, H_{t+s}\right) \tag{17}$$

The period utility function is of CRRA form and features intra-temporal non-separability between consumption and hours following Greenwood et al. (1988):

$$\mathcal{U}(C_t, H_t) = \log\left(C_t - \chi_1 \frac{H_t^{1+\chi_2}}{1+\chi_2}\right) \tag{18}$$

The consumer maximizes the discounted stream of utility subject to the sequence of budget constraints:

$$C_t + \int_0^1 b_{j,t} dj \leq H_t W_t + \int_0^1 R^b_{j,t} b_{j,t-1} dj + Div_t + T_t \tag{19}$$

where W_t is the competitive wage rate, Div_t are lump-sum transfers of bank dividends, and T_t are any remaining lump-sum transfers/taxes from the government.

2.4 Monetary and Prudential Policy

Monetary authority. The central bank sets the nominal interest rate on a zero-net-supply riskless bond following a rule in the spirit of Taylor (1993):

$$r^N_t = \bar{r} + \varphi_\pi \pi_t + v_t \tag{20}$$

where $\bar{r} = \frac{1}{\beta} - 1$ is the steady state net real interest rate,[12] r^N_t is the net nominal rate, $v_t \sim \mathcal{N}(0, \sigma^2_m)$ is the monetary shock, and $\varphi_\pi > 1$ is the weight on inflation. In the steady state, $v_t = 0$. The real interest rate is determined by the Fisher equation: $1 + r_t = \frac{1+r^N_t}{1+\pi_{t+1}}$, where r_t and π_{t+1} are the net real rate and the net inflation rate, respectively.

The nominal interest rate and the stochastic discount factor are linked via the Euler equation for bonds:

$$\Lambda_{t+1} \frac{R^N_t}{\Pi_{t+1}} = 1 \tag{21}$$

Prudential policy rule. Prudential regulation follows a systematic rule. Importantly, we allow for *micro*-prudential policy that sets the prudential instrument $\lambda_{j,t}$ for every bank in the distribution. The regulator observes the cross-section of default costs, $s_{j,t}$, in any given period and sets policy in order to respond to deviations of $s_{j,t}$ from the steady state,

[12] We focus on a zero-inflation steady state.

s_j. The rule takes the following form:

$$\lambda_{j,t+1} = \lambda_j \left(\frac{s_{j,t+1}}{s_j} \right)^{\phi}, \quad \phi > 0 \tag{22}$$

with $\phi > 0$. Thus, any bank whose $s_{j,t}$ is above its own steady-state value will face a micro-prudential tightening. We choose $s_{j,t}$, i.e., default costs, as the target variable since this is the relevant measure of financial instability in the model. Substituting it with the probability of insolvency, $\varphi_{j,t}$, yields qualitatively similar results because, as we will see later in the paper, the aggregate likelihood and cost of default tend to co-move in response to shocks.

Quantitatively, we will be assuming that—in response to any aggregate shock that shifts default costs from the steady state—the regulator will change $\lambda_{j,t}$ on a specific subset of the banking distribution. We refer to this as *micro*-prudential policy. In particular, we will be denoting by micro-pru *large* a policy change that affects only banks in the top 25% of bank assets, i.e., the largest institutions. In contrast, micro-pru *small* will be referencing a policy change that impacts only banks in the bottom 75% of bank assets, i.e. excluding the largest institutions. Finally, a policy change that affects all banks in the distribution will be labeled as *macro*-prudential.

2.5 Market Clearing and Equilibrium

To close the model, we now impose clearing conditions in several markets. First, the credit market clears:

$$\int n_t^*(n,\xi)d\Gamma_{t-1}(n,\xi) + \int b_t^*(n,\xi)d\Gamma_t(n,\xi) = Q_t \int l_t^*(n,\xi)d\Gamma_t(n,\xi) \tag{23}$$

where variables x^* signify optimal choices and $\Gamma_t(n,\xi)$ denotes the measure giving the joint cross-sectional distribution of banks over the two idiosyncratic states, net worth and returns. Second, the market for deposit savings must clear:

$$B_{h,t+1} = \int b_t^*(n,\xi)d\Gamma_t(n,\xi) \tag{24}$$

where $B_{h,t+1}$ is the aggregate supply of deposits from the household. Third, the market for capital must clear:

$$K_{t+1} = \int l_t^*(n,\xi)d\Gamma_t(n,\xi) \tag{25}$$

Finally, the goods market clearing condition implicitly defines aggregate consumption:

$$Y_t = C_t + \underline{Y}_t + \Theta_t \tag{26}$$

where Θ_t are retailers' price adjustment costs and $\underline{Y}_t$ denotes the costs of bank default in units of the final good. Finally, the labor market clears by Walras law.

Given an initial distribution Γ_0, the equilibrium is defined by a sequence of value functions $\{V_t(n, \xi)\}$; bank decision rules $\{l_t(n, \xi),\ b_t(n, \xi),\ n_{t+1}(n, \xi)\}$; measure of banks $\{\Gamma_t\}$; deposit rate schedules $\{R^b_t(n, \xi)\}$; bank default probabilities $\{\varphi_t(n, \xi)\}$; and aggregate prices $\{p_t, \Pi_t, Q_t, W_t, R^k_t, R^N_t, R_t, \Lambda_t\}$ such that (i) banks, firms, and household optimize, (ii) deposit rates, the wage rate, the rental rate of capital, and default risk are priced competitively, (iii) the distribution of banks is consistent with decision rules, (iv) the monetary and macro-prudential authorities follow their respective policy rules, and (v) all markets clear. For the steady state, we consider a stationary distribution in which all aggregate variables, including the measure Γ, are time-invariant.

3 Sequence-Space Representation

To study aggregate transition dynamics, we cast our framework in the sequence-space domain by leveraging the methodological contributions from Boppart et al. (2018) and Auclert et al. (2021a). We show how the full general-equilibrium macroeconomic response to a monetary policy shock can be tractably summarized with a small number of sufficient statistics following the insights from Auclert (2019). We then discuss the workings of the bank lending channel of monetary policy through the lenses of these sufficient statistics.

From (10), it is clear how the banks' policy functions in period t are fully pinned down by the future paths $\{r^k_s, r_s, \lambda_s\}^{\infty}_{s=t}$. To solve for the full sequence of the lending choice, $\{l_s\}^{\infty}_{s=t}$, one requires the following three aggregate time-varying sequences: the return on aggregate capital, the *real* interest rate, and prudential policy. Every other component of (10) is determined as part of the dynamic bank lending problem. The deposit rate is pinned down by (9). The default risk probability, which is necessary to know the deposit rate, as per (5), depends on the forecast of next-period net worth, which is known at the time lending is determined. Bank deposits can always be recovered from the balance sheet constraint, and the price of capital is independent of bank-specific characteristics. Thus, the three sequences $\{r^k_s, r_s, \lambda_s\}^{\infty}_{s=t}$ are necessary and sufficient to solve for the bank lending choice at time t. Given the banks' sequence of policy functions and an initial condition for the banking distribution Γ_t, one can then recover the full sequence of aggregate bank loans, deposits, and net worth.

The sequence for the real interest rate, $\{r_s\}_{s=t}^{\infty}$, is determined endogenously by the combination of the nominal interest rule (20) and the New Keynesian Phillips Curve. In addition, $\{r_s^k\}_{s=t}^{\infty}$ depends on the sequences of aggregate capital, the price of capital, labor supply, and the default cost. Finally, recall that the prudential policy parameter λ is a time-varying policy choice. Thus, it is also necessary to pin down bank lending at date t.

3.1 Aggregate Lending and Default Functions

Next, we define an aggregate lending function $\mathcal{L}_t(.)$ as a mapping from input sequences into aggregate lending at date t, L_t. Starting from the steady-state distribution $\Gamma_0 = \Gamma_{ss}$, aggregate lending can be expressed as follows:

$$L_t = \mathcal{L}_t\Big(\Big\{r_s^k(K_s, Q_s, S_s, H_s), r_s, \lambda_s\Big\}_{s=0}^{\infty}\Big) \tag{27}$$

Similarly, we define the aggregate default cost function that maps input sequences into the default cost at date t: $S_t = \mathcal{S}_t\Big(\Big\{r_s^k(K_s, Q_s, S_s, H_s), r_s, \lambda_s\Big\}_{s=0}^{\infty}\Big)$. Computing S_t via $\mathcal{S}_t$ is necessary in order to back out aggregate consumption from the resource constraint. Note that all equilibrium interactions between *heterogeneous* banks are always implied in the functions $\mathcal{L}_t$ and $\mathcal{S}_t$, which map aggregate sequences into other aggregate sequences, the key advantage of the sequence-space method (Auclert et al., 2021a).

Equation (27) is very general. A stripped down version can be obtained in a few simple steps. First, as in the standard macro-banking model, set λ as a parameter, i.e invariant across time and the cross-section. Second, shut down costly default. Third, assume that labor is supplied inelastically and, for example, normalized to unity. Fourth and finally, suppose that the capital good producer's problem is such that the price of capital is always unity. In such simplified case, the lending function relationship becomes: $L_t = \mathcal{L}_t\Big(\Big\{r_s^k(K_s), r_s\Big\}_{s=0}^{\infty}\Big)$.

3.2 Linearization

To solve for the general-equilibrium macroeconomic response to shocks, we recall the market clearing condition for capital (25). Combined with (27), the aggregate lending function becomes:

$$K_{t+1} = \mathcal{L}_t\Big(\Big\{r_s^k(K_s, Q_s, S_s, H_s), r_s, \lambda_s\Big\}_{s=0}^{\infty}\Big) \tag{28}$$

Equation (28) contains a fixed point in aggregate capital. After imposing a number of standard technical assumptions, we linearize (28) and obtain:[13]

$$\mathrm{d}\mathbf{K} = \left(\mathbf{I} - \mathbf{F}_K\right)^{-1}\left(\mathbf{F}_r \mathrm{d}\mathbf{r} + \mathbf{F}_\lambda \mathrm{d}\boldsymbol{\lambda}\right) \tag{29}$$

where bold-face letters denote infinite vectors of deviations from steady state, e.g., $\mathrm{d}\mathbf{K} = (\mathrm{d}K_0, \mathrm{d}K_1, \dots)$, and entries of $\mathbf{F}_K$ are: $[\mathbf{F}_K]_{t,s} = \frac{\partial \mathcal{L}_t}{\partial r^k_{s+1}}\left(\frac{\partial r^k_{s+1}}{\partial K_s} + \frac{\partial r^k_{s+1}}{\partial Q_s}\frac{\partial Q_s}{\partial K_s} + \frac{\partial r^k_{s+1}}{\partial S_s}\frac{\partial S_s}{\partial K_s} + \frac{\partial r^k_{s+1}}{\partial H_s}\frac{\partial H_s}{\partial K_s}\right)$, entries of $\mathbf{F}_r$ are: $[\mathbf{F}_r]_{t,s} = \frac{\partial \mathcal{L}_t}{\partial r_{s+1}}$, entries of $\mathbf{F}_\lambda$ are $[\mathbf{F}_\lambda]_{t,s} = \frac{\partial \mathcal{L}_t}{\partial \lambda_s}$, and $\mathbf{L}$ is a lag operator. These matrices capture partial-equilibrium responses at time t (corresponding to each row) to exogenous shocks to at horizon s (corresponding to each column). For example, entry $[F_r]_{0,0}$ is the impact response of bank lending to an unanticipated increase in the real rate. Generally, only four of these objects need to be computed numerically, namely the derivatives, or *Jacobians*, $\frac{\partial \mathcal{L}_t}{\partial r^k_{s+1}}, \frac{\partial \mathcal{L}_t}{\partial r_{s+1}}, \frac{\partial \mathcal{L}_t}{\partial \lambda_s}, \frac{\partial S_s}{\partial K_s}$. All the remaining derivatives can be solved for analytically around the steady state.

Once the equilibrium path of capital $\mathrm{d}\mathbf{K}$ is computed, we can construct the response of bank default costs as follows:

$$\mathrm{d}\mathbf{S} = \mathbf{X}_K \mathrm{d}\mathbf{K} + \mathbf{X}_r \mathrm{d}\mathbf{r} + \mathbf{X}_\lambda \mathrm{d}\boldsymbol{\lambda} \tag{30}$$

where entries of $\mathbf{X}_K$ are: $[\mathbf{X}_K]_{t,s} = \frac{\partial \mathcal{S}_t}{\partial r^k_{s+1}}\left(\frac{\partial r^k_{s+1}}{\partial K_s} + \frac{\partial r^k_{s+1}}{\partial Q_s}\frac{\partial Q_s}{\partial K_s} + \frac{\partial r^k_{s+1}}{\partial S_s}\frac{\partial S_s}{\partial K_s} + \frac{\partial r^k_{s+1}}{\partial H_s}\frac{\partial H_s}{\partial K_s}\right)$, entries of $\mathbf{X}_r$ are: $[\mathbf{X}_r]_{t,s} = \frac{\partial \mathcal{S}_t}{\partial r_{s+1}}$, and entries of $\mathbf{X}_\lambda$ are $[\mathbf{X}_\lambda]_{t,s} = \frac{\partial \mathcal{S}_t}{\partial \lambda_s}$. The following Jacobians need to be computed numerically: $\frac{\partial \mathcal{S}_t}{\partial r^k_{s+1}}, \frac{\partial \mathcal{S}_t}{\partial r_{s+1}}$, and $\frac{\partial \mathcal{S}_t}{\partial \lambda_s}$. The general-equilibrium path $\mathrm{d}\mathbf{K}$ and all the analytically-computed derivatives are unchanged.

Interest rate and the NK Phillips Curve. The path $\mathrm{d}\mathbf{r}$ requires an inner fixed point, as mentioned previously, because of nominal rigidities and the nominal interest rate rule. Recall that the real interest rate is determined by a Fisher equation, which in vector notation is:

$$\mathrm{d}\mathbf{r} = \left(\phi_\pi \mathbf{I} - \mathbf{J}\right)\mathrm{d}\boldsymbol{\pi} \tag{31}$$

where $\mathbf{J}$ is the lead operator. The inflation sequence $\mathrm{d}\boldsymbol{\pi}$ is governed by the New Keynesian block from Section 2.2 and takes the vector form of $\mathrm{d}\boldsymbol{\pi} = \kappa \mathrm{d}\mathbf{M} + \beta \mathbf{Q}\mathrm{d}\boldsymbol{\pi}$ with $\kappa \equiv \frac{\gamma - 1}{a}$ and

[13] Following Auclert et al. (2021a), we assume that the economy is initially at the stationary steady state. Furthermore, we consider only bounded perturbations in capital, $\mathrm{d}\mathbf{K}$, the real rate, $\mathrm{d}\mathbf{r}$, price of capital, $\mathrm{d}\mathbf{Q}$, labor supply, $\mathrm{d}\mathbf{H}$, default costs, $\mathrm{d}\mathbf{S}$, and prudential policy, $\mathrm{d}\boldsymbol{\lambda}$. In addition, we also assume that the lending and default functions are differentiable around the steady state.

d$\mathbf{M}$ denoting the sequence of (log) marginal costs. Following Auclert et al. (2023), this yields the New Keynesian Phillips curve that, in the sequence space, reads:

$$\mathrm{d}\boldsymbol{\pi} = \mathbf{P}\mathrm{d}\mathbf{M} \tag{32}$$

with

$$\mathbf{P} \equiv \kappa \begin{pmatrix} 1 & \beta & \beta^2 & \ddots \\ 0 & 1 & \beta & \ddots \\ 0 & 0 & 1 & \ddots \\ \ddots & \ddots & \ddots & \ddots \end{pmatrix} \tag{33}$$

Combining equations (31) and (32) delivers the endogenous response of the real rate: $\mathrm{d}\mathbf{r} = \left(\phi_\pi \mathbf{I} - \mathbf{J}\right)\mathbf{P}\mathrm{d}\mathbf{M}$.

3.3 Equilibrium Construction

We finalize this section by emphasizing the force of the sufficient statistics approach in our framework. Equations (29) and (30), along with the paths of d$\mathbf{r}$ and d$\boldsymbol{\lambda}$ are sufficient to construct every other object in the economy. All complex cross-sectional distributions of the banking sector, the NK block, behaviors of firms and households, are succinctly summarized in these simple impulse response functions. Below, we sketch an algorithm of how to recover other variables of interest once the sufficient statistics are known.

First, start by computing the Jacobians $\mathbf{F}_\mathrm{K}$, $\mathbf{F}_\mathrm{r}$, $\mathbf{F}_\lambda$ and $\mathbf{X}_\mathrm{K}$, $\mathbf{X}_\mathrm{r}$, $\mathbf{X}_\lambda$. These objects are model-specific and need to be computed just once. Second, compute the sequence of capital, d$\mathbf{K}$, given initial guesses for d$\mathbf{r}$ and d$\boldsymbol{\lambda}$. Given d$\mathbf{K}$, compute the sequence of default costs, d$\mathbf{S}$. Solve for d$\mathbf{r}$ and d$\boldsymbol{\lambda}$ as fixed points. Third, given the *general-equilibrium* sequences of d$\mathbf{K}$, d$\mathbf{S}$, d$\mathbf{r}$, and d$\boldsymbol{\lambda}$ recover every other object of interest. The price of capital, d$\mathbf{Q}$ can be computed from the capital producer block, given d$\mathbf{K}$. Recover labor supply, d$\mathbf{H}$, from the GHH structure. Compute the response of aggregate output, d$\mathbf{Y}$, using the responses of capital and labor. Build the responses of the marginal cost, the real wage, and the inflation rate. Leverage the NK Phillips Curve in its sequence-space form. Finally, recover aggregate consumption, d$\mathbf{C}$, net of banks' default costs and retailers' price adjustment costs.

4 Monetary Policy Transmission

In this section, we explain the workings of the monetary transmission mechanism in our model.

4.1 Bank Lending Block

Equations (29) and (30) are the key equations we need in order to understand the monetary transmission mechanism in our setup. Those equations summarize the general-equilibrium economic responses of capital and default costs to the monetary impulse, conditional on the automatic interest rate and prudential rule reactions. We start by re-writing the equation for capital as follows:

$$\mathrm{d}\mathbf{K} = \Big(\underbrace{\mathbf{I} - \mathbf{F}_{\mathrm{K}}}_{\text{GE Multiplier}} \Big)^{-1} \Big(\underbrace{\mathbf{F}_{\mathrm{r}} \mathrm{d}\mathbf{r}}_{\text{Monetary Policy}} + \underbrace{\mathbf{F}_{\lambda} \mathrm{d}\boldsymbol{\lambda}}_{\text{Prudential Reaction}} \Big) \tag{34}$$

Suppose that the central bank unexpectedly hikes the nominal rate, $\mathrm{d}\mathbf{r}^{\mathbf{N}}$. The general-equilibrium response of aggregate capital to this shock consists of three components: the interest rate rule reaction, the prudential policy reaction, and the general-equilibrium (GE) effect. The relevant margin for all agents in the economy is the *real* rate. Thus, parameters of the interest rate rule—in conjunction with the NK block—determine the equilibrium real interest rate path, $\mathrm{d}\mathbf{r}$, following (31) and (32).

The sensitivity of aggregate quantities to changes in $\mathrm{d}\mathbf{r}$ is captured by $\mathbf{F}_{\mathrm{r}}$. We refer to $\mathbf{F}_{\mathrm{r}}$ as the *monetary policy Jacobian*. This important object captures the partial-equilibrium reaction of aggregate capital to the real interest rate change, everything else equal. Alternatively, this represents the partial-equilibrium impact of the monetary policy shock on capital. Economic theory would predict that $\mathbf{F}_{\mathrm{r}}$ is generally negative, at least on impact, implying a negative causal impact of higher real interest rates on bank lending and firm investment.

An important channel of monetary policy transmission in HBANK is the systematic reaction of the prudential authority. This is captured by the equilibrium path of $\mathrm{d}\boldsymbol{\lambda}$. This path depends on the prudential policy rule (22), i.e. on the strength of the cyclicality of the desired shifts in leverage regulation in response to deviations of bank default costs from their steady-state levels. The sensitivity of aggregate capital to changes in $\mathrm{d}\boldsymbol{\lambda}$ is captured by the prudential policy Jacobian F_{λ} which, similarly to $\mathbf{F}_{\mathrm{r}}$, stands for the matrix of partial-equilibrium responses. Ex ante, it is most likely that most terms in F_{λ} are negative since the impact of tighter leverage regulation has an immediate negative effect on bank assets.

The third and final channel is the GE multiplier term. This term $\mathbf{F}_K$ captures the fixed point aspect of the response of bank lending supply to shocks to aggregate capital. Due to the concavity of the aggregate production function, entries of $\mathbf{F}_K$ are most surely negative. In other words, since bank lending increases with aggregate returns, R_t^k, it falls with the level of aggregate capital. This logic is similar to the indirect effect of the response of non-financial firms to monetary shocks in Ottonello and Winberry (2020).

4.2 Financial Stability Block

In the absence of endogenous bank default risk, Equation (34) would be sufficient to completely characterize the bank lending channel. However, there is also the financial stability block of the model, that can be summarized with the following response function:

$$d\mathbf{S} = \underbrace{\mathbf{X}_K d\mathbf{K}}_{\text{Equilibrium Capital}} + \underbrace{\mathbf{X}_r d\mathbf{r}}_{\text{Monetary Policy}} + \underbrace{\mathbf{X}_\lambda d\boldsymbol{\lambda}}_{\text{Prudential Reaction}} \tag{35}$$

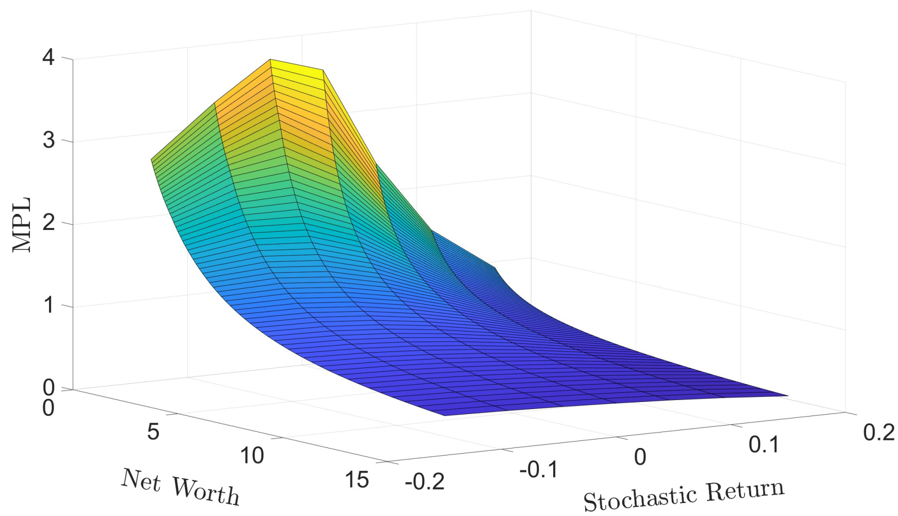
As before, the impulse is an anticipated nominal interest rate shock, and $d\mathbf{r}$, $d\boldsymbol{\lambda}$, and $d\mathbf{K}$ are GE paths for the real rate, prudential regulation, and aggregate capital. The Jacobian $\mathbf{X}_r$ represents the partial-equilibrium effect of monetary policy on financial stability. The sign of its entries is ex-ante ambiguous. A higher cost of external financing for banks—induced by the monetary contraction—should decelerate balance sheet growth, bringing more banks closer to the zero-net-worth bound. However, the risk-taking channel of monetary policy posits that low interest rates, while increasing credit supply, can also stimulate more risk taking, which can, in turn, fuel financial instability (Bruno and Shin, 2015). Quantitatively, as we show later in the paper, the balance sheet channel dominates and the entries of $\mathbf{X}_r$ are generally negative.

The indirect channel of the financial stability block is comprised of two parts. The systematic prudential policy response to the monetary impulse spills over onto $d\mathbf{S}$ via the $\mathbf{X}_\lambda$ Jacobian. To the extent that a prudential tightening curbs risk-taking behavior, we expect the entries of $\mathbf{X}_\lambda$ to be negative. Finally, the equilibrium path of capital influences default costs through the $\mathbf{X}_K$ Jacobian. Diminishing marginal returns to capital suggest that $\mathbf{X}_K$ should contain negative terms.

4.3 Bank Heterogeneity

It may appear, at first glance, that the distribution of banks is absent in equations (34) and (35). However, crucially, the dynamics of the distribution, Γ_t, along the transition

Figure 1: MPL Heterogeneity



Notes: Bank-specific marginal propensities to lend as a function of net worth and idiosyncratic returns.

path following exogenous monetary shocks, is fully operational in the background. It is taken care of and captured by the bank lending and default cost Jacobians, **F** and **X**. These Jacobians collect aggregate, economy-wide responses to shocks. Any changes in the micro-foundations of the model, such as the structure of financial markets or nominal rigidities, translate into changes in the relevant Jacobian matrices.

In partial equilibrium, aggregate sensitivity of the banking sector towards exogenous shocks is best summarized with the average marginal propensity to lend (MPL). In the absence of any cross-sectional variation, such as in the RBANK special case of our framework, the average MPL would be sufficient to understand the partial-equilibrium pass-through of shocks to bank lending decisions as it would simply correspond to the MPL of the representative agent. However, the average MPL can mask rich underlying heterogeneity across individual banks.

Figure 1 depicts bank-level MPLs, which were defined in (12), as a function of net worth size, n, and stochastic returns, ξ. MPLs vary systematically across both idiosyncratic states, as banks that are small and less profitable generally have a higher MPL. In other words, the sensitivity to shocks is concentrated in the left tail of the distributions of bank size and profitability. This is consistent with the empirical literature that finds that the elasticity of bank credit supply to monetary shocks, to the extent that those impact net worth, is higher for precisely these types of banks (Kashyap and Stein, 1995; Kishan and Opiela, 2000). We will return to the question of the heterogeneous transmission of shocks in Section 6.2.

In summary, suppose again that the central bank unexpectedly hikes the nominal rate, $\mathrm{d}\mathbf{r}^{\mathbf{N}}$. The real interest rate initially increases one-to-one and, through $\mathbf{F}_{\mathrm{r}}$, raises the cost of external capital for banks, leading to a reduction in the credit supply to firms. The extent of the decline in credit depends explicitly on the distribution of MPLs, i.e. the elasticity of lending to exogenous shocks. Less aggregate credit leads to a decline in capital, which feeds back into the bank lending problem via the GE multiplier channel, $\mathbf{F}_{\mathrm{K}}$. The aggregate marginal cost adjusts, because both the rental rate of capital and the real wage have shifted, thereby lowering the inflation rate. The interest rate rule creates an additional feedback effect via $\mathbf{F}_{\mathrm{r}}$ and, in equilibrium, the real rate responds by more than one-to-one to the original shock. Compression of bank balance sheet size and asset valuation effects change the risk profile of the financial sector, raising both the probability and the costs of default in a heterogeneous fashion across banks. The prudential authority reacts automatically by tightening prudential regulation which, via $\mathbf{F}_{\lambda}$, also feeds back into the bank lending problem. All of these forces are taken into account in general equilibrium. We now turn to the parameterization of our framework in order to inspect the transmission of monetary policy in HBANK quantitatively.

5 Parameterization

In this section, we calibrate the model and quantitatively inspect its steady-state and transitional properties.

5.1 Calibration

The model period is one quarter. We fix a subset of parameters exogenously and calibrate the remaining ones in order to target select moments in the data. Table 1 summarizes the parameterization of the model. We start with the description of the households block. The discount factor, β, is set to 0.996 in order to target a steady-state real interest rate of 1.6%. The Frisch labor supply elasticity, χ_2, is set at unity following the literature (Kaplan et al., 2018). The labor dis-utility parameter is calibrated in order to target unit labor supply.

For the banking block, we set the survival rate, σ, to 0.973 following Gertler and Kiyotaki (2010). The stochastic process for the idiosyncratic return shock, ξ, is calibrated in two steps. First, persistence ρ_ξ is set to 0.553 following Jamilov and Monacelli (2025) who estimate a linear panel fixed-effects model with AR (1) disturbances from the U.S. Reports of Condition and Income, commonly known as Call Reports. Second, volatility σ_ξ is calibrated in order to target an average quarterly probability of bank default of 2%.

Table 1: Model Parameterization

Parameter	Description	Value	Target/Source
		Households	
β	Discount factor	0.996	Internally calibrated
χ_1	Labor disutility	1.82	Labor supply = 1
χ_2	Labor supply elasticity	1	Kaplan et al. (2018)
		Banks	
σ	Bank survival rate	0.973	Gertler and Kiyotaki (2010)
ζ_1	Non-interest expense, linear	0.0024	Non-interest cost to assets ratio = 0.05
ζ_2	Non-interest expense, quadratic	2	Normalization
ρ_ξ	Idiosyncratic risk, persistence	0.553	Call Reports
σ_ξ	Idiosyncratic risk, volatility	0.04	Average default probability = 2%
ω_1	Default cost, linear	0.28	Granja et al. (2017)
ω_2	Default cost, quadratic	2	Normalization
ψ	Resource cost of default	0.0086	Default cost to output ratio = 2.5%
		Firms	
α	Capital share	0.36	Standard
a	Production technology	2.65	Steady-state capital price = 1
b	Production technology	0.25	Price elasticity of lending = 0.25
γ	Demand elasticity	10	Standard
θ	Price adjustment cost	90	Slope of the Phillips curve = 0.1
		Monetary and Prudential Policy	
φ_π	Taylor rule coefficient	1.25	Standard
$\bar{r}$	Steady-state real rate target	1.6% p.a.	Standard
ϕ	Prudential policy rule	10	Internally calibrated
λ	Steady-state leverage policy	0.02	Average bank leverage ratio = 10

We arrive at this number by computing z-scores for the cross-section of U.S. banks. The z-score is a commonly used proxy measure for default risk (Laeven and Levine, 2009). Section 7.1 provides more details on the empirical approach. The average z-score across banks and time suggests an average probability of insolvency of 2% per quarter. We discretize the idiosyncratic return process into seven grid points using the Tauchen (1986) method.

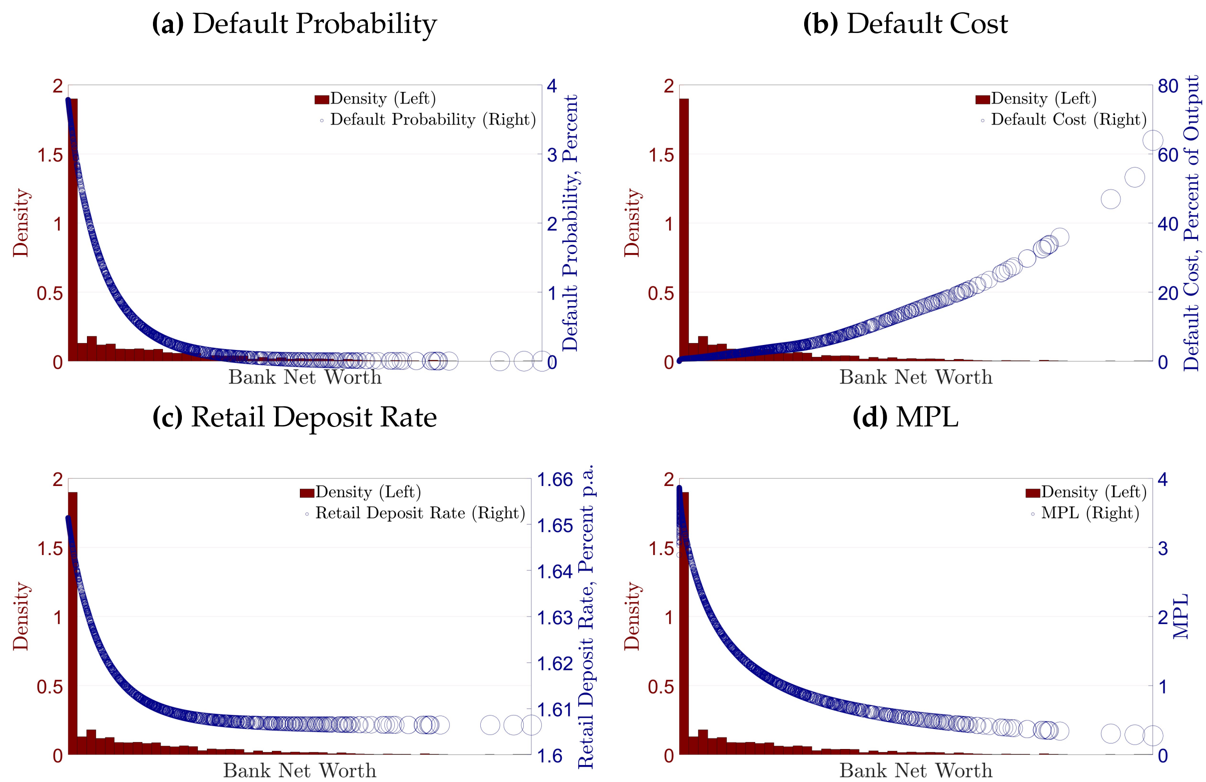
The costs of bank default are calibrated in three steps. First, we set ω_1 to 0.28, which is the average FDIC loss from bank failures in the U.S. (Granja et al., 2017). Second, the parameter ψ is calibrated to match the share of quarterly aggregate output lost due to bank failures. In the data, this value is around 2.5%. The historical macro-finance literature estimates that the long-run, or "steady-state", output losses from banking crises and panics, are around 10% per year (Laeven and Valencia, 2018; Baron et al., 2021; Jamilov et al., 2024). Third, parameter ω_2 is normalized to 2, implying a quadratic default cost function. As Section 7.1 demonstrates, this is an appropriate representation of the data. The empirical relationship between U.S. bank size and measures that capture the macro impact of bank-specific distress, such as CoVaR, is increasing and highly convex (Adrian and Brunnermeier, 2016). Finally, the non-interest expense parameter, ζ_1, calibrated to target a non-interest cost to assets ratio of 5%. In the Call Reports, it is in the region of 1.5%-6.5% depending on the precise definition of bank assets and the time period.

We continue with the firms block. The capital share, α, demand elasticity, γ, and the price adjustment cost parameter, θ, are set to standard values. The implied slope of the Phillips curve 0.1, in line with Kaplan et al. (2020), among others. This is in the ballpark albeit slightly on the higher end of the recent micro empirical estimates Hazell et al. (2022). One of our robustness tests will be to show that our main quantitative results are impervious to the slope of the Phillips curve. The capital goods production function takes on the following functional form: $\Phi(x) = ax^{1-b}, \quad a > 0, b > 0$. The parameter b is set in order to target the capital price elasticity of bank lending of 0.25. This value corresponds to the elasticity of the price of capital to firm investment that is usually estimated in the data (Gertler and Gilchrist, 1994). The technology parameter a is calibrated to target the steady-state price of capital, Q_t, of unity.

We conclude with the policy block. The Taylor rule coefficient, φ_π, is set to the standard value of 1.25. The steady-state real interest target is 1.6% p.a. The steady-state leverage regulation parameter, λ, is calibrated internally in order to target the average bank leverage ratio, defined as the ratio of assets over net worth, of 10. This target is the same across all banks in the steady state, however, micro-prudential policy targeting individual banks will be implemented along the transition paths following aggregate shocks. The prudential policy rule parameter, ϕ, is internally calibrated to 10 in order to match the empirically estimated elasticity of bank lending to changes in capital buffer requirements of 1.5 percentage points (Behn et al., 2025). Importantly, unless stated otherwise, we will be considering *micro-prudential* policy rules that target only the top 25% largest banks by assets. In other words, if the prudential authority determines—according to the policy rule (22)—that policy should be tightened, the tightening will apply only to the top 25% of largest banks. The remaining banks will face the same λ as implied by the steady state.

Finally, when considering different versions of the model—such as RBANK or HBANK without default risk—we will always re-calibrate the model appropriately in order to

Figure 2: Stationary Distributions



Notes: Cross-sectional distributions of select banking variables in the stationary steady state of the model.

target the same moments as in the baseline.

5.2 Stationary distributions

We begin the presentation of the quantitative results with the analysis of the steady-state properties of the model. The four panels of Figure 2 plot the stationary distribution of net worth, n_j, overlayed with (i) bank default probability, φ_j, (ii) the cost of bank default, s_j, (iii) the retail deposit rate, R^b_j, and (iv) the MPL_j.

As can be seen from the Figure, the distribution of bank size features pronounced right-skewness, which is in line with the data. There is a small number of very large intermediaries, and the banking industry is concentrated. Second, the probability of bank default is decreasing in net worth. This is due to the fact that smaller banks are closer to the zero-net-worth insolvency threshold and it takes a small negative return draw to bring them towards bankruptcy. Third, the *cost* of bank default is increasing in bank net worth. Recall that the realized default costs are convex in assets. This pattern captures closely the real-world systemic-risk features of any modern banking sector: the macroeconomic impact of bank failure is disproportionally larger for big institutions.

Third, the distribution of default probabilities is priced competitively into the cross-section of retail deposit rates. This can be seen from Panel (c). Large intermediaries have outgrown insolvency risk, and their deposits pay essentially the risk-free rate. Small banks, on the other hand, because they are riskier in the eyes of the household, must pay a premium to attract external debt financing. Fourth and finally, because small banks face a higher cost of funds, they are more sensitive to fluctuations in net worth. In line with this logic, in Panel (d), we observe that MPLs are almost monotonically falling with size.

Note that uninsured idiosyncratic return shocks, ξ, that induce the aforementioned equilibrium cross-section of bank size, cause banks to undertake precautionary savings in the form of net worth accumulation. This resembles the standard income fluctuations problem in the case of households-consumers. Figure A.1 in the Online Appendix illustrates this mechanism by showing the net worth policy function, $n^*(n, \xi; \Gamma)$, in the baseline HBANK model, in the version of the model without default risk, and for the RBANK special case with no bank heterogeneity. Going from RBANK to HBANK without default risk we observe a significant increase in net worth accumulation that is due to market incompleteness and idiosyncratic shocks. Furthermore, the introduction of costly insolvency is an additional source of risk, which induces banks to attempt to amass even more net worth.

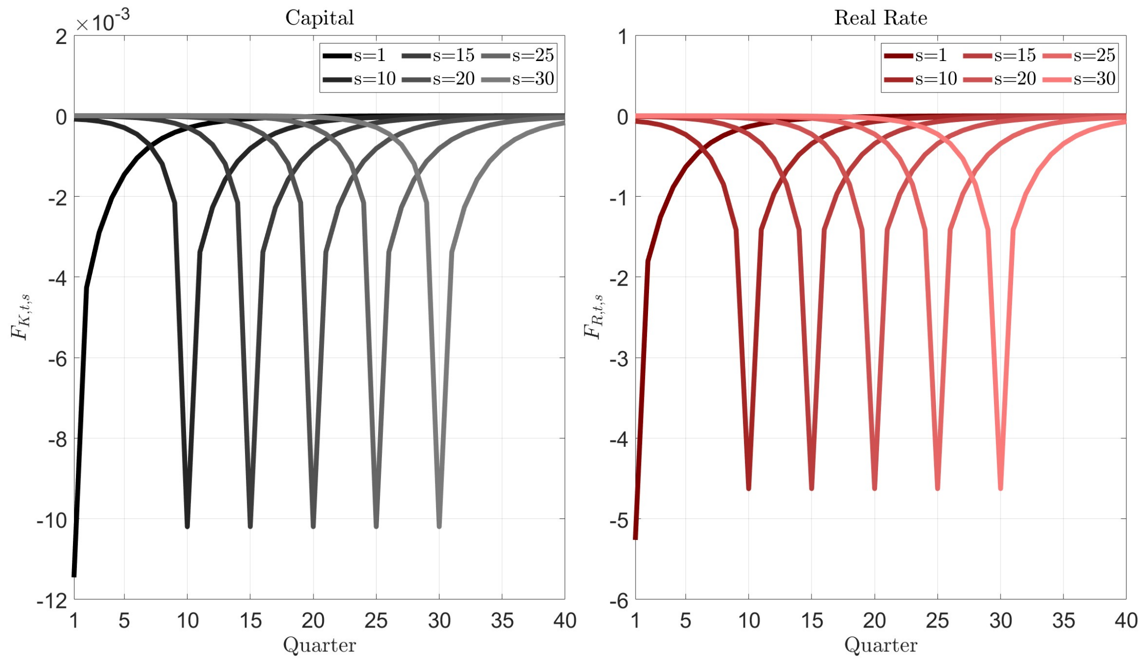
Overall, our HBANK framework delivers a very realistic picture of the banking cross-section. The distribution of size is heavily concentrated. Big banks are less likely to default, but the cost of their rare default is much larger. Small banks' insolvency risk forces them to pay higher rates on the retail deposit market. As a result, small banks have a higher MPL and are more sensitive to exogenous shocks. In Section 7, we provide direct empirical support for the model's steady-state predictions using U.S. micro-data.

5.3 Model Jacobians

Having characterized the steady-state properties of our calibrated model, we now move on to aggregate dynamics. As discussed above, the full transition path following an exogenous "MIT shock" can be computed with a small number of sufficient statistics. All of the presented Jacobians are truncated at 40 quarters.

Figure 3 shows the two key Jacobians for capital: $\mathbf{F}_K$ and $\mathbf{F}_r$. We display six response paths, respectively for shocks at horizons 1, 10, 15, 20, 25, and 30. Recall that $\mathbf{F}_r$ captures the partial-equilibrium response of aggregate bank lending (and thus, by the market clearing condition, of capital) to a real interest rate shock. As expected, entries of this matrix are negative, implying that higher real interest rates reduce bank lending to non-financial

Figure 3: Capital Jacobians, $\mathbf{F}$



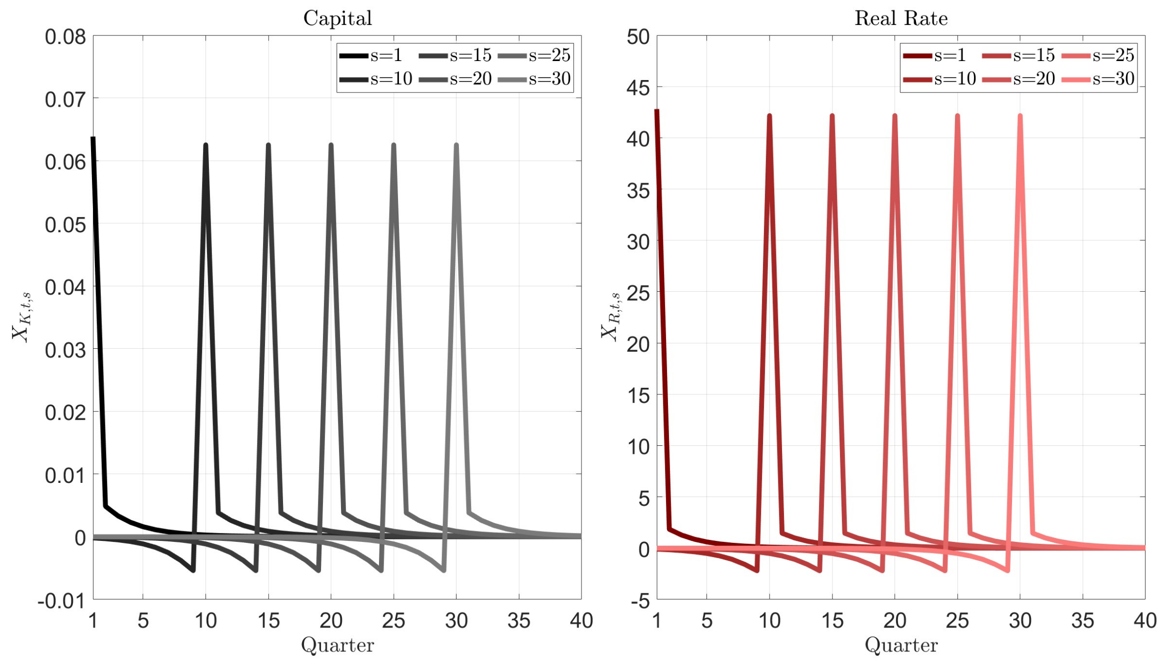
Notes: Jacobians of aggregate capital with respect to capital (left panel) and the real interest rate (right panel).

firms. This constitutes the fundamental logic of the bank lending channel of monetary policy transmission. The magnitude of the entries in $\mathbf{F}_{\mathrm{r}}$ is large, implying elasticities of the order of 4-5. In other words, a one percentage-point increase in the real interest rate reduces bank lending by around 4.5% percentage points. Notice that the columns of $\mathbf{F}_{\mathrm{r}}$ quickly converge to a stable pattern after about ten quarters. This property is shared among all the Jacobians in Figure 3 as well as in (4).

Recall that the object $\mathbf{F}_{\mathrm{K}}$ captures indirect, general-equilibrium effects. Entries of this matrix, as per Figure 3, are negative, as previously expected. A positive change in aggregate capital induces a fall in the return on capital, which in turn reduces bank credit supply. This can be seen immediately from the lending policy function (11). Since the entries of this matrix are negative, we can deduce that the general-equilibrium channel in HBANK acts as a dampener of real rate shocks. The magnitude of the entries in $\mathbf{F}_{\mathrm{K}}$ is very small, suggesting that this channel plays a quantitatively minute role.

Figure 4 now plots the two key Jacobians for aggregate bank default costs: $\mathbf{X}_{\mathrm{r}}$ and $\mathbf{X}_{\mathrm{K}}$. The policy Jacobian $\mathbf{X}_{\mathrm{r}}$ summarizes the partial-equilibrium reaction of default costs to a positive real rate shock. The positive entries of this matrix are intuitive: a higher real interest rate compresses balance sheet growth and reduces distance to default. The matrix $\mathbf{X}_{\mathrm{K}}$, finally, captures general-equilibrium spillovers onto default costs from the capital

Figure 4: Default Cost Jacobians, $\mathbf{X}$



Notes: Jacobians of bank default costs with respect to aggregate capital (left panel) and the real interest rate (right panel).

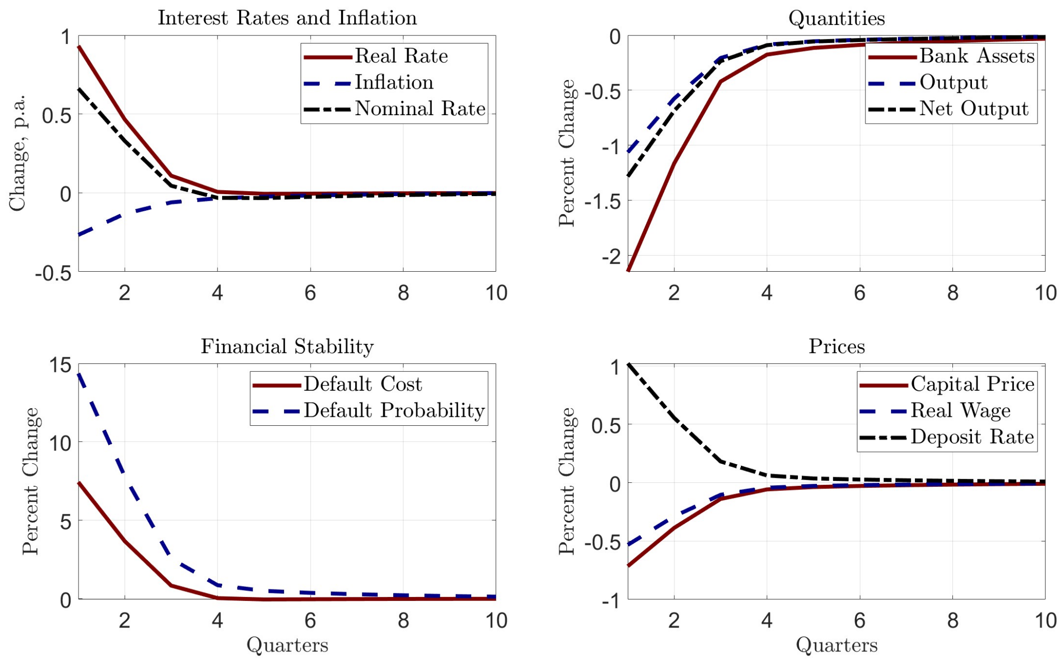
adjustment. The entries of $\mathbf{X}_{\mathrm{K}}$ are positive for the same reasons as to why the entries of $\mathbf{F}_{\mathrm{K}}$ are positive. We once more observe that the entries of $\mathbf{X}_{\mathrm{r}}$ and $\mathbf{X}_{\mathrm{K}}$ are very large and small, respectively, suggesting that the partial-equilibrium channel is potentially more important quantitatively.

In the Online Appendix, Figure A.2 presents the two key Jacobians for micro-prudential policy, $\mathbf{F}_{\lambda}$ and $\mathbf{X}_{\lambda}$. Recall that these capture, respectively, the partial-equilibrium responses to changes in λ of capital and default costs. Tighter micro-prudential regulation causes a decline in aggregate capital, as banks de-lever. Furthermore, it leads to a decline in default costs, an outcome that is consistent with de-leveraging. Thus, at least in partial equilibrium, we confirm that a prudential policy tightening reduces bank risk-taking and brings down the cost of default.

6 Quantitative Analysis of Monetary Policy

In this section, we analyze the transmission of monetary policy in HBANK quantitatively. First, we present the responses of aggregate macroeconomic and financial variables to a transitory interest rate shock. Second, we inspect the responses to a monetary shock

Figure 5: Aggregate Responses to a Monetary Policy Shock



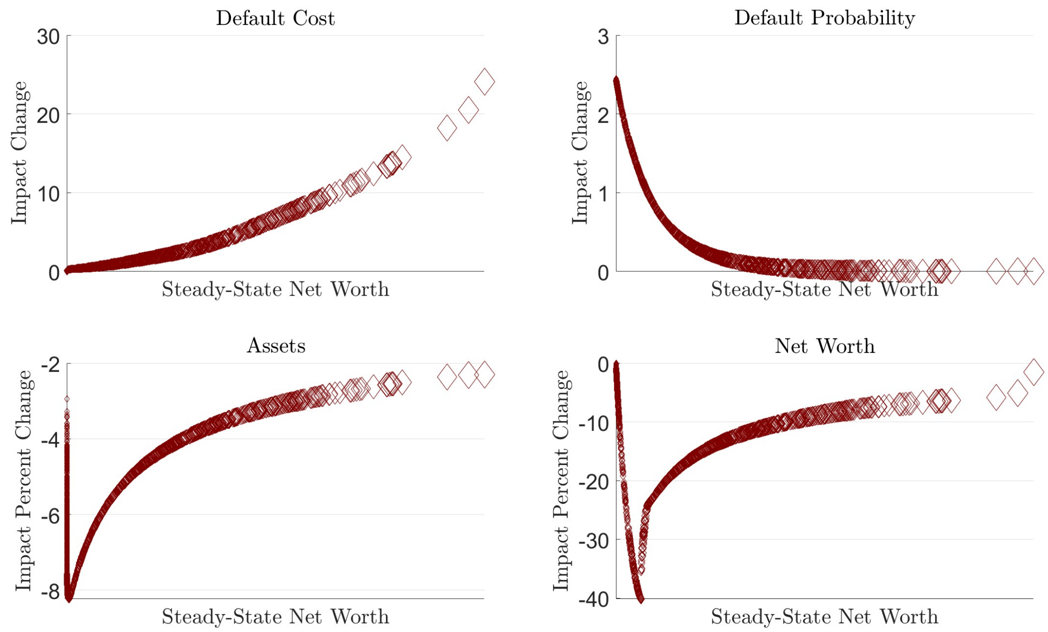
Notes: Impulse response functions to a monetary policy shock that increases the nominal interest rate by 0.25 percent on impact, with quarterly persistence of 0.5.

across the entire distribution of banks. Third, we decompose total responses into direct and indirect effects. Finally, we compare the performance of HBANK to the RBANK special case by shutting down bank heterogeneity. Throughout this section, we consider an experiment where there is a positive innovation to the Taylor rule of $v_0 = 0.25$ percent at $t = 0$, mean-reverting at the rate of 0.5. Unless stated otherwise, we will be considering the automatic micro-pru large policy which impacts only the largest 25% banks.

6.1 Aggregate Response to Monetary Policy

Figure 5 displays impulse responses of macroeconomic and financial aggregates. In response to a contractionary monetary policy shock, the real interest rate rises. It depresses aggregate demand and causes deflation, as well as lowering the real wage. The feedback effect from the interest rate rule is captured by the less than one-to-one increase of the nominal rate. The higher cost of external financing for banks leads to a decline in lending to non-financial firms. The ensuing fall in the price of capital, then, further hits banks' net-worth negatively, setting in motion a "financial acceleration" effect. As a result, aggregate output contracts. Because bank balance sheets shrink, the economy is more fragile. Both

Figure 6: Heterogeneous Responses to Monetary Policy



Notes: On-impact responses to the contractionary monetary policy shock across the distribution of banks.

the probability of bank defaults and their realized resource costs are now substantially higher. Consequently, the decline in *net* output, i.e. after the subtraction of default costs, is more severe. Finally, the retail deposit rate increases almost to the same degree as the real rate. Notice that the *pass-through* from the policy rate to the deposit rate is not exact, due to the adjustment of the default probability, that is priced into the retail rates.

Overall, the magnitudes of these responses are consistent with empirical evidence from standard macro VARs. Note that underneath the aggregate responses lies the distribution of bank-level behavior. In addition, the prudential regulator responds endogenously in this experiment by tightening the limit on the leverage multiple of the largest banks. We now turn to these two key dimensions of monetary policy transmission.

6.2 Heterogeneous Responses to Monetary Policy

Figure 6 presents the heterogeneous responses to the contractionary monetary policy shock. We show the full distribution of responses—that is, for every individual bank in our economy. For tractability, the figure plots changes on impact, i.e., in period $t = 0$. It is evident that aggregate responses mask the rich heterogeneity present at the individual bank level. The transmission of monetary policy varies systematically with bank size:

small banks experience a substantially larger decline in both assets and net worth. This pattern is consistent with the well-established empirical regularity that small banks are more responsive to monetary policy shocks (Kashyap and Stein, 1995, 2000). The behavior of retail deposits is quantitatively very similar to that of assets (not shown).

The financial stability picture is particularly interesting. The probability of default of small banks increases by much more than for the average intermediary. Interestingly, the cost of default behaves quite differently: it is the large banks whose realized default costs rise by more. Even a slight increase in the likelihood of failure of a large, systemically important institution is sufficient to drive up resource costs significantly. In other words, big banks in HBANK are essentially "too big to fail".

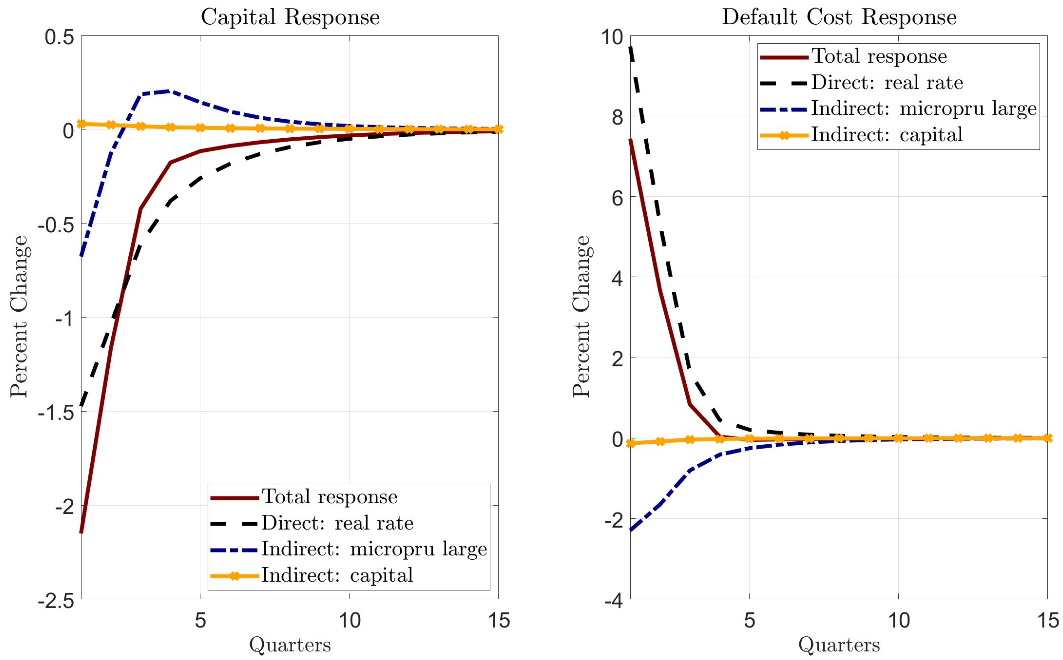
Overall, the above discussion shows that inspecting the bank lending channel of monetary transmission through the lenses of a representative-intermediary model could be very limiting. The bank lending channel operates heterogeneously across the banking distribution. Moreover, different macroeconomic and financial variables propagate across the distribution in markedly different ways. Monetary policy shocks induce reallocation of asset and deposit holdings across the individual banks with strong implications for financial stability. Fully understanding how monetary policy impacts the financial sector's quantities, prices, and fragility metrics requires a framework with bank-level balance sheet and risk heterogeneity.

6.3 Direct and Indirect Effects Decomposition

As discussed previously in Section 4, the total macroeconomic response to monetary policy is a combination of various partial and general equilibrium channels. We now provide a quantitative decomposition of the aggregate responses to a contractionary monetary policy shock into direct and indirect effects. Figure 7 presents the results of this exercise. We focus on the total responses of aggregate capital, $d\mathbf{K}$, and default costs, $d\mathbf{S}$, which are sufficient to recover every other endogenous object in the model. Each total response is decomposed into the direct effect that is due to the interest rate hike, and two indirect effects that are due to the endogenous prudential policy response and the general-equilibrium adjustment in capital. Recall that we are still considering a systematic *micro*-prudential policy rule which impacts the limit on leverage of only the largest 25% banks.

We begin with the response of capital. Recall that equation (34) summarizes the general-equilibrium response of capital to the monetary policy impulse. The left panel of Figure 7 plots the three components of (34) in an additively-separate manner. The direct effect of the policy shock, captured by $\mathbf{F}_r d\mathbf{r}$, is large and negative. This was already

Figure 7: Decomposing the Monetary Transmission Mechanism



Notes: Decomposition of the total response to a contractionary monetary policy shock into direct and indirect effects.

evident from the earlier analysis of the $\mathbf{F}_r$ Jacobian. The indirect response from the prudential authority, captured by $\mathbf{F}_\lambda d\lambda$ is negative and considerably large. As the aggregate impulse responses have previously demonstrated, both the probability and costs of bank defaults rise following the interest rate hike. The prudential regulator, therefore, reacts by increasing the limit on the leverage multiple for the largest banks. Tighter prudential regulation, in turn, further reduces aggregate bank lending and production. Thus, the indirect effect from the prudential policy response by itself *amplifies* the monetary policy shock, especially on impact. Finally, the indirect effect that is due to the general-equilibrium adjustment in capital, $\mathbf{F}_K$, is negative. This channel, as expected, is quantitatively very small.

Moving on to the default cost response, the right panel of Figure 7 decomposes $d\mathbf{S}$ into direct and indirect effects. Recall that its GE path is summarized by equation (35). The direct effect from the real interest rate hike is large and positive. Intuitively, higher borrowing costs for banks compress balance sheet growth and reduce the distance to default. The prudential policy response is negative, which is consistent with the automatic micro-prudential tightening in reaction to heightened default costs and rising financial fragility. Finally, the indirect effect that is due to the GE capital adjustment is negative and immaterially small.

In summary, so far we have derived two main results on the transmission of monetary policy shocks in HBANK. First, the total macroeconomic response in HBANK is driven largely by the *direct* effect of the monetary policy shock. Specifically, the total response of bank lending—which is sufficient to pin down capital, labor supply, and output—is 75% the direct effect, 24% the indirect effect from micro-prudential policy that targets large banks, and 1% the indirect effect from capital adjustment. This finding echoes the conclusions in Ottonello and Winberry (2020) on the importance of *both* direct and indirect effects for the investment channel of monetary policy with heterogeneous *non*-financial firms. The intuition for this result is the much-higher interest rate sensitivity of banks. In the models of Auclert (2019) and Kaplan et al. (2018), the indirect effect is generally more potent because households, unlike banks, are less price-sensitive due to consumption smoothing motives.

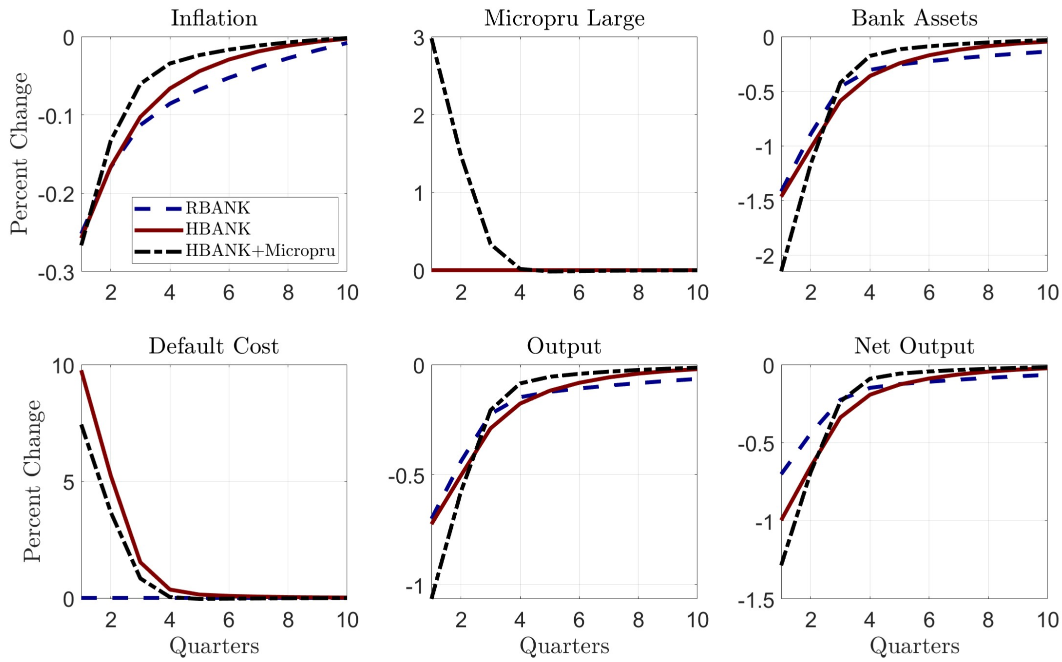
Second, we have shown that the endogenous micro-prudential policy response, while successfully reducing financial fragility and resource costs of bank default, amplifies the economic contraction that is caused by the interest rate hike. This observation points to systematic prudential policy being a potentially important channel of monetary transmission in HANK-type models like ours. It also points to possible *trade-offs* between macroeconomic and financial stabilization functions of the policy-maker. We return to this question in much greater detail in Section 6.5 when we discuss systematic monetary policy.

6.4 Monetary Policy Transmission in HBANK and RBANK

What are the aggregate implications of bank heterogeneity and insolvency risk for the transmission of monetary policy? To answer this question, we now compare the quantitative performance of our baseline calibrated HBANK framework with two special cases. First, to replicate the RBANK benchmark, we set the volatility of idiosyncratic shocks, σ_ξ, to zero. The distribution of banks now collapses to a single, representative agent. Second, to achieve the special case of HBANK without insolvency risk, we set φ_t(n, xi) to zero for all banks in the distribution. This also implies that there is no cost of default, at the bank level or on the aggregate.

We begin by comparing the macroeconomic response to a monetary contraction in HBANK and RBANK. Figure 8 reports the results. Quantitatively, the responses of bank assets, output, and net output are at least 30% greater in the baseline HBANK model with endogenous micro-pru policy than in RBANK. This magnitude is comparable to the literature (see, for example, Kekre and Lenel (2022) in the context of risk-taking house-

Figure 8: Monetary Transmission in HBANK and RBANK

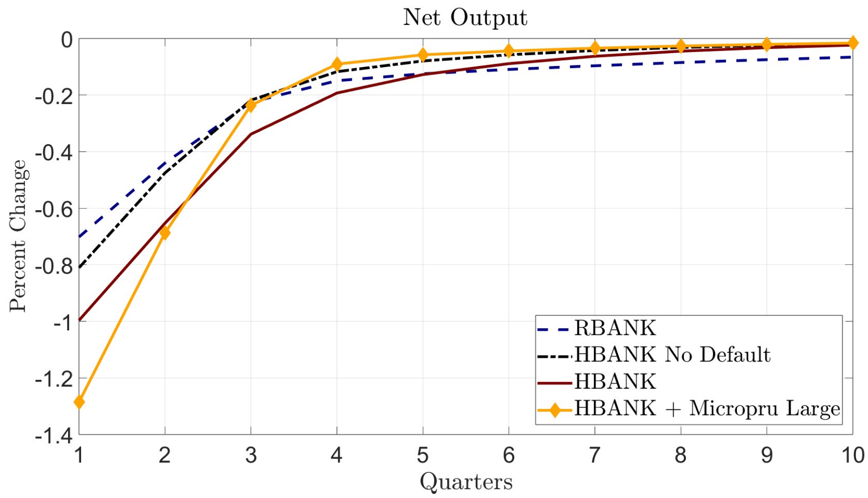


Notes: Responses to a contractionary monetary policy shock in HBANK and RBANK.

holds). Along the transition path back to the steady state, the recovery is somewhat faster in HBANK. This is due to the micro-foundations of the banking problem and specifically the precautionary savings motive that incentivizes banks to accumulate net worth faster. However, Figure A.3 in the Online Appendix shows that the *cumulative* macroeconomic response is still substantially larger in HBANK. Importantly, the prudential authority reacts to the monetary contraction by tightening leverage regulation on the largest banks. This policy reaction induces a second-round negative effect on credit supply and output, reinforcing the original shock.

We also find that *bank heterogeneity by itself*, i.e. without endogenous prudential policy, magnifies the bank lending channel of monetary policy. The amplification is particularly strong on impact and especially for net output. The mechanism behind this result lies within the financial stability block of the model. To isolate this channel, Figure 9 shows how the macroeconomic response changes in the special case of the model without default risk. In RBANK, the financial stability channel is absent entirely and banks do not face an idiosyncratic income fluctuation problem. Thus, the macroeconomic response is the weakest on impact. Introducing bank heterogeneity amplifies the economic response due to the addition of a large mass of small banks with high MPL. Furthermore, adding

Figure 9: Monetary Transmission in HBANK with and without Default Risk



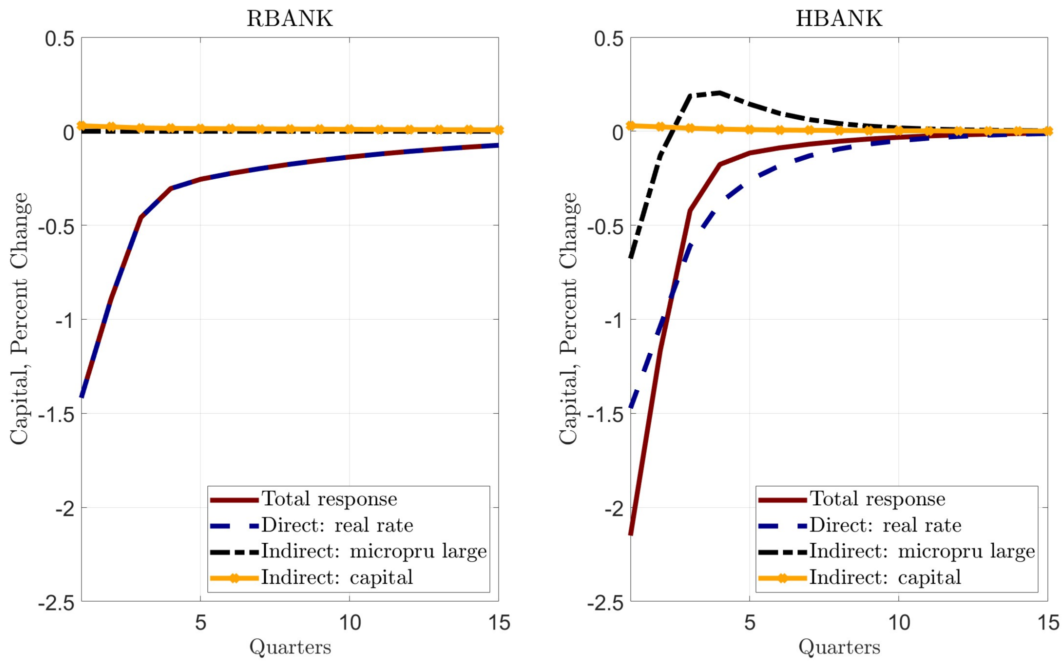
Notes: Response of net output to a contractionary monetary policy shock in RBANK as well as HBANK with and without endogenous default and endogenous micro-prudential policy targeting large banks.

endogenous and costly default further magnifies the response. Finally, as noted earlier, allowing for an endogenous micro-prudential policy reaction adds further amplification to the response, especially on impact.

To further shed light on the amplification result, we also decompose the monetary transmission mechanism into direct and indirect effects for RBANK. Figure 10 shows that the total response of capital in RBANK is driven almost entirely by the direct effect of the real interest rate hike. As mentioned before, the financial stability block is inactive and there is no endogenous prudential policy reaction. The indirect response from the capital adjustment, instead, is very small. In contrast, the transmission of monetary policy in HBANK is more complex and features a powerful, endogenous micro-prudential policy response to heightened financial fragility. Incidentally, even the direct effect by itself is greater in HBANK than in RBANK. These are the precise sources of the total amplification result.

To conclude, in this section we have quantitatively revisited the bank lending channel of non-systematic monetary policy. We have shown that an unexpected monetary policy contraction leads to a decline in financial aggregates, an economic recession, and rising financial instability as captured by the probability and costs of bank default. Beyond the aggregates, we have also demonstrated empirically-consistent heterogeneity in the responses to monetary policy across the distribution of individual banks. A critical factor in shaping the total macroeconomic effects of monetary shocks is the *endogenous micro-prudential* policy reaction function. In HBANK, the prudential authority responds to rising

Figure 10: Decomposing Monetary Transmission in HBANK and RBANK



Notes: Decomposition of the total response to a contractionary monetary shock into direct and indirect effects, in HBANK and RBANK.

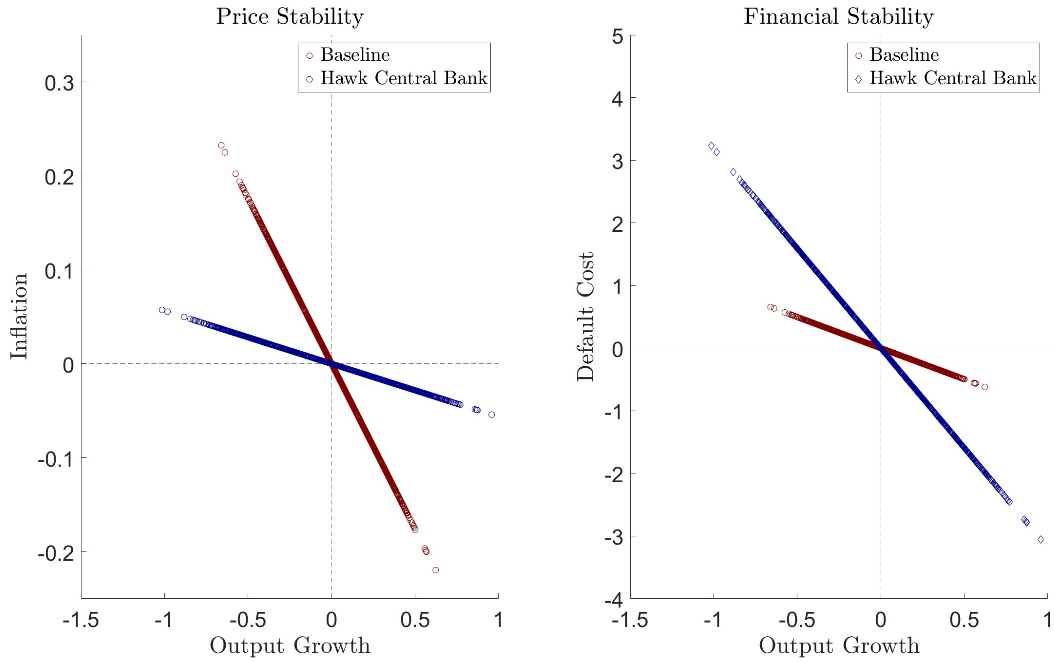
financial fragilities by tightening the constraint on leverage of the largest, systematically-important banks. This amplifies the original monetary policy shock by 1.3 times.

6.5 Macroeconomic-Financial Stabilization Trade-off

In this section, we employ our HBANK framework to study the systematic conduct of monetary policy. First, we show that the presence of costly bank default risk poses a trade-off between stabilizing the macroeconomy and financial stability for the central bank. Second, we demonstrate how automatic micro-prudential regulation that targets the largest banks can effectively alleviate the trade-off.

Taking advantage of sequence-space methods, we start by simulating the baseline model for 10,000 periods and discarding the first 2,000. The only aggregate shock in the simulation is a TFP shock, A_t, with volatility 0.01 and persistence 0.9. Observe that aggregate productivity only impacts the bank lending problem via the return on aggregate capital. The direct effect of the lending response requires a new Jacobian, $\mathbf{F}_A$. Its entries are: $[F_A]_{t,s} = \frac{\partial \mathcal{L}_t}{\partial r^k_{s+1}} \frac{\partial r^k_{s+1}}{\partial A_s}$. Similarly, the default cost response has a new Jacobian $\mathbf{X}_A$, whose

Figure 11: Macroeconomic-Financial Stabilization Trade-Off



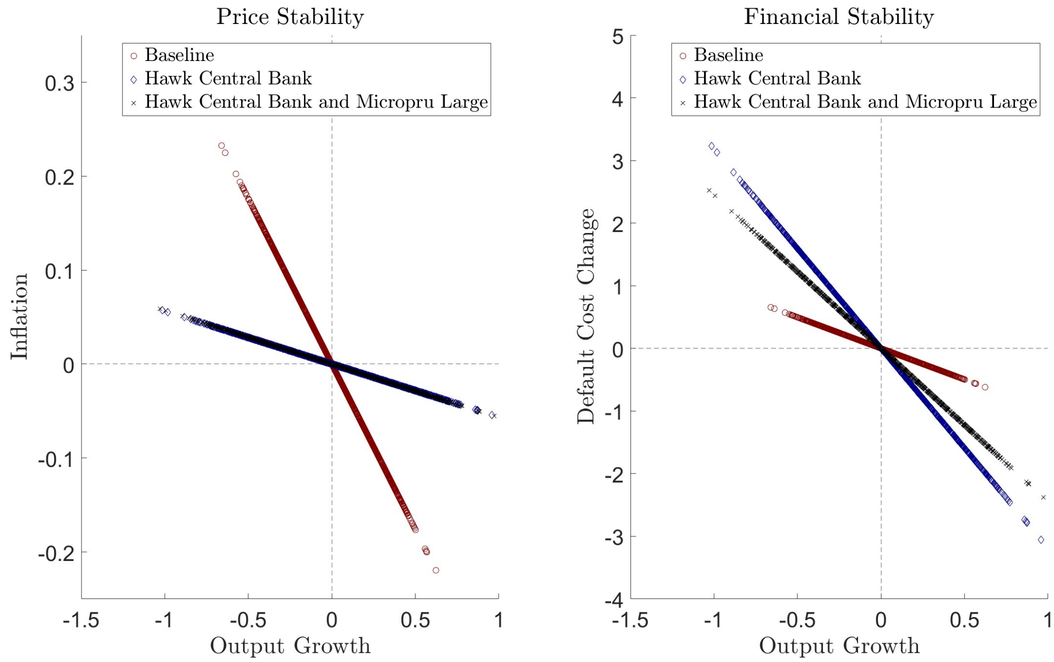
Notes: Output growth, inflation, and default cost realizations from a long stochastic simulation of the model driven by aggregate TFP shocks.

entries are: $[\mathbf{X}]_{t,s} = \frac{\partial \mathcal{S}_t}{\partial r^k_{s+1}} \frac{\partial r^k_{s+1}}{\partial A_s}$. The new derivatives can be computed analytically around the steady state.

Figure 11 presents the outcome of the stochastic simulation. Figure A.4 in the Appendix presents the underlying impulse response functions. The panels show output growth, inflation, and default cost realizations. Each dot on the plots represents a particular quarter. The baseline economy corresponds to the case with $\varphi_\pi = 1.25$. In the case of the "hawk central bank", the monetary reaction to inflation is more aggressive: $\varphi_\pi = 6$.

The left panel of Figure 11 shows that a more hawkish central bank can successfully enhance price stability over the long run. The output-inflation curve flattens considerably, implying that for any given realization of output growth, prices deviate much less from the steady state. However, as the right panel demonstrates, the unintended consequence of the central bank's more hawkish stance on inflation is increased financial instability. The output-default cost curve *steepens*, suggesting that fluctuations in bank default costs over the business cycle become more volatile. In particular, during negative output realizations—i.e., supply-driven recessions that coincide with higher inflation—default costs are consistently and substantially elevated.

Figure 12: The Role of Systematic Micro-prudential Policy for Big Banks



Notes: Output growth, inflation, and default cost realizations from a long stochastic simulation of the model with systematic micro-prudential policy that targets big banks.

This result highlights a *trade-off* for the central bank: targeting price stability comes at the cost of greater financial fragility. A classic example of the Tinbergen principle: the central bank has only one tool to combat inflation—the nominal interest rate—yet financial stability represents a second target, which requires a separate instrument. This is where systematic prudential policy comes into play.

6.6 Systematic Micro-Prudential Policy Targeting Big Banks

We now consider *systematic micro-prudential* policy that targets the largest quartile of banks. The prudential policy rule is still parameterized by $\phi = 10$. Specification of the stochastic simulation with aggregate TFP shocks is the same as before. Figure 12 presents the results from this experiment.

The Figure now introduces a third case. From the right panel, we observe how the output-default cost curve *flattens* relative to the case of the hawkish central bank. In other words, systematic micro-prudential policy mitigates the macroeconomic-financial stabilization trade-off. Specifically, the prudential authority tightens leverage regulation in supply-driven recessions, when financial fragility is high and relaxes regulation in booms, when financial fragility is low. Fluctuations of bank default costs over the business cycle are now considerably smaller. In this regard, the functioning of systematic micro-prudential policy is reminiscent of an automatic stabilizer (McKay and Reis, 2016).

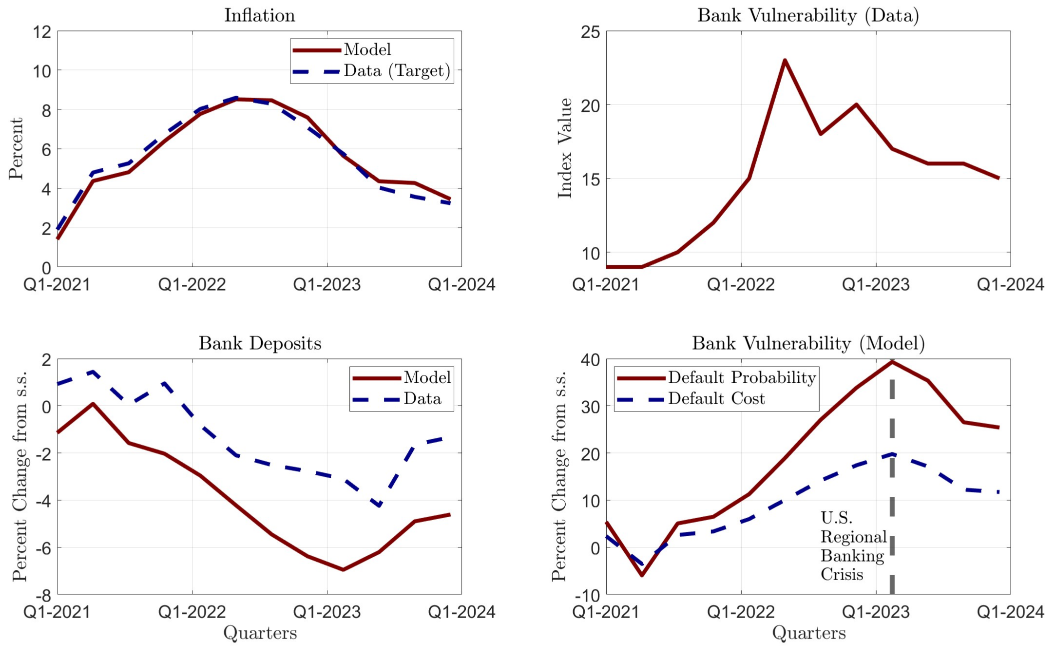
Robustness. For completeness, Figure A.5 in the Appendix plots the results with systematic prudential policy targeting the bottom 75% of banks. The trade-off between macroeconomic and financial stabilization is practically unaffected. This signifies the systemic importance of the top quartile of the banking distribution. In practice, it may be substantially easier to regulate a small number of large banks in a concentrated banking system (Beck et al., 2006). Thus, automatic micro-prudential policy that targets large banks may be not only more effective but far less costly than the alternatives.

Recent empirical evidence points to the flattening of the Phillips Curve (PC) (Hazell et al., 2022). How does this trend impact the macroeconomic-financial stabilization trade-off and the role of micro-pru policy? Figure A.6 in the Online Appendix shows the results from the simulation with the slope of the PC set to 0.006, instead of 0.1. We observe that the trade-off, if anything, is starker for the flatter PC, as seen from the right panel. In addition, the potency of the endogenous micro-pru large policy is still quantitatively large.

Finally, Figure A.7 in the Online Appendix considers fluctuations driven by a demand shock with the same volatility 0.01 and persistence 0.9 as before. There is now a natural positive (negative) co-movement between output growth and inflation (bank default costs). A hawkish central bank manages to flatten the output-inflation curve but steepens the output-financial stability curve. In other words, the trade-off between macroeconomic and financial stabilization is present also in the case of demand shocks, albeit it is quantitatively less pronounced. Systematic micro-pru policy targeting large banks again flattens both of the curves, implying that it successfully mitigates the trade-off.

To conclude, in this section we have shown that the central bank faces a trade-off between price stability and financial stability. A similar trade-off arises in the model of systemic risk by Coimbra and Rey (2023), where financial intermediaries are constrained by value-at-risk. Our novel contribution is to emphasize the role of systematic prudential policy as an instrument for the automatic stabilization of business cycles and mitigation of such trade-offs. Since the banking industry is heavily concentrated, micro-prudential regulation that targets the right tail of the distribution is particularly effective. It is the large banks whose marginal impact on the real economy and financial fragility is strongest. Section 6 has shown that automatic micro-prudential policy is a source of amplification for non-systematic monetary shocks. In this section, we have instead shown that it can

Figure 13: The 2021-2023 U.S. Experience According to HBANK



Notes: The path of select U.S. macroeconomic and financial aggregates over 2021-2024 in the data and in the model.

stabilize the financial sector without interfering with the central bank's systematic conduct of monetary policy and the price stability mandate.

6.7 The 2021-2023 U.S. Experience According to HBANK

In this section, we apply our calibrated HBANK framework to the U.S. experience from 2021 to 2023. In 2022, the Federal Reserve implemented significant monetary tightening to combat an unprecedented surge in inflation. Many have argued that rising interest rates contributed to delayed financial instability—most notably, the 2023 regional banking crisis and the run on Silicon Valley Bank (Drechsler et al., 2023; Jiang et al., 2024). However, the crisis did not escalate into a systemic event affecting the entire financial sector; instead, it was limited to a few individual banks. Our quantitative laboratory is therefore perfectly-suited to study this natural experiment, as it captures the real effects of monetary policy, endogenous financial fragility, and bank-level heterogeneity.

Bernanke and Blanchard (2025) showed that most of the inflationary pressure originated in the goods market, driven by sharp increases in relative prices—such as commodity prices and prices in sectors where strong demand met supply constraints. Therefore, modeling this as a "cost-push" shock is a reasonable approach. We proceed by reverseengineering the path of a cost-push shock that would match U.S. CPI inflation over 2021-2024 in the data and our baseline model. In addition, to allow for an equilibrium delay between peak inflation and financial instability in the model, we also allow for a backward-looking component in the Taylor rule.

Figure 13 presents the results of this experiment. The top-left panel displays the matched path of inflation, while the remaining panels show model-implied trajectories and empirical series of select macroeconomic and financial aggregates. The cost-push show causes persistent inflation, peaking at around 8% in mid-2022. In response to the large shock, the central bank raises the interest rate. Lower aggregate demand pushes the price of capital down, and asset valuation effects negatively impact bank balance sheets, leading to a decline in lending. As a result, aggregate credit supply, capital, and output all fall.

The bottom-left panel shows the path of bank deposits in the model and the data. We de-trend the empirical series since quarterly deposit trend growth is 2% while there is no trend growth in the model. Both series display qualitatively similar patterns: gradual declines and lows around the first quarter of 2023.

The top-right panel of the Figure presents an empirical proxy of financial instability around this period. Our favorite measure is the fire sale index from Duarte and Eisenbach (2021), which captures general banking-sector vulnerability.[14] As the central bank was raising the interest rate, the index was gradually rising until it peaked around mid-2022 and remained considerably high until 2024.

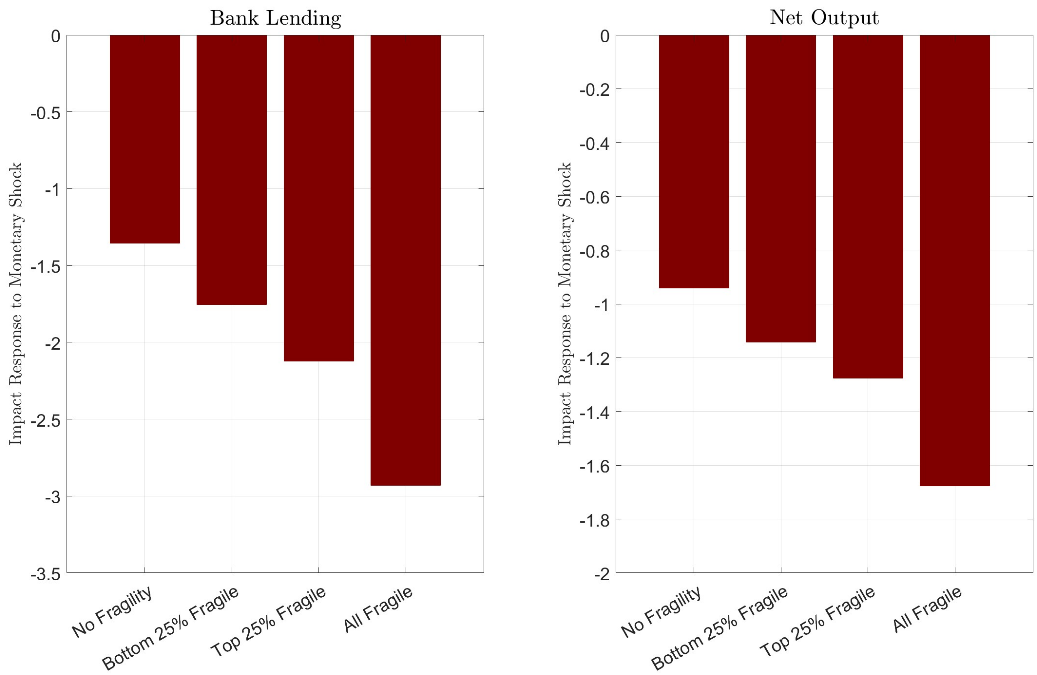
The bottom-right panel of the Figure shows the two proxies of financial instability in HBANK. Both the average probability and costs of bank default increase gradually, peaking at approximately the same time as the 2023 U.S. banking crisis. Because the interest rate rule now includes a backward-looking component, the central bank maintains a higher nominal rate for longer than usual. Consequently, the peak in financial fragility occurs about one year after the inflation peak in 2022, consistent with the observed timeline. The overall pattern of the paths of banking fragility in the model and the data are very similar. Importantly, the banking sector does not collapse entirely, as it would in a representative-bank counterfactual.[15] Instead, a fraction of banks default while others remain solvent, offering a more realistic depiction of what actually occurred.

In summary, using our calibrated HBANK framework, we replicate the U.S. experience from 2021 to 2023 by modeling a cost-push inflation shock and a monetary policy

[14] Similar pictures are obtained if we use alternative measures from Duarte and Eisenbach (2021), such as the capital vulnerability or the run vulnerability indices.

[15] In other words, in a representative-bank model, either the whole banking sector is insolvent or it is not; either there is a run on the entire sector, or there is none at all. There is no margin for bank-specific failures.

Figure 14: Monetary Transmission and Distributional State-Dependency



Notes: On-impact response to a contractionary monetary policy shock conditional on different underlying distribution of banks.

tightening that followed. The model captures key macro-financial dynamics, including a delayed rise in bank defaults following interest rate hikes. It reproduces the observed sequence of inflation peaking in 2022 and financial instability emerging in 2023, while allowing for partial, bank-specific failures rather than systemic collapse, thereby aligning closely with the real-world outcome.

6.8 Distributional State-Dependency of Monetary Policy

In our HBANK framework, the transmission of monetary policy shocks may exhibit *state dependence*, in the sense that it is shaped by the underlying degree of *financial fragility* within the banking sector. In our final quantitative exercise, we investigate whether monetary policy becomes more or less effective depending on the distributional configuration of bank balance sheets—specifically, whether the economy is in a state of heightened vulnerability, with a larger share of fragile institutions. This analysis allows us to assess how the macroeconomic impact of monetary policy is mediated by the resilience (or lack thereof) of the financial intermediary sector.

In Figure 14, we show on-impact responses of bank lending and net output to a contractionary monetary policy shock conditional on different underlying distributions of banks. The first column in both panels shows the baseline response when the distribution initiates from the steady state. In the second column, we exogenously increase the probability of default, $\varphi_{j,t}$, by 2.5% (10% per annum) only for the smallest 25% of banks as defined by steady-state net worth. In the third column, we perturb only the largest 25% of banks. In the fourth column, the probability of default for all banks is ex-ante higher by 2.5%. One can think of these perturbations as being due to an exogenous financial shock that occurs before the interest rate hike.

As can be clearly seen from the Figure, monetary policy is more powerful when the distribution of banks is more fragile than in the steady state. Intuitively, a financial shock that increases the probability of default for all or some banks reduces distance to default and reallocates bank net worth towards banks with high marginal propensity to lend. As a result, the sensitivity of the economy (bank ending and output) towards interest rate shocks is higher if the underlying distribution is ex-ante more fragile. In particular, this effect is more pronounced if it is the large banks that are more fragile, rather than the small ones. This observation is consistent with our earlier discussion of the systemically important role of big banks for the interplay between financial stability and monetary policy.

7 Empirical Evidence

To empirically assess the theoretical predictions of our model, in this final section we turn to micro-data on U.S. commercial banks. We present *four key findings* that provide empirical validation for our framework. First, we document a negative relationship between bank size and default risk, consistent with the notion that larger banks are more resilient. Second, we find that the aggregate cost of bank defaults is convex in bank size, implying that the failure of larger institutions carries disproportionately higher systemic consequences. Third, we show that small banks' balance sheets respond more strongly to monetary policy shocks than those of large banks, reflecting a higher MPL. Finally, we document that default risk rises in response to contractionary monetary policy shocks, with the effect being particularly pronounced for smaller institutions. We describe our data sources and construction in more detail and run a battery of robustness checks in Online Appendix B.

7.1 Default Risk and Default Cost in the Cross-Section of Banks

We begin by documenting how default risk and default costs vary in the cross-section of U.S. banks. Following Laeven and Levine (2009), we use the *z-score* as a proxy of default risk. This measure is particularly convenient in our setting because it can be constructed from income-statement and balance-sheet items alone, making it readily available for all banks and time periods in our sample.[16] By contrast, market-based measures of default risk are only available for publicly traded banks, which constitute a small fraction of the universe of all U.S. commercial banks.[17] Under the assumption of normally distributed profits, the z-score equals the inverse of the probability of insolvency of a given bank.

Panel (a) of Figure 15 presents a binned scatter plot of bank size —defined as total book assets— against the default probability as proxied by the inverse z-score. We find a robust negative cross-sectional relationship between bank size and default risk: a 10% increase in bank size is associated with a decrease in the probability of default of 2 basis points, approximately 1% of the sample mean. This empirical pattern confirms a central prediction of our model: default risk systematically declines with bank size, as illustrated in Figure 2.

Next, we examine the relationship between bank size and the systemic cost of bank default, another core feature of our theory. To proxy for the systemic cost of default of an individual institution, we use the ΔCoVaR measure developed in Adrian and Brunnermeier (2016). This metric is computed as the change in the overall financial sector's value-at-risk conditional on a bank moving from its median state to a distressed state. Thus, it measures the maximum potential systemic loss triggered by an individual bank's distress.

Panel (b) of Figure 15 plots our proxy for the aggregate default cost against bank size and provides strong evidence of a convex relationship. This convexity indicates that failures of large banks disproportionately affect the overall economy, validating our model calibration discussed in Section 5. In Figure B.2 of Appendix B, we show that this convex relationship persists even when considering probability-weighted default costs.

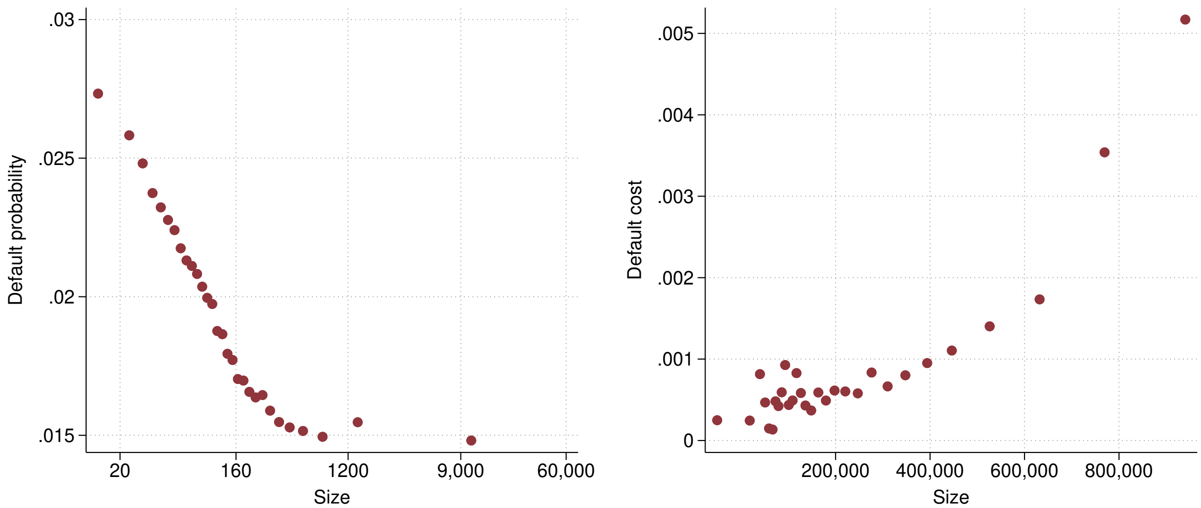
7.2 Monetary Policy Transmission in the Cross-Section of Banks

Having validated the stationary properties of our HBANK framework, we now test its dynamic predictions. Our empirical analysis proceeds in two stages. First, we study

[16] Online Appendix B provides details on measurement and computation.

[17] The z-score is highly correlated with market-based measures of default risk such as Nagel and Purnanandam (2019), albeit the sample size shrinks considerably due to the fact that many banks are not listed.

Figure 15: Size, Default Risk, and Default Cost in the Cross-Section of Banks



(a) Default probability

(b) Default cost

Notes: binned scatter plots of default probability (panel (a)) and default cost (panel (b)) against bank size. We proxy default probability with the inverse z-score (Laeven and Levine, 2009), default cost with the 95% dollar CoVaR from Adrian and Brunnermeier (2016), and bank size with total assets. Both axes are residualized from time fixed effects.

the aggregate response of assets and default risk to a monetary policy shock for the U.S. commercial banking sector.[18] Second, we investigate the heterogeneity of these responses across individual banks.

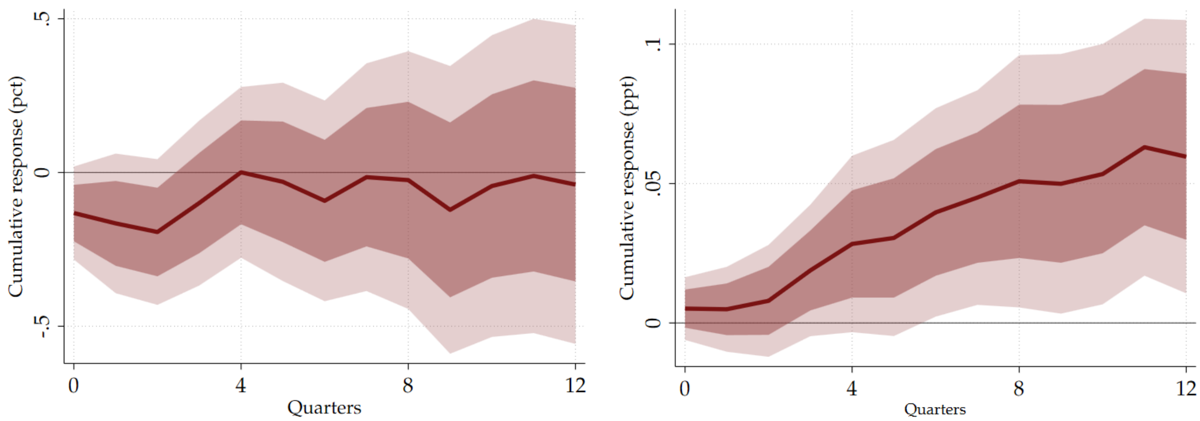
Aggregate responses. We start by studying the response of total real assets and default risk of the average bank to a contractionary monetary policy shock. To do so, we rely on the following lag-augmented panel local projection (Jordà, 2005; Montiel Olea and Plagborg-Møller, 2021):

$$\Delta Y_{it+h} \;=\; \alpha_{ih} + \psi_h \varepsilon_t + \sum_{\ell=1}^{4} \gamma_{h\ell} \Delta Y_{it-\ell} + \sum_{\ell=1}^{4} \phi_{h\ell} X_{t-\ell} + u_{iht} \tag{36}$$

where ε_t is the "poor man's" monetary policy shock from Jarociński and Karadi (2020) and $X_{t-\ell}$ is a vector of controls which includes the CPI, real GDP, the return on the S&P 500 index, the 1-year treasury rate, and the excess bond premium (Gilchrist and Zakrajšek, 2012). We define ΔY_{it+h} as the cumulative level change of Y_{it} between quarters $t-1$ and

[18]Unfortunately, we can not estimate the conditional response of default costs to a monetary shock because time variation in the CoVaR measure of Adrian and Brunnermeier (2016) is obtained via projection methods and thus only captures variation in underlying aggregate macroeconomic variables.

Figure 16: Aggregate Response of the Banking Sector to Monetary Policy Shocks



(a) Assets

(b) Default Risk

Notes: estimated ψ_h from (36) to a one-standard-deviation contractionary monetary shock. The y-axis represents the cumulative percentage change in total real assets in panel (a) and the cumulative level change in default probability —as proxied by the inverse z-score— in panel (b). The x-axis represents quarters elapsed since the shock. Errors are two-way clustered at the time and bank level. Lightly (darkly) shaded areas represent 90% (68%) confidence intervals.

$t + h$ for the probability of default and the cumulative log-change for assets.[19] We follow Almuzara and Sancibrián (2024) and use two-way clustered standard errors by bank and time.

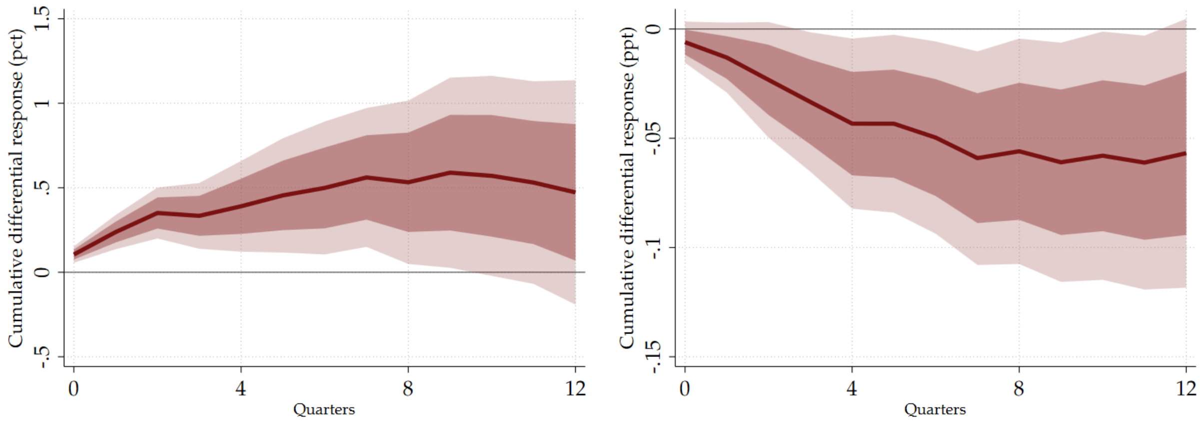
Figure 16 plots the estimated $\widehat{\psi_h}$. Panel (a) shows the response of total book real assets. Following a one-standard deviation contractionary monetary policy shock, aggregate assets fall by around .15% .[20] This is in line with our model's predictions in Figure 5 and with the large literature on the bank lending channel of monetary policy. The wide standard errors around these point estimates are a well-documented pattern in the literature and reflect increased drawdowns of existing credit facilities (Greenwald et al., 2025) as well as the fact that we only observe book, rather than market, asset holdings. Moreover, we will show momentarily how this aggregate response masks substantial cross-sectional differences, which motivates the focus on bank heterogeneity in our theoretical framework.

Panel (b) plots the response of default risk. A monetary policy tightening makes the aggregate banking sector more fragile. In particular, the default probability of the average bank increases by around 50 basis points after a one-standard deviation contractionary monetary shock. This identified moment is directly in line with the model's predictions,

[19] Similarly, $\Delta Y_{it-\ell}$ denotes cumulative level changes between quarters $t-\ell-1$ and $t-1$ for the probability of default and cumulative past log-changes for assets.

[20] To put this number into perspective, a one-standard-deviation monetary contraction corresponds to roughly a 18 basis points hike in the policy rate.

Figure 17: Heterogeneous Response of the Banking Sector to Monetary Shocks



(a) Assets

(b) Default Risk

Notes: estimated β_h from (37) to a one-standard-deviation contractionary monetary shock. The y-axis represents the cumulative percentage change in total real assets in panel (a) and the cumulative level change in default probability —as proxied by the inverse z-score— in panel (b) for banks in the top 10% of the asset distribution, relative to those in the bottom 90%. The x-axis represents quarters elapsed since the shock. Errors are two-way clustered at the time and bank level. Lightly (darkly) shaded areas represent 90% (68%) confidence intervals.

in particular the impulse responses in Figure 5. It also places structure on the Jacobian $\mathbf{X}_r$.

Figures B.3 and B.4 in Appendix B show that the above results are robust to using total net book loans as the outcome variable instead of total book assets, as well as to different variants of the monetary shock measure.

Heterogeneous responses. A key prediction of our theory is that the aggregate behavior of the banking sector only provides a partial characterization of the transmission of monetary policy. Accordingly, we now examine the heterogeneous effects of monetary policy in the *cross-section* of U.S. banks. To do so, we run the following regression:

$$\Delta Y_{it+h} = \underbrace{\alpha_{ih} + \delta_{th}}_{\text{Fixed effects}} + \underbrace{\beta_h \times D_{it} \times \varepsilon_t}_{\text{Size interaction}} + \underbrace{\phi_h D_{it}}_{\text{Interaction controls}} + \underbrace{\sum_{\ell=1}^{4} \gamma_{h\ell} \Delta Y_{it-\ell}}_{\text{Lagged controls}} + u_{iht} \tag{37}$$

where δ_{th} is a time fixed effect and D_{it} is a dummy variable which is equal to unity only for those banks that were in the top 10% of the size distribution in the quarter preceding the shock. $\Delta Y_{i,t+h}$, $\Delta Y_{i,t-\ell}$, α_{ih}, and ε_t are defined as above.

Figure 17 plots the estimated $\hat{\beta}_h$ coefficients. While the time fixed effect δ_{th} absorbs the average response to the monetary shock, our focus here is on the differential response across banks. This is precisely what the coefficient β_h captures. Specifically, β_h can be interpreted as the differential response of the dependent variable for banks in the top 10% of the asset distribution *relative* to those in the bottom 90%.

Panel (a) of Figure 17 shows that, compared to small banks, large banks tend to contract their balance sheets by less following a monetary tightening—consistent with the findings of Kashyap and Stein (1995, 2000). After a one-standard-deviation contractionary monetary policy shock, banks in the top decile of the size distribution experience a decline in book assets that is up to 0.5% smaller than that of banks in the bottom 90%. This estimate is statistically significant for at least eight quarters. This differential lending response to monetary policy is in line with the model's predictions, particularly the responses shown in Figure 6. It also corresponds to Figure 1 and the steady-state MPL heterogeneity in HBANK. In our model, large banks are less sensitive to net-worth fluctuations and—since monetary policy affects bank net worth—are therefore less responsive to monetary shocks. The results in Figure 17 validate this key testable prediction.

Panel (b) of Figure 17 plots the estimated $\hat{\beta}_h$ for default risk. Following a monetary contraction, the probability of default increases by up to 6 basis points less for banks in the top decile of the size distribution compared to those in the bottom 90%. This finding also supports our model's prediction, as shown in Figure 6. Altogether, these results underscore the importance of accounting for bank heterogeneity—particularly in size and default risk—when assessing the transmission of monetary policy, which is the central message of our HBANK framework.

Figures B.3, B.5 and B.6 in Appendix B show that the above findings also hold for total book loans and are robust to different monetary shock measures as well as to different definitions of the dummy D_{it}.

8 Conclusions

There is a trade-off between macroeconomic and financial stabilization faced by the central bank. This trade-off arises in our newly developed HBANK framework—a dynamic general equilibrium model with heterogeneous banks, costly endogenous bank insolvency, and nominal rigidities. A manifestation of the Tinbergen principle, this trade-off cannot be readily managed through interest rate policy alone. Instead, automatic prudential regulation serves as a credible secondary instrument that addresses financial fragility without compromising the central bank's price stability mandate. In particular, endogenous micro-prudential policy targeting only the top quartile of banks proves especially effective—more so than policies targeting the bottom three quartiles and nearly as effective as full macro-prudential regulation.

Our paper revisits the canonical bank lending channel of monetary transmission by moving beyond the representative-bank benchmark. In a setting characterized by rich heterogeneity in bank balance sheets and exposures, we characterize the full general-equilibrium response to monetary policy shocks using sufficient statistics and tractable sequence-space methods. This approach allows us to quantify how the distribution of financial fragility across banks shapes the aggregate effects of monetary policy.

The flexibility of our HBANK framework opens several promising avenues for future research. One natural extension is to embed the model in an open-economy context, allowing for cross-border bank lending and international policy spillovers. Another is to incorporate banking networks, where interconnectedness may amplify financial fragility through balance sheet linkages. Both extensions would deepen our understanding of how monetary and prudential policy interact in complex financial environments, especially in the presence of systemic risk.

References

Abad, Jorge, David Martinez-Miera, and Rafael Repullo, "A Macroeconomic Model of Banks' Systemic Risk Taking," *CEPR DP19532*, 2024.

Abadi, Joseph, Markus Brunnermeier, and Yann Koby, "The Reversal Interest Rate," *American Economic Review*, August 2023, *113* (8), 2084–2120.

Adrian, Tobias and Markus K Brunnermeier, "CoVaR," *American Economic Review*, July 2016, *106* (7), 1705–41.

Aggarwal, Rishabh, Adrien Auclert, Matthew Rognlie, and Ludwig Straub, "Excess Savings and Twin Deficits: The Transmission of Fiscal Stimulus in Open Economies," *NBER Macroeconomics Annual*, 2023, *37*, 325–412.

Almuzara, Martin and Victor Sancibrián, "Micro Responses to Macro Shocks," *Federal Reserve Bank of New York Staff Reports*, 2024, *190.*

Auclert, Adrien, "Monetary Policy and the Redistribution Channel," *American Economic Review*, June 2019, *109* (6), 2333–67.

__ , "HANK: A New Core of Usable Macroeconomics," *American Economic Association Papers and Proceedings*, 2025, *Forthcoming.*

__ , Bence Bardóczy, Matthew Rognlie, and Ludwig Straub, "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models," *Econometrica*, 2021, *89* (5), 2375–2408.

__ , Matthew Rognlie, and Ludwig Straub, "Micro Jumps, Macro Humps: Monetary Policy and Business Cycles in an Estimated HANK Model," *NBER Working Paper*, 2020, *26647.*

__ , __ , and __ , "Fiscal and Monetary Policy with Heterogeneous Agents," *Annual Review of Economics*, 2024, *Forthcoming.*

__ , __ , and __ , "The Intertemporal Keynesian Cross," *Journal of Political Economy*, 2024, *132* (12), 4068–4121.

__ , __ , Martin Souchier, and Ludwig Straub, "Exchange Rates and Monetary Policy with Heterogeneous Agents: Sizing up the Real Income Channel," *NBER Working Paper 28873*, 2021.

__ , Rodolfo Rigato, Matthew Rognlie, and Ludwig Straub, "New Pricing Models, Same Old Phillips Curves?," *The Quarterly Journal of Economics*, 2023, *139* (1), 121–186.

Baqaee, David R., Emmanuel Farhi, and Kunal Sangani, "The Supply-Side Effects of Monetary Policy," *Journal of Political Economy*, 2024, *132* (4), 1065–1112.

Baron, Matthew, Emil Verner, and Wei Xiong, "Banking Crises Without Panics," *Quarterly Journal of Economics*, 2021, *136* (1), 51–113.

Beck, Thorsten, Asli Demirgüç-Kunt, and Ross Levine, "Bank concentration, competition, and crises: First results," *Journal of Banking Finance*, 2006, *30* (5), 1581–1603.

Begenau, Juliane and Tim Landvoigt, "Financial Regulation in a Quantitative Model of the Modern Banking System," *The Review of Economic Studies*, 12 2021, *89* (4), 1748–1784.

__ , Saki Bigio, Jeremy Majerovitz, and Matias Vieyra, "A Q-Theory of Banks," *Review of Economic Studies*, 2025, *Forthcoming.*

Behn, Markus, Stijn Claessens, Leonardo Gambacorta, and Alessio Reghezza, "Macroprudential and monetary policy tightening: more than a double whammy?," *ECB Working Paper 3043*, 2025.

Benhabib, Jess, Alberto Bisin, and Mi Luo, "Wealth Distribution and Social Mobility in the US: A Quantitative Approach," *American Economic Review*, May 2019, *109* (5), 1623–47.

— **and** — , "Skewed Wealth Distributions: Theory and Empirics," *Journal of Economic Literature*, December 2018, *56* (4), 1261–91.

Bernanke, Ben and Alan Blinder, "Credit, Money, and Aggregate Demand," *The American Economic Review*, 1988, *78* (2), 435–439.

— **and** — , "The Federal Funds Rate and the Channels of Monetary Transmission," *The American Economic Review*, 1992, *82* (4), 901–921.

— **and Mark Gertler**, "Agency Costs, Net Worth, and Business Fluctuations," *American Economic Review*, 1989, *79(1)*, 14–31.

— **and** — , "Inside the Black Box: The Credit Channel of Monetary Policy Transmission," *Journal of Economic Perspectives*, December 1995, *9* (4), 27–48.

— **and Olivier Blanchard**, "What Caused the U.S. Pandemic-Era Inflation?," *American Economic Journal: Macroeconomics*, 2025, *Forthcoming.*

— **, Mark Gertler, and Simon Gilchrist**, "The financial accelerator in a quantitative business cycle framework," *Handbook of Macroeconomics*, 1999, *1.*

Bianchi, Javier and Saki Bigio, "Banks, Liquidity Management, and Monetary Policy," *Econometrica*, 2022, *90* (1), 391–454.

Bigio, Saki and Yuliy Sannikov, "A Model of Credit, Money, Interest, and Prices," *NBER Working Paper 28540*, 2021.

Bocola, Luigi, "The Pass-Through of Sovereign Risk," *Journal of Political Economy*, 2016, *124* (4), 879–926.

Boissay, Frédéric, Fabrice Collard, and Frank Smets, "Booms and Banking Crises," *Journal of Political Economy*, 2016, *124* (2), 489–538.

Boppart, Timo, Per Krusell, and Kurt Mitman, "Exploiting MIT shocks in heterogeneous-agent economies: the impulse response as a numerical derivative," *Journal of Economic Dynamics and Control*, 2018, *89*, 68–92.

Brunnermeier, Markus and Lasse H. Pedersen, "Market Liquidity and Funding Liquidity," *Refiew of Financial Studies*, 2009, *22*, 2201–2238.

Brunnermeier, Markus K. and Yuliy Sannikov, "A Macroeconomic Model with a Financial Sector," *American Economic Review*, February 2014, *104* (2), 379–421.

Bruno, Valentina and Hyun Song Shin, "Capital flows and the risk-taking channel of monetary policy," *Journal of Monetary Economics*, 2015, *71*, 119–132.

Coimbra, Nuno and Hélène Rey, "Financial Cycles with Heterogeneous Intermediaries," *The Review of Economic Studies*, 04 2023, *91* (2), 817–857.

Corbae, Dean and Pablo D'Erasmo, "Capital Buffers in a Quantitative Model of Banking Industry Dynamics," *Econometrica*, 2021, *89* (6), 2975–3023.

Davila, Eduardo and Andreas Schaab, "Optimal Monetary Policy with Heterogeneous Agents: Discretion, Commitment, and Timeless Policy," *NBER Working Paper*, 2023, *30961.*

Dempsey, Kyle, "Capital Requirements with Non-Bank Finance," *Review of Economic Studies*, 2024, *Forthcoming.*

Di Tella, Sebastian, "Optimal Regulation of Financial Intermediaries," *American Economic Review*, January 2019, *109* (1), 271–313.

Drechsler, Itamar, Alexi Savov, Philipp Schnabl, and Olivier Wang, "Deposit Franchise Runs," Working Paper 31138, National Bureau of Economic Research April 2023.

Duarte, Fernando and Thomas M. Eisenbach, "Fire-Sale Spillovers and Systemic Risk," *The Journal of Finance*, 2021, *76* (3), 1251–1294.

Egan, Mark, Ali Hortaçsu, and Gregor Matvos, "Deposit Competition and Financial Fragility: Evidence from the US Banking Sector," *American Economic Review*, January 2017, *107* (1), 169–216.

Elenev, Vadim, Tim Landvoigt, and Stijn Van Nieuwerburgh, "A Macroeconomic Model With Financially Constrained Producers and Intermediaries," *Econometrica*, 2021, *89* (3), 1361–1418.

Faccini, Renato, Seungcheol Lee, Ralph Luetticke, Morten Ravn, and Tobias Renkin, "Financial Frictions: Macro vs Micro Volatility," *CEPR DP*, 2024, *15133.*

Galí, Jordi, "Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications," *Princeton University Press*, 2008.

Gerali, Andrea, Stefano Neri, Luca Sessa, and Federico Signoretti, "Credit and Banking in a DSGE Model of the Euro Area," *Journal of Money, Credit, and Banking*, 2010, *42(s1).*

Gertler, Mark and Nobuhiro Kiyotaki, "Financial Intermediation and Credit Policy in Business Cycle Analysis," *Handbook of Monetary Economics*, 2010, *3*, 547–599.

_ and Peter Karadi, "A Model of Unconventional Monetary Policy," *Journal of Monetary Economics*, 2011, *58(1)*, 17–34.

_ and Simon Gilchrist, "Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms," *The Quarterly Journal of Economics*, 05 1994, *109* (2), 309–340.

_ , Nobuhiro Kiyotaki, and Andrea Prestipino, "Wholesale Banking and Bank Runs in Macroeconomic Modelling of Financial Crises," *Handbook of Macroeconomics*, 2016, *2.*

_ , _ , and _ , "A Macroeconomic Model with Financial Panics," *The Review of Economic Studies*, 05 2019, *87* (1), 240–288.

Gilchrist, Simon and Egon Zakrajšek, "Credit Spreads and Business Cycle Fluctuations," *American Economic Review*, June 2012, *102* (4), 1692–1720.

Goldstein, Itay, Alexandr Kopytov, Lin Shen, and Haotian Xiang, "Bank heterogeneity and financial stability," *Journal of Financial Economics*, 2024, *162*, 103934.

González, Beatriz, Galo Nuño, Dominik Thaler, and Silvia Albrizio, "Firm Heterogeneity, Capital Misallocation and Optimal Monetary Policy," *CEPR DP 18695*, 2023.

Granja, JoÃo, Gregor Matvos, and Amit Seru, "Selling Failed Banks," *The Journal of Finance*, 2017, *72* (4), 1723–1784.

Greenwald, Daniel, John Krainer, and Pascal Paul, "The Credit Line Channel," *The Journal of Finance*, 2025, *Forthcoming.*

Greenwood, Jeremy, Zvi Hercowitz, and Gregory W. Huffman, "Investment, Capacity Utilization, and the Real Business Cycle," *The American Economic Review*, 1988, *78* (3), 402–417.

Hazell, Jonathon, Juan Herreño, Emi Nakamura, and Jón Steinsson, "The Slope of the Phillips Curve: Evidence from U.S. States," *The Quarterly Journal of Economics*, 02 2022, *137* (3), 1299–1344.

He, Zhiguo and Arvind Krishnamurthy, "Intermediary Asset Pricing," *American Economic Review*, April 2013, *103* (2), 732–70.

Jamilov, Rustam and Tommaso Monacelli, "Bewley Banks," *Review of Economic Studies*, 2025, *Forthcoming.*

— **, Tobias König, Karsten Müller, and Farzad Saidi**, "Two Centuries of Systemic Bank Runs," *CEPR Discussion Paper*, 2024, *19382.*

Jarociński, Marek and Peter Karadi, "Deconstructing Monetary Policy Surprises - The Role of Information Shocks," *American Economic Journal: Macroeconomics*, April 2020, *12* (2), 1–43.

Jeenas, Priit, "Firm Balance Sheet Liquidity, Monetary Policy Shocks, and Investment Dynamics," *Working Paper*, 2024.

Jermann, Urban and Vincenzo Quadrini, "Macroeconomic Effects of Financial Shocks," *American Economic Review*, February 2012, *102* (1), 238–71.

Jiang, Erica Xuewei, Gregor Matvos, Tomasz Piskorski, and Amit Seru, "Monetary tightening and U.S. bank fragility in 2023: Mark-to-market losses and uninsured depositor runs?," *Journal of Financial Economics*, 2024, *159*, 103899.

Jordà, Oscar, "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, March 2005, *95* (1), 161–182.

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante, "Monetary Policy According to HANK," *American Economic Review*, March 2018, *108* (3), 697–743.

— **, Kurt Mitman, and Giovanni L. Violante**, "The Housing Boom and Bust: Model Meets Evidence," *Journal of Political Economy*, 2020, *128* (9), 3285–3345.

Kashyap, Anil K. and Jeremy C. Stein, "The impact of monetary policy on bank balance sheets," *Carnegie-Rochester Conference Series on Public Policy*, 1995, *42*, 151–195.

— **and** — , "What Do a Million Observations on Banks Say about the Transmission of Monetary Policy?," *American Economic Review*, June 2000, *90* (3), 407–428.

Kekre, Rohan and Moritz Lenel, "Monetary Policy, Redistribution, and Risk Premia," *Econometrica*, 2022, *90* (5), 2249–2282.

Kishan, Ruby P. and Timothy P. Opiela, "Bank Size, Bank Capital, and the Bank Lending Channel," *Journal of Money, Credit and Banking*, 2000, *32* (1), 121–141.

Kiyotaki, Nobuhiro and John Moore, "Credit Cycles," *Journal of Political Economy*, 1997, *105* (2), 211–248.

Laeven, Luc and Fabian Valencia, "Systemic Banking Crises Revisited," Working Paper WP/18/206, International Monetary Fund Sep 2018.

— **and Ross Levine**, "Bank governance, regulation and risk taking," *Journal of Financial Economics*, 2009, *93* (2), 259–275.

Luetticke, Ralph, "Transmission of Monetary Policy with Heterogeneity in Household Portfolios," *American Economic Journal: Macroeconomics*, 2021, *13* (2), 1–25.

Mankiw, N. Gregory and Ricardo Reis, "Sticky Information in General Equilibrium," *Journal of the European Economic Association*, 2007, *5* (2/3), 603–613.

McKay, Alisdair and Ricardo Reis, "The Role of Automatic Stabilizers in the U.S. Business Cycle," *Econometrica*, 2016, *84* (1), 141–194.

— **, Emi Nakamura, and Jón Steinsson**, "The Power of Forward Guidance Revisited," *American Economic Review*, 2016, *106* (10), 3133–3158.

Mendicino, Caterina, Kalin Nikolov, Juan Rubio-Ramirez, Javier Suarez, and Dominik Supera, "Twin Default Crises and Bank Capital Requirements," *The Journal of Finance*, 2024, *Forthcoming.*

Nagel, Stefan and Amiyatosh Purnanandam, "Banks' Risk Dynamics and Distance to

Default," *The Review of Financial Studies*, 10 2019, *33* (6), 2421–2467.
Nuño, Galo and Carlos Thomas, "Bank Leverage Cycles," *American Economic Journal: Macroeconomics*, April 2017, *9* (2), 32–72.
Olea, José Luis Montiel and Mikkel Plagborg-Møller, "Local Projection Inference Is Simpler and More Robust Than You Think," *Econometrica*, 2021, *89* (4), 1789–1823.
Ottonello, Pablo and Thomas Winberry, "Financial Heterogeneity and the Investment Channel of Monetary Policy," *Econometrica*, 2020, *88* (6), 2473–2502.
Ríos-Rull, José-Víctor., Tamon Takamura, and Yaz Terajima, "Banking Dynamics, Market Discipline and Capital Regulations," *Manuscript*, 2020.
Schaab, Andreas and Stacy Y. Tan, "Monetary and Fiscal Policy According to HANK-IO," *Working Paper*, 2023.
Stein, Jeremy C., "An Adverse-Selection Model of Bank Asset and Liability Management with Implications for the Transmission of Monetary Policy," *The RAND Journal of Economics*, 1998, *29* (3), 466–486.
Tauchen, George, "Finite state markov-chain approximations to univariate and vector autoregressions," *Economics Letters*, 1986, *20* (2), 177–181.
Taylor, John B., "Discretion versus policy rules in practice," *Carnegie-Rochester Conference Series on Public Policy*, 1993, *39*, 195–214.
Werning, Iván, "Incomplete Markets and Aggregate Demand," *NBER Working Paper*, 2015, *21448.*
Wolf, Christian K., "Interest Rate Cuts versus Stimulus Payments: An Equivalence Result," *Journal of Political Economy*, 2025, *133* (4).
Woodford, Michael, "Interest and Prices: Foundations of a Theory of Monetary Policy," *Princeton University Press*, 2003.

Online Appendix for

"HBANK: Monetary Policy with Heterogeneous Banks"

Marco Bellifemine Rustam Jamilov Tommaso Monacelli

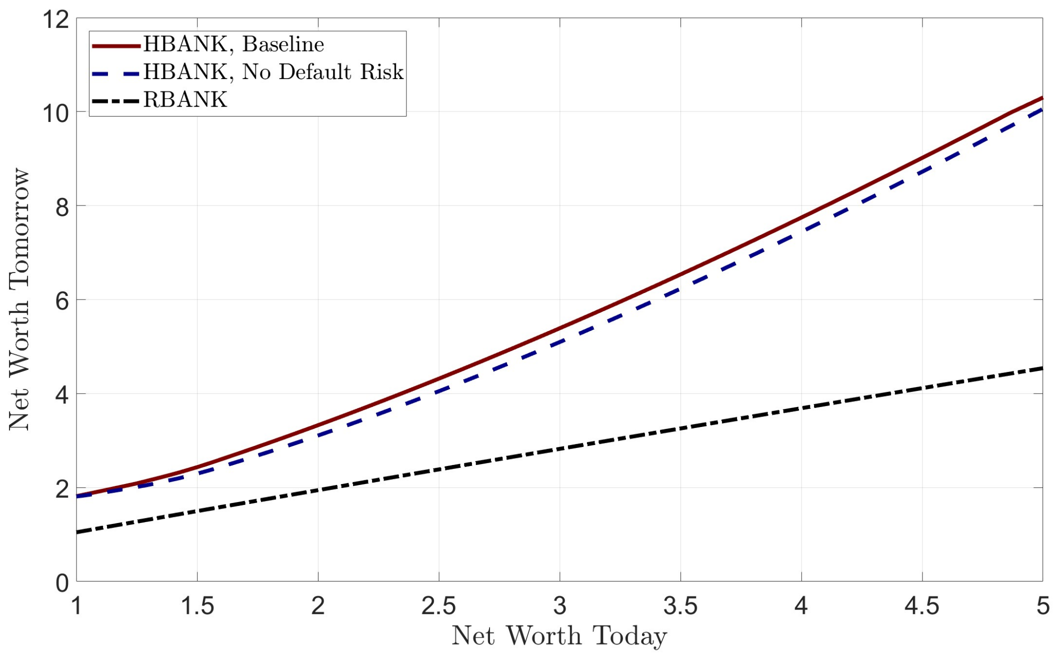
May 2025

Contents

A Model Appendix

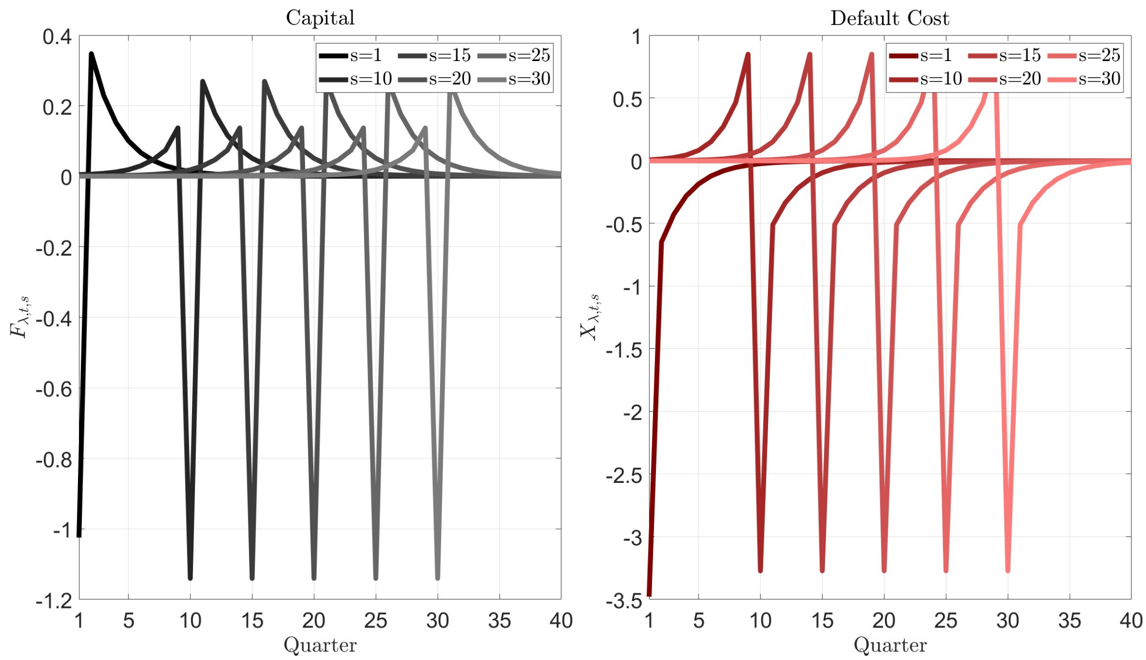
A.1 Additional Quantitative Results

Figure A.1: Illustration of the Precautionary Saving Motive



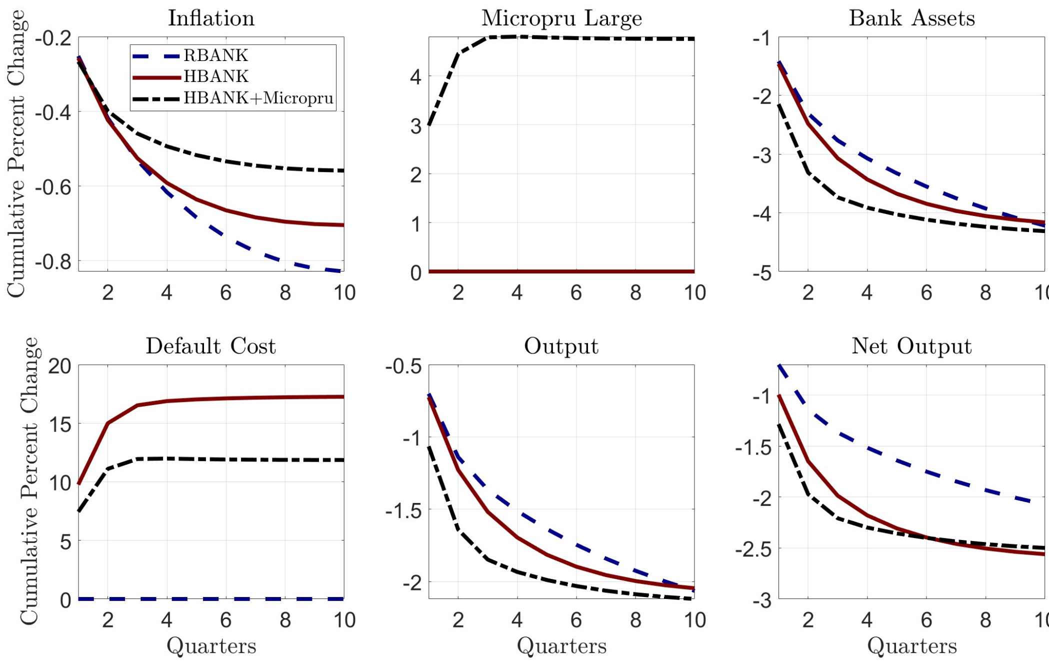
Notes: Net worth policy functions $n_t^*(n, \xi)$ in the baseline HBANK model, in the version of the model without default risk, and in the RBANK special case with no bank heterogeneity.

Figure A.2: Micro-prudential Policy Jacobians, $\mathbf{F}_\lambda$ and $\mathbf{X}_\lambda$



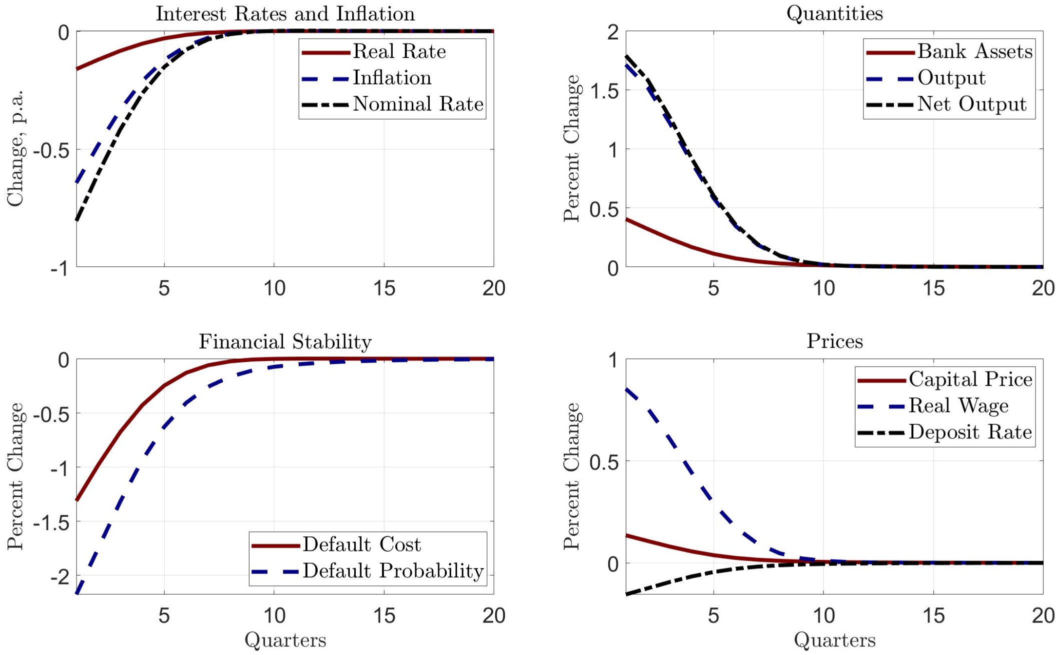
Notes: Jacobians of aggregate capital (left) and default costs (right) with respect to changes in microprudential policy that affects only large banks.

Figure A.3: Cumulative Responses to Monetary Policy in HBANK and RBANK



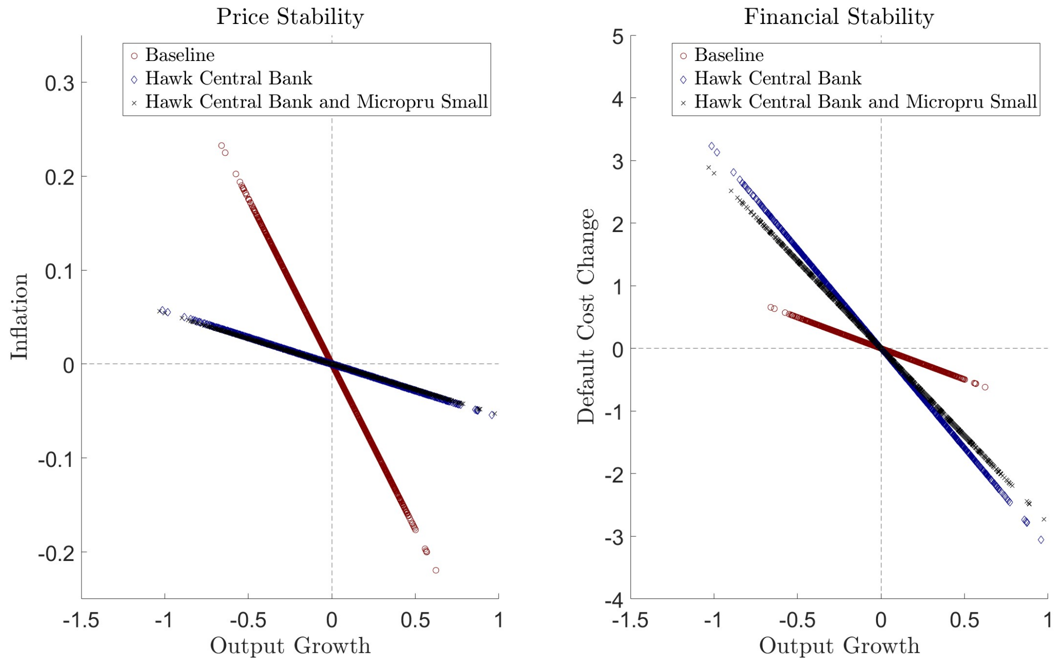
Notes: Cumulative impulse responses to a monetary shock that increases the nominal interest rate by 0.25 percent on impact, with quarterly persistence of 0.5.

Figure A.4: Impulse Response to TFP Shocks



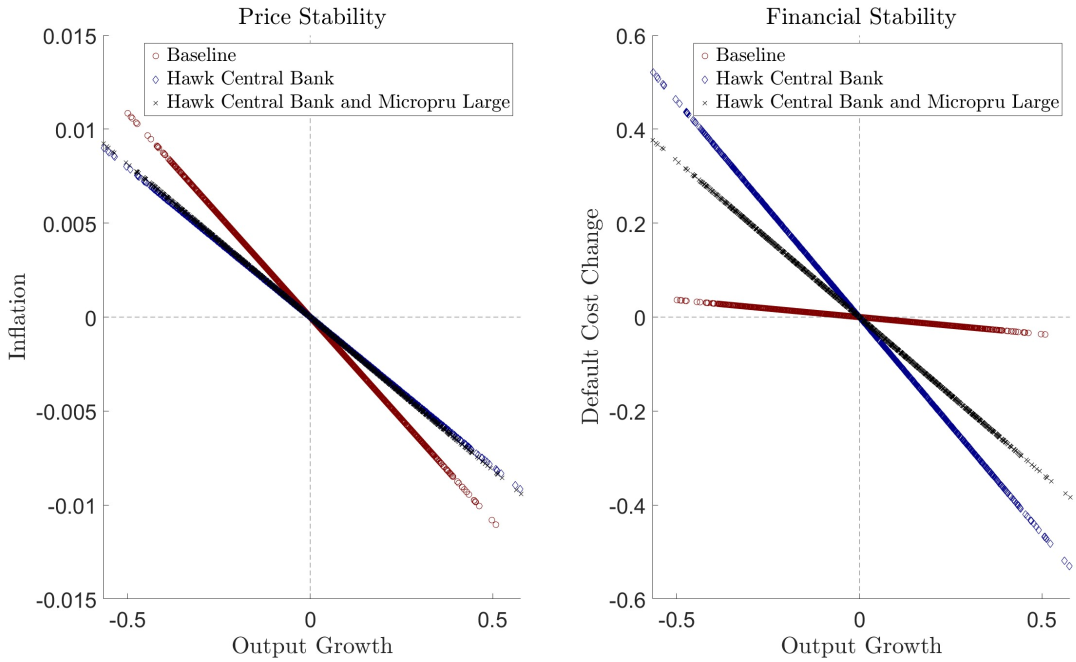
Notes: Impulse responses to a TFP shock that increases by 1 percent on impact, with quarterly persistence of 0.9.

Figure A.5: Macroeconomic-Financial Stabilization Trade-Off with Micropru Policy for Small Banks



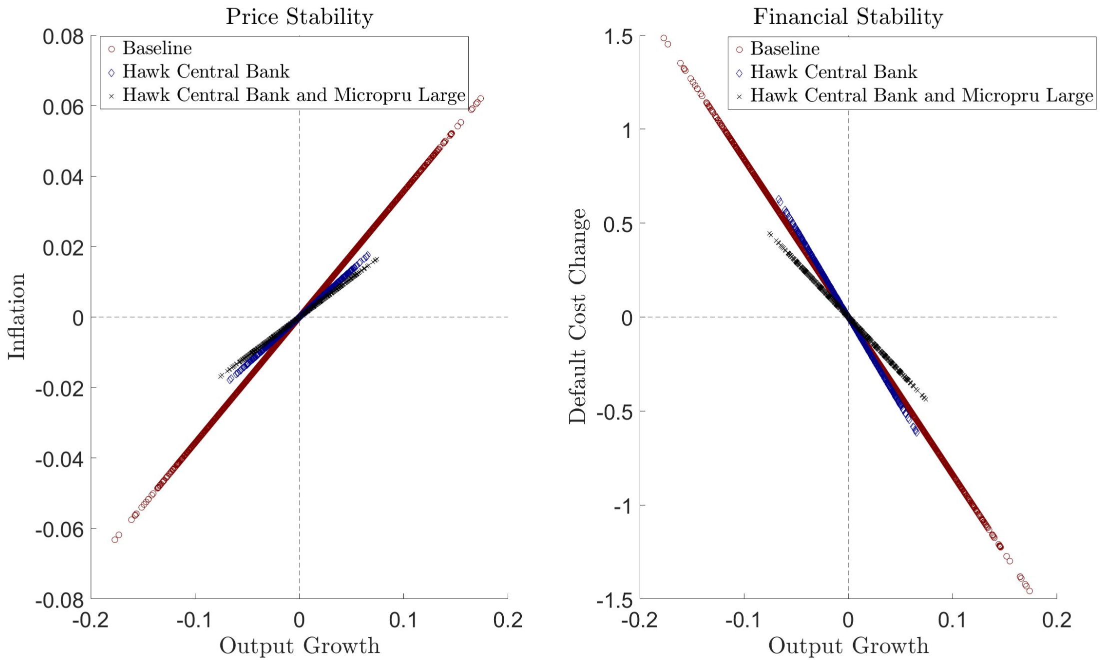
Notes: Output growth, inflation, and default cost realizations from a long stochastic simulation of the model with systematic microprudential policy that targets the smallest 75% banks.

Figure A.6: Macroeconomic-Financial Stabilization Trade-Off with a Flatter Phillips Curve



Notes: Output growth, inflation, and default cost realizations from a long stochastic simulation of the model with the slope of the Phillips Curve of 0.006 instead of 0.1.

Figure A.7: Macroeconomic-Financial Stabilization Trade-Off with Demand Shocks



Notes: Output growth, inflation, and default cost realizations from a long stochastic simulation of the model generated by non-systematic monetary shocks instead of TFP shocks.

B Empirical Appendix

B.1 Data Sources

In this Section, we describe the data sources for the variables used in our empirical analysis. We list all the variables with the associated data sources in Table B.1.

Call reports. Our main data source is the Federal Reserve Consolidated Reports of Condition and Income (Call reports), which provide quarterly financial statements for the universe of FDIC-insured U.S. banks. This dataset includes both income statement and balance sheet variables at quarterly frequency. Our sample covers the period 1990q1-2019q4. Our main variables of interest are total (book) assets and equity as well as total loans from the balance sheet and net income from the income statement. Following Corbae and D'Erasmo (2021), we aggregate bank level information to the Bank Holding Company level.

CoVaR. To proxy for bank-specific aggregate default costs, we use the ΔCoVaR measure developed in Adrian and Brunnermeier (2016). This metric captures the systemic impact of an individual bank's financial distress by measuring the change in the overall financial sector's value at risk (VaR) when bank i moves from its median state to a state of distress. Thus, ΔCoVaR_i can be interpreted as capturing the maximum potential loss faced by the broader financial sector conditional on the default of an individual bank. Specifically, we use the dollar CoVar measure, since we are interested in an absolute rather than relative cost of default. We use the 95% CoVaR measure, i.e., the one computed for the 95$^{\text{th}}$ quantile of equity returns of a given bank.

Monetary surprises. To capture monetary policy surprises, we follow the high-frequency identification approach. Specifically, we use the information-adjusted poor man's shock series from Jarociński and Karadi (2020) as our baseline instrument for monetary shocks. For robustness, following Gurkaynak et al. (2005) and Gertler and Karadi (2015), we also consider the unadjusted high-frequency shock defined as the change in the 3-month ahead Fed Funds futures within a 30 minute window around FOMC announcements as well as the Jarociński and Karadi (2020) information adjusted shock obtained with the median rotation that implements the sign restriction. Throughout our analysis, we normalize the sign of the measure of monetary shocks ε_t such that positive values are associated with *contractionary* shocks. Moreover, we also normalize ε_t to have unitary standard deviation.

Table B.1: Variable Details and Sources

Variable	Details	Source
GDP	U.S. real Gross Domestic Product, chained 2012 dollars	FRED (GDPC1)
Inflation	Consumer price index for all urban consumers: all items in U.S. city average	FRED (CPIAUCSL)
1 year treasury rate	Market Yield on U.S. Treasury Securities at 1-Year Constant Maturity, Quoted on an Investment Basis	FRED (GS1)
S&P500 return	Quarterly return on the S&P500 index	Compustat Global
Excess bond premium	Gilchrist and Zakrajšek (2012) excess bond premium	Favara et al. (2016)
Monetary surprise	Raw high-frequency, poor man's and sign-restricted shocks	Jarociński and Karadi (2020)
Assets	Total book assets	Call Reports (RCFD2170)
Loans	Loans and leases net of unearned income and allowance for loan and lease losses	Call Reports (RCFD2122-RCFD3123)
Equity	Total book equity	Call Reports (RCFD3210)
Net income	Net income	Call Reports (RIAD4340)
z-score	Construction procedure described in Appendix B.2	Authors' calculation
ΔCoVaR	Conditional value at risk	95% dollar CoVaR from Adrian and Brunnermeier (2016)

Notes: This table summarizes every empirical series used throughout the paper.

Macroeconomic aggregates. Finally, we obtain aggregate macroeconomic variables from the St. Louis Federal Reserve (FRED database), with the exception of the excess bond premium, which comes from Favara et al. (2016), and the return on the S&P500 index, which we download from Compustat Global.

B.2 *Z*-Score Construction

Following Laeven and Levine (2009), we use the z-score as our proxy for default probability. The z-score is defined as the ratio of the return on assets (RoA) plus the inverse leverage, divided by the standard deviation of RoA:

$$z_{it} = \frac{\text{RoA}_{it} + \text{Leverage}_{it}^{-1}}{\text{SD(RoA)}_{it}} \tag{B.1}$$

where RoA is defined as net income over total assets, leverage is defined as book assets over book equity, and SD(RoA) is a moving average of the bank-level standard deviation of RoA over some time window.[1]

[1] We choose a 5 years moving average of RoA for our baseline specification, but results are robust to different windows.

B.3 Additional Details on Figure 15

We now explain more in detail our procedure to construct the binned scatter plots in Figure 15. We split our sample into 30 equally-sized bins based on real assets, each including roughly 27,000 observations for panel (a) and 900 observations for panel (b). We then residualize the variables on both axes from a time fixed effect. Finally, for each bin we display average residualized real book assets against the average residualized default probability and default cost within each bin. We trim default probability, i.e., the inverse z-score, at the 2.5 and 97.5% level. As a proxy of aggregate default cost we use the 95% dollar CoVaR measure from Adrian and Brunnermeier (2016). The x-axis represents log real book assets for panel (a) and real book assets in level for panel (b).[2] For readability, we exclude the top 1% of observations in terms of real book assets from panel (b).

B.4 Additional Details on Local Projections

We now describe more in detail our procedure to estimate the panel local projections (36) and (37). We define our dependent variable as $\Delta Y_{i,t+h} \equiv \ln(Y_{i,t+h}) - \ln(Y_{i,t-1})$ for real book assets and $\Delta Y_{i,t+h} \equiv Y_{i,t+h} - Y_{i,t-1}$ for default probability (inverse z-score). Similarly, we define the lagged dependent variable as $\Delta Y_{it-\ell} \equiv \ln(Y_{i,t-1}) - \ln(Y_{i,t-\ell-1})$ for assets and $\Delta Y_{it-\ell} \equiv Y_{i,t-1} - Y_{i,t-\ell-1}$ for default probability. We trim our dependent variable in cumulative changes at the 1% and 99% level. Our sample covers the period 1990q1-2019q2.

B.5 Robustness and Additional Empirical Results

In this Section, we report robustness results and additional results for our empirical analysis.

Figure B.1 plots the behavior of default probability over time and in the cross-section of banks, together with the time series for the federal funds rate. Large banks consistently show a lower default probability than small ones. Unconditionally, default risk turns out to be positively correlated with the federal funds rate, and the correlation becomes stronger once one allows for some lag between the two series.

Figure B.2 replicates panel (b) of Figure 15 using CoVar weighted by default probability as the y-axis variable. This shows that also ex-post —that is, probability-weighted— aggregate default costs display a convex behavior in bank size, thus lending further support to our model calibration.

[2] We have also run a regression of log default costs on log real assets and the estimated coefficient is significantly above 1, thus confirming our convexity result.

Panel (a) of Figure B.3 plots estimates for the local projection specification in (36), using real "Loans and leases net of unearned income and allowance for loan and lease losses" as the dependent variable, as opposed to real book assets. Panel (b) does the same for specification (37).

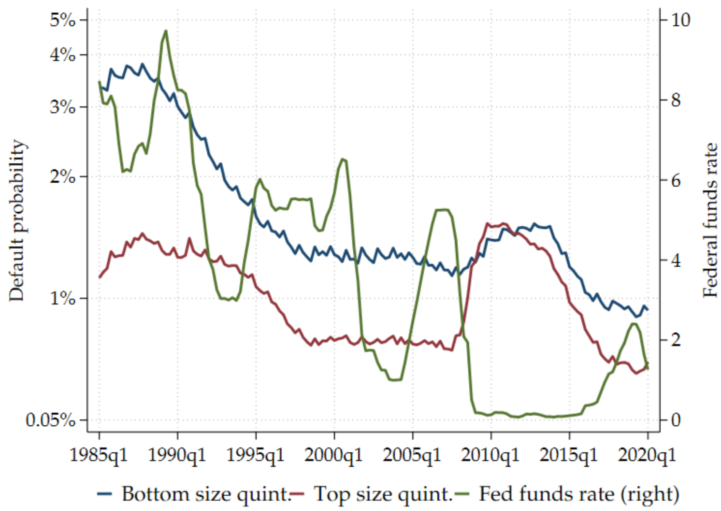
Figure B.4 shows that our estimates in the main text for the response of average bank assets and default probability to monetary shocks are robust to different measures of monetary surprises. In particular, we use the raw high frequency monetary shock as defined in Gertler and Karadi (2015), and updated by Jarociński and Karadi (2020) as well as the Jarociński and Karadi (2020) information adjusted shock obtained with the median rotation that implements the sign restriction. Figure B.5 does the same for the estimated heterogeneous responses.

Figure B.6 shows that our estimated heterogeneous responses of size and default probability in the cross-section of bank size are robust to different definition of the interaction term. Panels (a) and (c) plot results obtained defining the dummy variable for large banks as capturing the top 5% and 20% of the bank size distribution, as opposed to the top 10% in our baseline specification. Panels (b) and (d) consider the case of a continuous interaction, where the monetary shock is interacted with (the log of) bank assets. In all cases the results are consistent with our baseline specification.

Figure B.7 decomposes our finding in panel (a) of Figure 15 that default probability is decreasing in bank size, by leveraging the definition of z-score in (B.1). We find that the volatility of RoA is decreasing in bank size, while the RoA is increasing in bank size. Together, these two components contribute to the overall finding that default probability is decreasing in bank size. On the other hand, it is well known that book leverage increases with bank size, hence we find that inverse leverage strongly decreases with size. This slightly dampens the previous two forces in shaping the behavior of default probability in the cross-section of bank assets, but is not enough to counteract them.

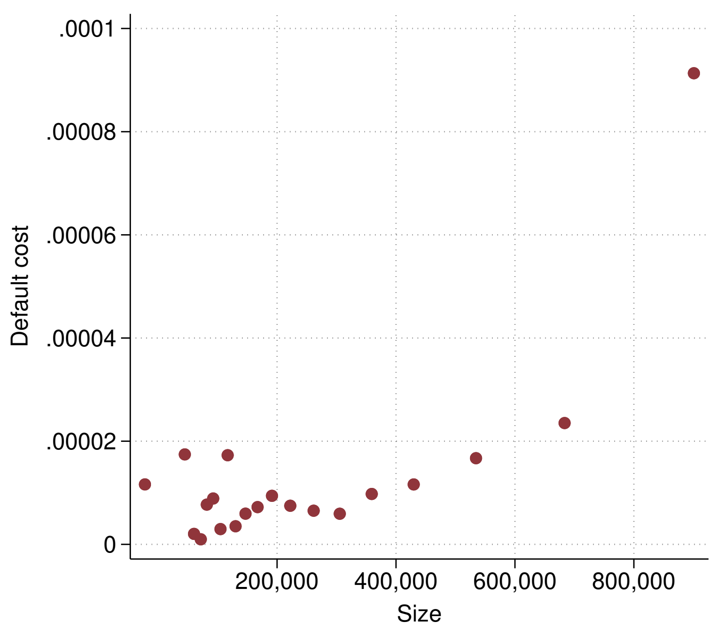
Figure B.8 performs a similar exercise and decomposes the overall heterogeneous response of default probability documented in panel (b) of Figure 17 into its different components. Remember that our overall finding is that default probability increases by less for large banks following a monetary contraction. Panels (a) shows that this is mainly driven by volatility of RoA increasing by less for large banks and, to a lesser extent, from the RoA decreasing by less, as showed in panel (b).

Figure B.1: Default probability over time and in the cross-section



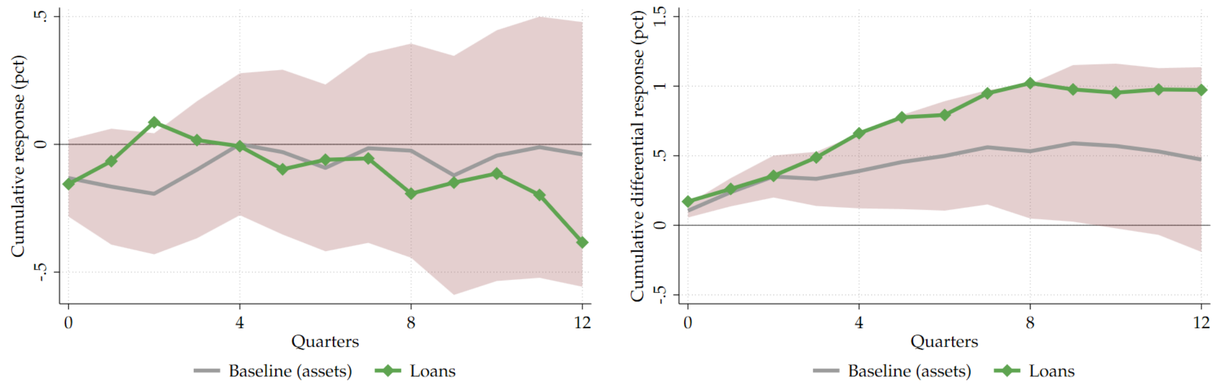
Notes: this figure plots the federal funds rate (on the right scale) together with average probability of default for banks in the bottom and top quintile of the asset distribution over time. The average probability of default is computed as the inverse z-score.

Figure B.2: Probability-weighted default cost in the cross-section of banks



Notes: binned scatter plot of default-probability-weighted default cost against bank size. We proxy default probability with the inverse z-score and default cost with the 95% dollar CoVaR from Adrian and Brunnermeier (2016). Both axis are residualized by time fixed effects. See Appendix B.3 for additional details.

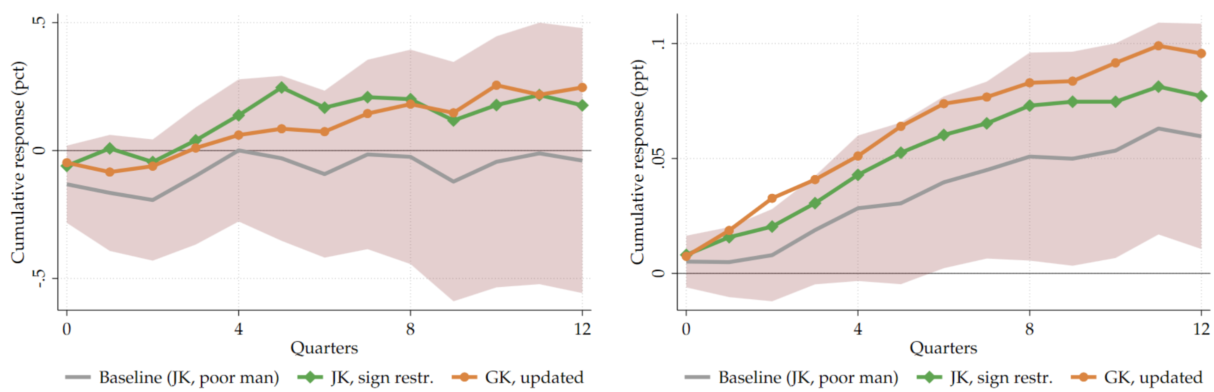
Figure B.3: Loans, average and heterogeneous response



(a) Average response **(b)** Heterogeneous response

Notes: panel (a) plots the estimated ψ_h from (36) to a 1 standard deviation contractionary monetary shock. The y-axis represents the cumulative percentage change in total real loans. Panel (b) plots the estimated β_h from (37) to a 1 standard deviation contractionary monetary shock. The y-axis represents the cumulative percentage change in total real loans relative to banks in the bottom 90% of the asset distribution. The x-axis represents quarters elapsed since the shock. Errors are two-way clustered at the time and bank level. Lightly (darkly) shaded areas represent 90% (68%) confidence intervals.

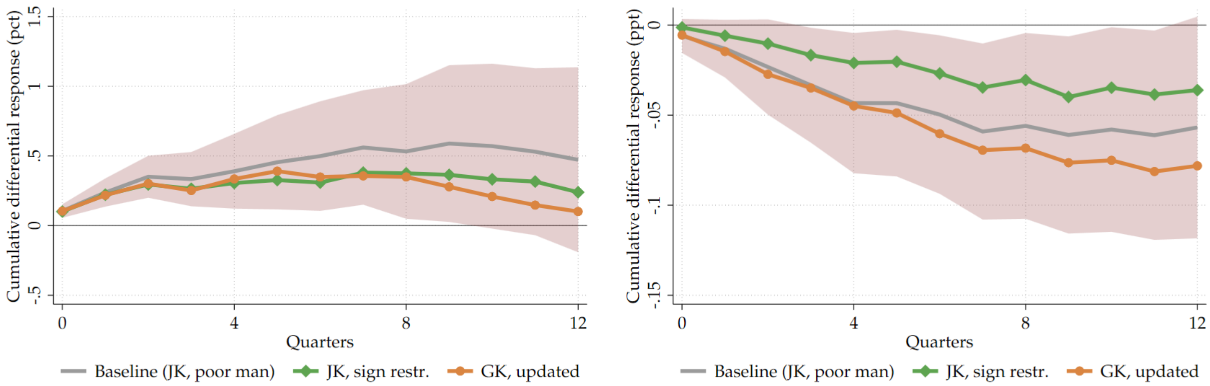
Figure B.4: Average response: robustness to different monetary shocks



(a) Assets **(b)** Default probability

Notes: estimated ψ_h from (36) to a 1 standard deviation contractionary monetary shock for different measures of monetary surprises. The y-axis represents the cumulative percentage change in total real assets in panel (a) and the cumulative level change in default probability —as proxied by the inverse z-score— in panel (b). The x-axis represents quarters elapsed since the shock. Errors are two-way clustered at the time and bank level. Lightly (darkly) shaded areas represent 90% (68%) confidence intervals.

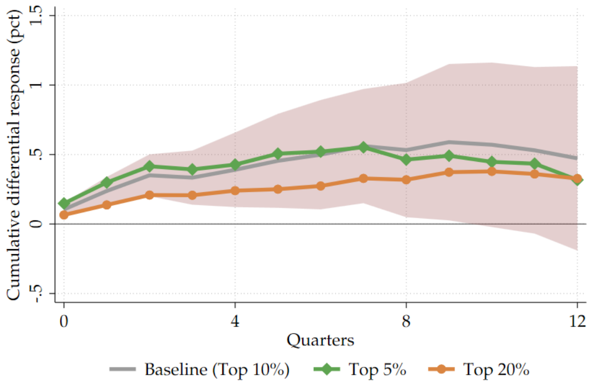
Figure B.5: Heterogeneous response: robustness to different monetary shocks



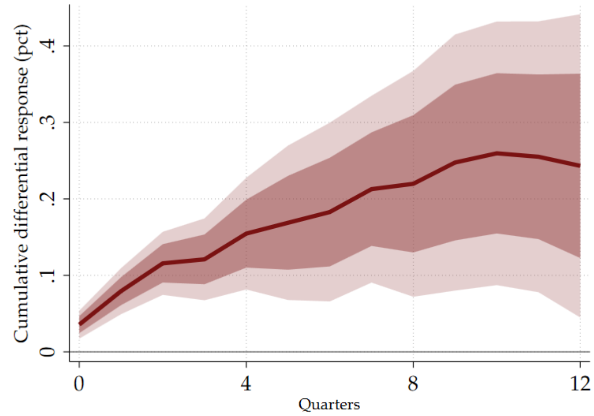
(a) Assets **(b)** Default Probability

Notes: estimated β_h from (37) to a 1 standard deviation contractionary monetary shock for different measures of monetary surprises. The y-axis represents the cumulative percentage change in total real assets in panel (a) and the cumulative level change in default probability —as proxied by the inverse z-score— in panel (b) relative to banks in the bottom 90% of the asset distribution. The x-axis represents quarters elapsed since the shock. Errors are two-way clustered at the time and bank level. Lightly (darkly) shaded areas represent 90% (68%) confidence intervals.

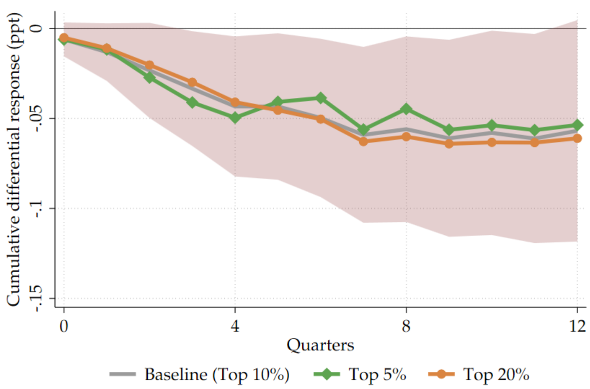
Figure B.6: Heterogeneous response: robustness to different size interactions



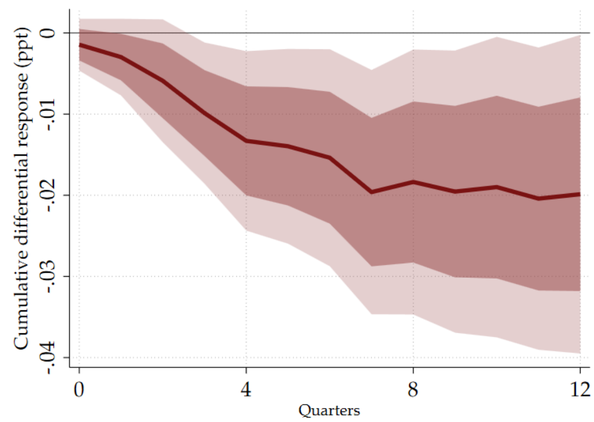
(a) Assets, different top dummies



(b) Assets, continuous interaction



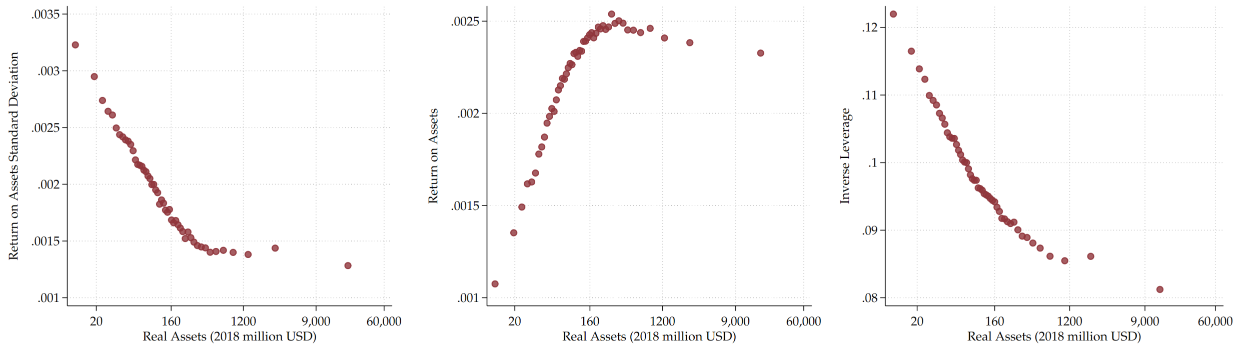
(c) Default probability, different top dummies



(d) Default probability, continuous interaction

Notes: estimated β_h from (37) to a 1 standard deviation contractionary monetary shock for different specifications of the size interaction terms. The y-axis represents the cumulative percentage change in total real assets in panels (a) and (c) and the cumulative level change in default probability —as proxied by the inverse z-score— in panels (b) and (d). The x-axis represents quarters elapsed since the shock. Errors are two-way clustered at the time and bank level. Lightly (darkly) shaded areas represent 90% (68%) confidence intervals. See text for more details.

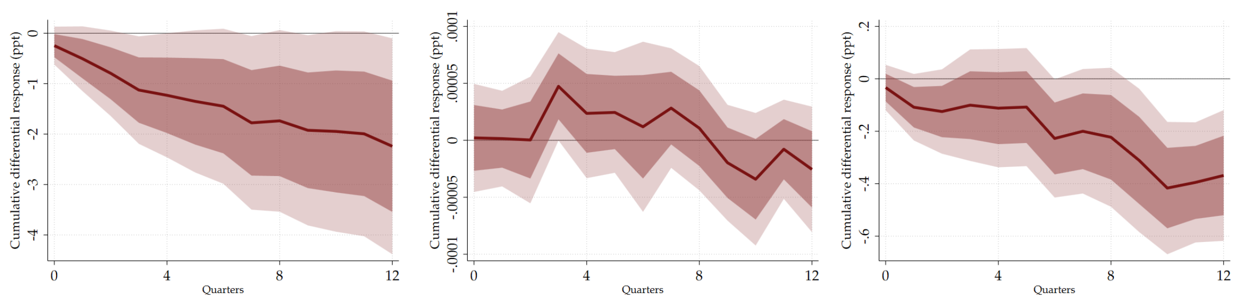
Figure B.7: Z-score Decomposition



(a) Standard Deviation of RoA **(b)** Return on Assets **(c)** Inverse Leverage

Notes: decomposition of the binned scatter plot in panel (a) of Figure 15 according to (B.1). Both axis are residualized by time fixed effects.

Figure B.8: Z-score Heterogeneous Response Decomposition



(a) Standard Deviation of RoA **(b)** Return on Assets **(c)** Inverse Leverage

Notes: estimated β_h from (37) to a 1 standard deviation contractionary monetary shock for the different components of the inverse z-score measure as constructed in (B.1). The y-axis represents cumulative level changes. The x-axis represents quarters elapsed since the shock. Errors are two-way clustered at the time and bank level. Lightly (darkly) shaded areas represent 90% (68%) confidence intervals.

References

Adrian, Tobias and Markus K Brunnermeier, "CoVaR," *American Economic Review*, July 2016, *106* (7), 1705–41.

Corbae, Dean and Pablo D'Erasmo, "Capital Buffers in a Quantitative Model of Banking Industry Dynamics," *Econometrica*, 2021, *89* (6), 2975–3023.

Favara, Giovanni, Simon Gilchrist, Kurt F Lewis, and Egon Zakrajšek, "Recession Risk and the Excess Bond Premium," *FEDS Notes*, 2016, *April 8.*

Gertler, Mark and Peter Karadi, "Monetary Policy Surprises, Credit Costs, and Economic Activity," *American Economic Journal: Macroeconomics*, January 2015, *7* (1), 44–76.

Gilchrist, S. and E. Zakrajšek, "Credit Spreads and Business Cycle Fluctuations," *American Economic Review*, 2012, *102(4).*

Gurkaynak, R., B. Sack, and E. Swanson, "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements," *International Journal of Central Banking*, 2005, *1(1).*

Jarociński, Marek and Peter Karadi, "Deconstructing Monetary Policy Surprises - The Role of Information Shocks," *American Economic Journal: Macroeconomics*, April 2020, *12* (2), 1–43.

Laeven, Luc and Ross Levine, "Bank governance, regulation and risk taking," *Journal of Financial Economics*, 2009, *93* (2), 259–275.