

HBANK: Monetary Policy with Heterogeneous Banks[†]

Marco Bellifemine
LSE

Rustam Jamilov
Oxford

Tommaso Monacelli
Bocconi

December 2023

Abstract

We study monetary, liquidity, and macroprudential policy transmission in a Heterogeneous Bank New Keynesian (HBANK) model that is solved in sequence space. Using a sufficient-statistic approach, we show that the combination of incomplete markets and costly bank insolvency breaks the “as-if” result, generating substantial amplification of policy shocks relative to the representative-bank benchmark. There is a trade-off between macroeconomic and financial stabilization: contractionary monetary policy worsens financial stability by raising the likelihood of bank insolvency in the lower tail of the bank size distribution. We enrich our baseline framework with departures from perfect deposit and credit market competition and apply it to the study of monetary, forward guidance, macroprudential, and reserve requirement policies. We validate our model empirically with novel cross-sectional and time-series facts on U.S. commercial banks.

[†]We thank Javier Bianchi, Saki Bigio, Sebastian Fanelli (discussant), Pablo D’Erasmus, Peter Karadi, Luc Laeven, Ben Moll, Dima Mukhin, Pablo Ottonello, Vincenzo Quadrini, H el ene Rey, Farzad Saidi, Ludwig Straub, Alejandro Van der Ghote, Ia Vardishvili (discussant), Ivan Werning, Russell Wong and seminar participants at numerous venues for valuable comments and suggestions. Bellifemine: m.bellifemine@lse.ac.uk. Jamilov: rustam.jamilov@all-souls.ox.ac.uk. Monacelli: tommaso.monacelli@unibocconi.it. First version: July 2022.

1 Introduction

The emphasis on the role of financial intermediaries in the transmission of monetary policy has strengthened after the 2007-08 financial and credit crisis. A large literature acknowledges that disruptions in financial intermediation can have significant effects on economic activity (Brunnermeier and Pedersen, 2009; He and Krishnamurthy, 2013; Brunnermeier and Sannikov, 2014). However, most of the recent work in macroeconomics on the link between financial intermediation and monetary policy still abstracts from the implications of heterogeneity and imperfect insurance that characterize the full distribution of financial intermediaries.

In this paper we study the impact of bank heterogeneity on the transmission of monetary policy in a Heterogeneous Bank New Keynesian (HBANK) model. Our general framework combines four main features: (i) incomplete financial markets and uninsured idiosyncratic bank rate of return risk, (ii) costly bank insolvency, (iii) deposit insurance, and (iv) nominal rigidities. Our setup nests both the canonical New Keynesian model (Woodford, 2003; Gali, 2008) and the Gertler and Kiyotaki (2010); Gertler and Karadi (2011) macro-banking framework as special cases.

We solve our model in the sequence space domain and characterize the general-equilibrium solution in terms of measurable sufficient statistics following a burgeoning methodological literature (Boppart et al., 2018; Auclert et al., 2021a). To first order, the bank lending channel (Bernanke and Gertler, 1995; Kashyap and Stein, 2000) can be conveniently summarized by two sufficient statistics: (i) a policy Jacobian and (ii) the intertemporal marginal propensity to lend (iMPL). The former collects the partial-equilibrium responses of bank lending to a policy intervention, e.g., monetary or macroprudential, holding the value of aggregate capital constant. The latter completely determines the general-equilibrium feedback effect on aggregate capital, which can either amplify or dampen the first-round effect. To the extent that non-financial firms depend on banks for external financing, the sequence of bank lending is sufficient to recover every other endogenous object in the model, such as inflation, output or bank deposits. The response of every aggregate variable to policy shocks can be immediately decomposed into direct and indirect components. All Jacobians and iMPLs objects are readily measurable.

Our approach offers substantial payoffs when it comes to extending the basic macro-banking framework with additional policy-relevant questions and frictions from the banking literature. First, we show that departures from the perfect credit and deposit market competition assumptions involve simply augmenting and re-computing the two standard sufficient statistics (policy Jacobian and iMPL). Second, our framework can study

the macroeconomic effects of macroprudential or liquidity policy interventions at very low costs. Raising capital requirements, for example, can be simulated in the model with the introduction of just two new terms: the macroprudential policy Jacobian and the path of capital requirements which is inputted by the econometrician. Third, our approach is especially useful for the analysis of *news* shocks of policy changes at some future horizon. Both sufficient-statistic objects are *matrices* with each column representing a policy shock at some horizon $s \geq 0$ and each row representing the time of the response. Thus, the approach is ideal for the study of forward guidance announcements.

The key theoretical channel at work in the model is costly bank *insolvency risk*, which is the main proxy for financial stability considerations. Due to market incompleteness and scale-variance, there is a risk that bank-level net worth can hit the zero-bound (insolvency) constraint due to a very large negative idiosyncratic return draw. Following a large and influential literature, we assume that this risk, despite the presence of deposit insurance, matters for the real economy because bank default is costly (Hoggarth et al., 2002; Laeven and Valencia, 2012).

Our first main result is that the HBANK framework with costly bank default generates *amplification* of monetary policy shocks relative to the representative-bank (RBANK, for short) benchmark. In the literature jargon, the model breaks the “as-if” result (Krusell and Smith, 1998; Werning, 2015), i.e., the model with heterogeneous banks does not behave in the aggregate as if it was governed by a representative bank. The intuition is simple: market incompleteness and idiosyncratic bank return risk generate a right-skewed stationary distribution of bank size with a non-trivial mass of small banks that feature low distance to default and high sensitivity to aggregate shocks. Ex post, a non-trivial share of banks defaults due to insolvency. Since insolvency is costly, there is a first-order link with aggregate profitability, so that the economy turns out to be more elastic and rate-sensitive than the RBANK benchmark. We show that the degree of amplification of policy shocks in a calibrated model that targets financial stability moments from the literature is substantial. This applies to multiple policy instruments: conventional monetary policy, forward guidance, macroprudential regulation, and interest rate on reserves.

The above finding is important for the following reason. It is well-known, since at least Krusell and Smith (1998), that in a model with heterogeneous agents, even if the *level* of macroeconomic aggregates differs, the *elasticity* of the same aggregates to shocks and/or policy changes does not necessarily differ from that of a representative-agent counterfactual. It has been shown in the literature that only some features, such as cyclical income risk, could break the as-if hurdle. Thus, our amplification result applies more generally to the influential literature on agent heterogeneity (Krueger et al., 2016;

Kaplan et al., 2018). In fact, and in order to illustrate this point in the clearest possible way, we show that our HBANK model *without* costly default generates impulse responses to non-systematic monetary policy shocks that are quantitatively very similar to those from the RBANK version of the model.

Our second result follows immediately from the above discussion. There is a trade-off for a monetary authority that wishes to stabilize inflation *and* preserve financial stability. This finding is similar to the result in Coimbra and Rey (2023) in the context of financial intermediaries that are ex-ante heterogeneous in the Value-at-Risk constraint. The elasticity of aggregate bank default risk to changes in the real interest rate is positive, while the elasticity of bank lending to changes in default risk is negative. Thus, suppose the economy is hit by a rise in aggregate demand, raising both output and inflation, and requiring a higher real interest rate in response. A higher real interest rate also 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.

Third, we extend our framework to the analysis of policy instruments beyond conventional monetary policy. We look at the effects of non-systematic changes in forward guidance, macroprudential policy, and minimum reserve requirements. The analysis of forward guidance shocks is particularly convenient in our framework because of the Sequence-Space representation of every “policy function”: the general equilibrium response at time t to an announced interest rate shock at time $t + 10$ is known immediately from the Jacobian matrices that summarize the response at t to shocks at every horizon $s \geq 0$. We show that, similarly to conventional monetary policy, forward guidance shocks are considerably amplified in a model with bank heterogeneity. The intuition for this result is that small banks, which are closer to default, are relatively more responsive to interest rate changes at any horizon than the average bank. At the same time, we show that our HBANK model does not suffer from the forward guidance puzzle: the size of the response of all macroeconomic variables to future interest rate shocks *decreases* with the horizon of the shocks.

Fourth, we show how our flexible framework can be extended to accommodate changes to the micro-foundations of the banking block. In particular, we relax the assumption of perfect banking competition and introduce endogenous and heterogeneous credit mark-ups and deposit mark-downs. An extensive literature has previously shown that bank market power has first-order effects on policy-making and the macroeconomy (Scharfstein and Sunderam, 2016; Drechsler et al., 2017, 2021; Corbae and D’Erasmus, 2021). We show that solving the extended, more complex model in sequence space involves sim-

ply re-computing the policy Jacobians and the iMPL objects. We find that imperfect competition in deposit (credit) markets dampens (amplifies) the macroeconomic effects of monetary policy shocks. While the former impact is quantitatively very large, the latter is minute.

Fifth and finally, we validate our model using micro-data on U.S. commercial banks. First, we show, both in the model and in the data, that bank size heterogeneity matters for the responsiveness to monetary policy shocks. Smaller banks' balance sheets are significantly more responsive than the ones of larger banks, a result that echoes previous evidence in [Kashyap and Stein \(1995\)](#). This result confirms that a complete theoretical and quantitative understanding of the bank lending channel of monetary policy transmission requires a model with some realistic degree of bank heterogeneity. Second, we document a robust link between bank *size* and insolvency-driven *default risk*. We measure bank default risk with two complementary indices: the so-called "Z-score" ([Laeven and Levine, 2009](#)) and distance to default ([Nagel and Purnanandam, 2019](#)). We show that smaller banks are considerably closer to insolvency, both in the cross section and in the time series. Therefore, our model delivers the right cross-sectional patterns and correlations between size and default riskiness. Third, we find strong evidence in the data that default risk increases conditional on identified positive monetary policy shocks. The sign of this conditional moment is a key mechanism required for the amplification of policy shocks in the model.

Literature review Our paper builds on several literature strands from macroeconomics, finance, and monetary economics. First, we are contributing to the so-called "macro-banking" literature which incorporates financial frictions into otherwise standard macroeconomic frameworks. There are two broad complementary directions in this literature. Some studies introduce *market*-based constraints on risk-taking that, generally speaking, generate counter-cyclical amplification of aggregate shocks. Papers in this strand include [Gertler and Kiyotaki \(2010\)](#), [Gertler and Karadi \(2011\)](#), [Brunnermeier and Sannikov \(2014\)](#), [He and Krishnamurthy \(2013\)](#), [Jermann and Quadrini \(2013\)](#), [Nuno and Thomas \(2017\)](#), [Bocola \(2016\)](#), [Gertler et al. \(2016, 2020\)](#), [Mendicino et al. \(2020\)](#), [Elenev et al. \(2021\)](#), among many others. On the other hand, studies such as [Brunnermeier and Pedersen \(2009\)](#), [Adrian and Shin \(2010\)](#), [Adrian and Boyarchenko \(2015\)](#) introduce book-based constraints on risk taking which thus do not differentiate between market or book leverage ratios. A notable exception is [Begenau et al. \(2021\)](#) who propose a unifying approach to modeling book vs market leverage of banks. Our paper adds to the market-based strand and particularly emphasizes the role of bank heterogeneity in a literature that otherwise

mostly embraces the representative intermediary assumption.

Second, we are building on the fast-expanding literature on heterogeneous financial intermediaries. This literature can be further divided into two subsets. The first set studies environments where intermediaries feature *permanent* ex-ante heterogeneity. For example, [Coimbra and Rey \(2023\)](#) develop a general equilibrium framework with endogenous entry and where financial intermediaries are heterogeneous in their Value-at-Risk constraints. [Begenau and Landvoigt \(2022\)](#) build a quantitative model with two banking sectors that approximate the empirically-documented divide between standard commercial and “shadow” banks. The second subset of the literature introduces some form of bank-level non-systematic risk such that intermediaries are generally ex-ante identical but heterogeneous ex post. For example, [Bianchi and Bigio \(2022\)](#) study the credit channel of monetary policy in an environment where bank deposits circulate in an unpredictable way and banks face deposit withdrawal shocks. [Rios Rull et al. \(2020\)](#) study aggregate effects of capital requirements in a quantitative model with non-diversifiable credit risk. Relative to these two literature strands our contribution is to incorporate stochastic bank returns heterogeneity into a canonical macro-banking environment, following our previous work in [Jamilov and Monacelli \(2023\)](#), and link these features with monetary, liquidity, and macroprudential policy in a New Keynesian framework.

Methodologically, our paper builds on the burgeoning literature that solves complex general equilibrium macroeconomic models with sequence-space methods ([Boppart et al., 2018](#); [Auclert et al., 2021a](#)). The sequence space approach has been recently applied to the case of household heterogeneity ([Auclert et al., 2020, 2023](#)), input and output frameworks ([Schaab and Tan, 2022](#)), exchange rates ([Auclert et al., 2021b](#)), regional heterogeneity ([Bellifemine 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 New-Keynesian macro-banking framework with heterogeneous intermediaries. In doing so, we point out a novel sufficient statistic for characterizing the general equilibrium implications of the bank lending channel of monetary policy transmission: the intertemporal Marginal Propensity to Lend (iMPL).

Overall, the tractability payoffs from the sequence-space approach are substantive. All policy-relevant banking frictions and all details that shape the model’s micro-foundations typically result in linear reduced-form vector-valued equations that can be managed with simplicity.

Finally, we are contributing to the vast literature that quantifies the role of heterogeneity, financial frictions, or both for monetary policy-making. [Lee et al. \(2020\)](#) introduce frictional financial intermediation into the canonical HANK literature ([Galí et al., 2007](#);

Bilbiie, 2008; Kaplan et al., 2018; Auclert, 2019; Bayer et al., 2019; Ravn and Sterk, 2020; Acharya et al., 2023). Their representative-bank friction, which is similar to the one that we impose in our set-up, amplifies monetary policy and gives rise to consumption inequality. Our approach is different but conceptually similar: we focus on heterogeneous intermediaries but keep the household block very simple. Bigio and Sannikov (2021) build an incomplete-markets environment with wage rigidities where the central bank controls credit spreads and interest rate targets via the supply of reserves. In important related work, Baqaee et al. (2023) uncover the supply side of monetary policy in a model with heterogeneity and endogenous product market power of non-financial firms. Ottonello and Winberry (2020) quantify the investment channel of monetary policy in the case of non-financial firms that are heterogeneous in their riskiness and distance to default. Kaplan et al. (2020) emphasize the role of housing and long-term mortgages in the dynamic of credit conditions leading up to the Great Financial Crisis. Lenel and Kekre (2022) study a HANK environment with heterogeneity in marginal propensity to take risk (MPR) and show how endogenous risk premia fluctuations amplify monetary shocks. Our contribution is to zoom in both empirically and quantitatively on the roles of *bank* balance sheet heterogeneity and default risk channels of monetary transmission in an otherwise textbook New Keynesian model with endogenous capital accumulation and financial frictions.

2 A New-Keynesian Model with Heterogeneous Banks

Below we lay out the baseline version of our model, featuring heterogeneous intermediaries, incomplete markets, costly insolvency risk, and nominal rigidities. Later we extend the baseline setup to the case of market power on both the asset and the liability side of banks' balance sheet.

2.1 Model Formulation

Capital Producers

Capital is required for the production of a final good. Capital good producers are cash-strapped and require bank financing in the form of equity-type claims $l_{j,t}$ ("loans"). We assume that these firms possess a technology to costlessly convert loans into differentiated units of capital $k_{j,t}$, which get immediately aggregated into the capital stock K_t . Capital depreciates fully every period. Competition is assumed to be perfect which means that all claims are priced to the marginal cost. We later introduce monopolistic loan-market

competition.

Households

The representative household supplies labor inelastically (normalized to unity) and derives utility from consumption. The household can save in the form of one-period deposits or mutual funds. The utility index, which is increasing and strictly concave in consumption, is defined as:

$$U(C_t) = \frac{C_t^{1-\psi}}{1-\psi} \quad (1)$$

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

$$C_t + \int_0^1 b_{j,t} dj + M_t \leq R_t M_{t-1} + W_t + \int_0^1 R_{j,t}^b b_{j,t-1} + Div_t + T_t \quad (2)$$

where M_t are mutual fund holdings, W_t is the real competitive wage rate, $R_{j,t}^b$ is the non-contingent bank-specific interest rate on deposits, R_t is the real risk-free interest rate, Div_t are lump-sum transfers of bank dividends, and T_t are lump-sum transfers/taxes. In the baseline model, there are no fundamental differences between mutual funds and deposits, which equalizes their returns. We introduce deposit market power later in the paper, which generates an equilibrium mark-down on the deposit rate. For now, competitive pricing ensures that $R_{j,t+1}^b = R_{t+1}$ for all j .

Bank Heterogeneity

Financial intermediaries accumulate net worth $n_{j,t}$ by sourcing household deposits $d_{j,t}$ at the state non-contingent rate $R_{j,t+1}^b$, and investing into firms' claims $l_{j,t}$ which are priced at $q_{j,t}$. At the moment, firm claims are the sole asset in the economy; we introduce reserves later in the paper along with a minimum reserve requirement and an interest rate on reserves which constitutes a new policy choice.

In exchange for purchasing claims on capital, banks receive the realized aggregate return on the capital stock R_{t+1}^k which is perturbed by a bank-specific component $\kappa_{j,t}$. Markets are incomplete and $\kappa_{j,t}$ represents uninsured idiosyncratic rate of return risk in the spirit of [Benhabib and Bisin \(2018\)](#); [Benhabib et al. \(2019\)](#). Bank-specific return on investment $R_{j,t}^T$ therefore equals $R_{j,t}^T = \kappa_{j,t} R_t^k$. We postulate that $\kappa_{j,t}$ follows an AR(1) process:

$$\kappa_{j,t} = \bar{\kappa} + \rho \kappa_{j,t-1} + \sigma_\kappa \epsilon_{j,t} \quad (3)$$

In order to eliminate scale invariance and make bank net worth a relevant state variable, we introduce convex asset adjustment costs that are governed by the dyad $\{\zeta_1, \zeta_2\}$.

Alternatively, one can view them as non-interest expenses. The law of motion of bank net worth is therefore:

$$n_{j,t+1} = R_{j,t+1}^T q_{j,t} l_{j,t} - R_{j,t+1}^b b_{j,t} - \zeta_1 l_{j,t}^{\zeta_2} \quad (4)$$

The balance sheet constraint must bind at all times:

$$b_{j,t} + n_{j,t} = q_{j,t} l_{j,t} \quad (5)$$

Bank Insolvency Risk

Due to the presence of market incompleteness and uninsured idiosyncratic risk, banks can become insolvent. The *ex-ante* probability of hitting the zero-net-worth bound can be defined as $\varphi_{j,t} = \mathbb{E}_t(\Pr(n_{j,t+1} < 0))$. Bank deposits are fully insured by the government and funded via lump-sum taxes on the household. Deposit insurance nullifies *ex-ante* bank insolvency risk, and deposits are still priced at the risk-free rate. The *ex-post* realized mass of insolvent banks is denoted by s_t . We assume that default is costly: a fraction ν of the final good gets eroded as a result of bank failure. The parameter ν represents the real cost of banking crises and will be used to target empirical estimates in the literature, such as in [Laeven and Valencia \(2012\)](#). Thus, total realized default costs can be written as:

$$S_t = \nu s_t Y_t \quad (6)$$

The realized return on aggregate capital is then $R_t^k = (1 - S_t)R_t^{K^*}$ where $R_t^{K^*}$ is the marginal product of capital, i.e., the return without default losses. Finally, we assume that insolvent banks are immediately replaced by new entrants whose size-return profile is the average of the stationary distribution.

Dynamic Bank Problem

In order to motivate macroprudential regulation and introduce a hard cap on leverage, we allow for a moral-hazard driven constraint of the [Gertler and Kiyotaki \(2010\)](#); [Gertler and Karadi \(2011\)](#) form:

$$\lambda q_{j,t} l_{j,t} \leq V_t(n_{j,t}, s_{j,t}) \quad (7)$$

where $V_t(n_{j,t}, s_{j,t})$ is the franchise value of bank j . Notice that the constraint above is occasionally binding.

The full dynamic problem of bank j can thus be summarized as:

$$V_t(n_{j,t}, s_{j,t}) = \max_{\{l_{j,t}, b_{j,t}, q_{j,t}, R_{j,t+1}^b\}} \mathbb{E}_t \left[\Lambda_{t+1} \left((1 - \sigma)n_{j,t+1} + \sigma V_{t+1}(n_{j,t+1}, s_{j,t+1}) \right) \right]$$

subject to:

$$\begin{aligned}
n_{j,t+1} &= R_{j,t+1}^T q_{j,t} l_{j,t} - R_{j,t+1}^b b_{j,t} - \zeta_1 l_{j,t}^{\zeta_2} \\
b_{j,t} + n_{j,t} &= q_{j,t} l_{j,t} \\
\lambda q_{j,t} l_{j,t} &\leq V_t(n_{j,t}, s_{j,t}) \\
R_{j,t}^T &= \kappa_{j,t} R_t^k \\
R_{j,t+1}^b &= R_{t+1}
\end{aligned}$$

where $1 - \sigma$ is the fixed dividend payout rule and $\Lambda_{t,t+1}$ is the stochastic discount factor which equates the marginal rate of substitution in the households block.

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

$$Y_t = \left(\int_0^1 y_{i,t}^{\frac{\gamma-1}{\gamma}} di \right)^{\frac{\gamma}{\gamma-1}} \quad (8)$$

where $\gamma > 1$ is the elasticity of substitution between differentiated goods. Each retailer rents labour and capital to produce intermediate goods using a constant returns to scale production technology.

$$y_{i,t} = A_t K_{i,t}^\alpha \quad (9)$$

Retailers set a relative price for their variety $p_{i,t}$ and pay quadratic adjustment costs $\frac{\varphi}{2} \left(\frac{p_{i,t}}{p_{i,t-1}} - 1 \right)^2 Y_t$. 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 relative 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. The final good gets consumed every period.

Retailers' symmetrical problem yields the conventional Phillips curve relationship:

$$\log \Pi_t = \frac{\gamma - 1}{\varphi} (\log MC_t - \log MC^*) + E_t [\Lambda_{t+1} \log \Pi_{t+1}] \quad (10)$$

We assume that the central bank follows a real interest rate rule, specifically setting R_t to a constant \bar{R} . This simplification avoids the feedback loop between nominal rates and inflation, simplifying the analysis in the following section without affecting any of our key insights.

2.2 Solution in Sequence Space

Intertemporal Constraints We now recast our model into the sequence space domain. First, we begin with the law of motion of bank net worth:

$$n_{j,t} = \underbrace{\left[R_{j,t}^T - R_{j,t}^b \right]}_{\text{Excess Return } X_{j,t}} q_{j,t} l_{j,t} + \underbrace{R_{j,t}^b}_{\text{Interest Expense}} n_{j,t-1} - \underbrace{E_{j,t}}_{\text{Non-Interest Expense}}$$

We derive its intertemporal version by imposing a version of the transversality condition:

$$n_{j,t} = \mathbb{E}_t \sum_{s=1}^{\infty} \left[\left(E_{j,t+s} - X_{j,t+s} q_{j,t+s} l_{j,t+s} \right) \prod_{\ell=1}^s R_{j,t+\ell}^{b^{-1}} \right] \quad (11)$$

Next, recall the household's period-by-period budget constraint from Equation 2. The intertemporal version can be shown to be the following:

$$\mathbb{E}_t \sum_{s=0}^{\infty} \left(C_{t+s} - W_{t+s} - Div_{t+s} - T_{t+s} \right) \prod_{l=0}^s R_{t+l}^{-1} = M_{t-1} + \int_0^1 b_{j,t-1} \quad (12)$$

Input and Output Sequences

From the intertemporal bank law of motion of net worth (Equation 11), we deduce that bank choices $\{q_{j,t}, l_{j,t}, b_{j,t}, R_{j,t}^b\}$ depend only on the full sequences of real rates and aggregate capital $\{R_s, K_s\}_{s=0}^{\infty}$, of which only the sequence of aggregate capital is an endogenous "state", in turn determining the common component of bank income.

The four bank outcomes can be written with the use of *outcome functions* $Y_{o,t} = \mathcal{Y}_{o,t}(\{X\}_{s=0}^{\infty})$, for each outcome o . Outcome functions map an *input sequence* $\{I_s\}_{s=0}^{\infty}$ into the output $Y_{o,t}$. We can now define a key object for our analysis - the aggregate lending function - as follows:

$$L_t = \mathcal{L}_t(\{R_s, K_s\}_{s=0}^{\infty}) \quad (13)$$

Clearly, from the bank balance sheet constraint (5), the choice $b_{j,t}$ is redundant: it is pinned down as soon as $q_{j,t}$ and $l_{j,t}$ are known, given the pre-determined state $n_{j,t-1}$. Note that the banking problem also depends on the household's stochastic discount factor Λ_t , which is pinned down by the sequence of household consumption, which is in turn an endogenous object. As we show below, it is not necessary to track Λ_t as a separate input sequence once $\{R_s, K_s\}_{s=0}^{\infty}$ are known.

As can be seen from Equation (12), household choices $\{b_{j,t}, C_t\}$ depend on the sequences $\{\{R_{j,s}^b\}_j, R_s, W_s, Div_s\}_{s=0}^{\infty}$. The sequence of wages is redundant and depends on capital, which is immediately known solely from $\{K_s\}_{s=0}^{\infty}$, given that labor supply is inelastic. Furthermore,

the bank dividend payout rule is a function of bank net worth only. In every period the following holds: $Div_t = (1 - \sigma)N_t$. Since $N_t = \int n_{j,t}dj$ and $n_{j,t}$ is pinned down by the bank's intertemporal law of motion, we can thus write $Div_t = \mathcal{D}_t(\{R_s, K_s\}_{s=0}^\infty)$.

Market Clearing and Linearization

Using the aggregate lending function from above, we impose the market clearing condition in the credit market:

$$\mathcal{L}_t(\{R_s, K_s\}_{s=0}^\infty) = K_t \quad (14)$$

This equation requires that credit demand always equates supply, i.e., there is no equilibrium rationing. Recall that our assumption is that firms produce capital goods with a technology that transforms one-to-one units of loans into units of capital. Next, consider bounded shocks dR and bounded perturbations in capital dK . Let us for a moment abstract from costly default to ease exposition. Assuming that \mathcal{L} is differentiable around the steady state yields a fixed point in lending:

$$\underbrace{(I - F)}_{\text{Indirect Effect}} \cdot dK = \underbrace{F^R}_{\text{Direct Effect}} \cdot dR \quad (15)$$

where I is an identity matrix, and F and F^R are Jacobians matrices with entries $F_{ts} = \frac{\partial \mathcal{L}_t}{\partial K_s}$, $F_{ts}^R = \frac{\partial \mathcal{L}_t}{\partial R_s}$. These matrices capture partial-equilibrium responses of bank lending at time t (corresponding to each row) to exogenous shocks to either capital or real rates at horizon s (corresponding to each column). Matrix F^R can be viewed to represent the "direct effect" of a policy intervention on lending, *holding the level of aggregate capital fixed*. Matrix F , on the other hand, captures indirect, general-equilibrium effects that happen through the endogenous adjustment in the quantity of aggregate capital.

To provide an intuition of the interplay between "direct" and "indirect" effects, consider for instance a contractionary monetary policy shock. For a given level of capital, the increase in the marginal cost of funds leads to a contraction in lending. In the general equilibrium, however, aggregate capital is reduced, thereby raising the marginal product of capital, and therefore encouraging lending. Thus a *negative* direct (partial equilibrium) effect interacts with a *positive* (general equilibrium) effect in determining the net response of aggregate lending. As argued above, once the sequence of aggregate lending, and therefore aggregate capital, is recovered, all other endogenous series can be readily derived.

Generalization with Costly Default

Let us turn to the baseline case with costly bank default. The market clearing condition

can be written in the following manner:

$$\mathcal{L}_t \left(\{R_s, K_s, S_s(R, K)\}_{s=0}^{\infty} \right) = K_t \quad (16)$$

where note the new term $\{S_s(R, K)\}_{s=0}^{\infty}$ which represents the input sequence of realized default costs. This sequence is an endogenous object and depends on the sequences of both real rates and aggregate capital. When linearizing, we must therefore take into account the propagating effects of real rate shocks on lending and capital accumulation through the default risk channel:

$$\underbrace{\left(\mathbf{I} - \mathbf{F} - \mathbf{F}^{d\mathcal{L}dS} \mathbf{F}^{dSdK} \right)}_{\text{Indirect Effect}} \cdot d\mathbf{K} = \underbrace{\left(\mathbf{F}^R + \mathbf{F}^{d\mathcal{L}dS} \mathbf{F}^{dSdR} \right)}_{\text{Direct Effect}} \cdot d\mathbf{R} \quad (17)$$

There are three differences relative to the basic case with no default in equation 15. First, there is an additional direct-effect term $\mathbf{F}^{d\mathcal{L}dS} \mathbf{F}^{dSdR}$ which represents the partial-equilibrium response of default costs to real rates shocks \mathbf{F}^{dSdR} , scaled by the response of lending to default cost shocks $\mathbf{F}^{d\mathcal{L}dS}$. Intuitively, real rate fluctuations impact the marginal cost of financial intermediaries and thus their distance to insolvency, which translates into the mass of realized defaults and the resource cost of banking crises S_t . In turn, changes in default costs - through their impact on the realized return on capital holdings - affect aggregate lending decisions of banks, as summarized in $\mathbf{F}^{d\mathcal{L}dS}$. Second, there is a new indirect-effect term $\mathbf{F}^{d\mathcal{L}dS} \mathbf{F}^{dSdK}$ which represents the general-equilibrium feedback loop from the response of default costs to changes in the capital stock (\mathbf{F}^{dSdK}), scaled again by the impact on aggregate lending $\mathbf{F}^{d\mathcal{L}dS}$. Finally, note that matrices \mathbf{F} and \mathbf{F}^R are newly re-computed conditional on the model with endogenous costly default.

Solving Equation 17 involves three simple steps. First, compute the partial-equilibrium truncated $T \times T$ matrices \mathbf{F} , \mathbf{F}^R , \mathbf{F}^{dSdR} , $\mathbf{F}^{d\mathcal{L}dS}$, and \mathbf{F}^{dSdK} . Second, feed the mean-reverting sequence of $d\mathbf{R}$ representing a monetary policy innovation. Finally, recover $d\mathbf{K}$. Once the sequence of $d\mathbf{K}$ is found, every other endogenous aggregate sequence can be calculated in one step. Equilibrium inflation π can be immediately computed since the New Keynesian block is a function of just the path of the aggregate capital stock, conditional on exogenous labor supply. The path of aggregate net worth follows the law of motion that, as has been shown, is a function of the input sequences. Given the path of assets and net worth, the balance sheet constraint pins down aggregate bank deposits. Finally, because capital depreciates every period, the goods market clearing condition $Y_t = C_t$ pins down consumption as a function of capital-determined output.

In what follows, to ease exposition we will collect matrices and work with compressed

formulas of the following form:

$$\underbrace{(I - F^*)}_{\text{Indirect Effect}} \cdot dK = \underbrace{F^{R*}}_{\text{Direct Effect}} \cdot dR \quad (18)$$

where $F^* \equiv F + F^{dLdS} F^{dSdK}$ and $F^{R*} \equiv F^R + F^{dLdS} F^{dSdR}$. In words, the star * superscript denotes Jacobians that take into account the endogenous bank default risk channel.

2.3 Model Extensions

We now discuss several extensions to the baseline HBANK model. We show that, with each extension, we either introduce a new Jacobian to Equation 18, or augment the existing Jacobians.

Macroprudential Policy

Our framework allows for a tractable analysis of macroeconomic effects of various policy interventions. We first discuss macroprudential regulation, specifically its non-systematic component. We consider bounded, mean-reverting shocks to the parameter that governs the leverage constraint $d\lambda$. A positive innovation represents a macroprudential policy tightening. Note that we still allow for the leverage constraint to bind occasionally. A linearized solution now contains a new Jacobian $F^{\lambda*}$ with entries representing partial-equilibrium responses to macropru shocks:

$$\underbrace{(I - F^*)}_{\text{Indirect Effect}} \cdot dK = \underbrace{F^{\lambda*}}_{\text{Macroprudential Policy}} \cdot d\lambda \quad (19)$$

where, as mentioned previously, a starred notation indicates that the default risk channel, i.e., the endogenous reaction of insolvency risk through changes in $d\lambda$, is taken into account in both F^* and $F^{\lambda*}$. In addition, notice that the iMPL matrix F^* is the same as computed before as the underlying micro-foundations of the model have not been changed.

Liquidity Policy

In practice, banks are required to hold a fraction of the stock of deposits in the form of reserves (Bianchi and Bigio, 2022). This buffer stock of cash capital is then used to withstand idiosyncratic fluctuations, such as the ones we allow for in our framework. We can parsimoniously introduce minimum reserves into HBANK in the following manner. Denote reserves as x_i and the minimum reserves ratio as ω . Thus the reserve requirement

reads as:

$$x_{j,t} \geq \omega b_{j,t} \quad (20)$$

The bank balance sheet is now: $b_{j,t} + n_{j,t} = q_{j,t}l_{j,t} + x_{j,t}$. The law of motion of net worth becomes: $n_{j,t+1} = R_{j,t+1}^T q_{j,t}l_{j,t} + R_{j,t+1}^x x_{j,t} - R_{j,t+1}^b b_{j,t} - \zeta_1 l_{j,t}^2$. The intertemporal version can be shown to be:

$$n_{j,t} = \mathbb{E}_t \sum_{s=1}^{\infty} \left[\left(E_{j,t+s} - X_{j,t+s} q_{j,t+s} l_{j,t+s} \right) \prod_{\ell=1}^s G_{j,t+\ell}^{-1} \right]$$

where $G_{j,t} \equiv \frac{R_{j,t}^b - R_{j,t+1}^x \omega}{1 - \omega}$ denotes the cum-reserves interest expense.

As our benchmark liquidity policy instrument we consider the interest rate on reserves R^x . Solving for general equilibrium with linearization we obtain:

$$\underbrace{(I - F_x^*)}_{\text{Indirect Effect}} \cdot dK = \underbrace{F_x^{R^{x*}}}_{\text{Interest on Reserves}} \cdot dR^x \quad (21)$$

where dR^x is a bounded shock to the reserves rate. Note that F_x^* now has an x subscript. Since the micro-foundations of the model have changed, i.e., financial intermediaries now face a new price and a new constraint on their behavior, we must re-compute the iMPL object conditional on the new assumptions. We also need to construct $F_x^{R^{x*}}$, which is a novel object and specific to the instrument of interest. Finally, once again, recall that all of this is conditional on the endogenous default risk channel since every object continues to have the star superscript.

Deposit Market Power

Our framework can be readily extended to the case of imperfect deposit market competition. To this end, we now assume that households derive utility from the holding of aggregate deposits in the spirit of the money-in-utility function approach (Sidrauski, 1967):

$$U(C_t, B_t) = \frac{1}{1 - \phi} C_t^{1 - \phi} + \nu_1 \frac{B_t^{1 - \nu_2}}{1 - \nu_2}$$

Furthermore, suppose that deposit franchises are imperfectly differentiated:

$$B_t = \left[\int_0^1 b_{j,t}^{\frac{\theta_b}{\theta_b + 1}} dj \right]^{\frac{\theta_b + 1}{\theta_b}}$$

where θ_b is the elasticity of substitution across deposit franchises. It can be shown that the deposit rate is now priced according to a Lerner-type equation that sets a *mark-down*

over the risk-free rate:

$$R_{j,t+1}^b = \left(1 - \frac{U_B(C_t, B_t)}{U_C(C_t, B_t)} \left(\frac{b_{j,t}}{B_t} \right)^{\frac{1}{\theta_b}} \right) R_{t+1}$$

where U_x represent marginal utilities. Hence the endogenous deposit mark-down depends on two arguments. First, a liquidity motive stemming from deposits appearing in the utility function, and captured by the marginal rate of substitution between deposits and consumption U_B/U_C . Second, the imperfect substitutability of banks' deposits, captured by the term $(b_{j,t}/B_t)^{\frac{1}{\theta_b}}$.

The deposit rate is now a function of other choice variables $\{C_t, B_t, b_{j,t}\}$ which, in turn, depend on the usual input sequences $\{R_s, K_s\}_{s=0}^{\infty}$. Thus, redefining the Jacobians from the basic model, and returning to conventional monetary policy shocks, we obtain a generalized version of the baseline formula 18:

$$(\mathbf{I} - \mathbf{F}_{DMP}^*) \cdot d\mathbf{K} = \mathbf{F}_{DMP}^{R*} \cdot d\mathbf{R} \quad (22)$$

where subscript DMP stands for deposit market power.

Credit Market Power

We can relax the assumption of perfect competition in the credit market as well. Suppose that aggregate capital K_t is now assembled according to the Kimball-type aggregator (Kimball, 1995):

$$\int_0^1 \Phi\left(\frac{k_{j,t}}{K_t}\right) dj = 1 \quad (23)$$

where $\Phi(x)$ is a strictly increasing and concave function. Firms solve the following problem:

$$\max_{k_{j,t}} \left[Q_t K_t - \int_0^1 q_{j,t} l_{j,t} dj \right]$$

subject to Equation 23 and the one-to-one conversion of bank loans and units of capital. The solution to the above problem yields the following inverse asset demand curve:

$$\frac{q_{j,t}}{Q_t} \mathcal{A}_t^k = \Phi' \left(\frac{k_{j,t}}{K_t} \right) \quad (24)$$

where $\mathcal{A}_t^k := \int_0^1 \Phi' \left(\frac{k_{j,t}}{K_t} \right) \frac{k_{j,t}}{K_t} dj$, and $Q_t = \int_0^1 q_{j,t} \frac{k_{j,t}}{K_t} dj$ is the aggregate capital price index. We adopt the Klenow and Willis (2016) parametric specification for Equation 24. The asset

market aggregator is thus:

$$\Phi\left(\frac{k}{K}\right) = 1 + (\theta_k - 1) \exp\left(\frac{1}{\epsilon_k}\right) \epsilon_k^{\frac{\theta_k}{\epsilon_k} - 1} \left[\Gamma\left(\frac{\theta_k}{\epsilon_k}, \frac{1}{\epsilon_k}\right) + \Gamma\left(\frac{\theta_k}{\epsilon_k}, \frac{\left(\frac{k}{K}\right)^{\epsilon_k/\theta_k}}{\epsilon_k}\right) \right] \quad (25)$$

where $\Gamma(.,.)$ is the incomplete Gamma function. Parameter θ_k helps control the average credit mark-up. Parameter ϵ_k helps determine the slope of the credit mark-up function. It can be easily shown that, under Kimball aggregation, the aggregate credit mark-up is a function of the distribution of relative bank sizes $\int \frac{l_{j,t}}{L_t}$, which are in turn functions of the same input sequences as before. Thus, the linearized solution of the model has the same shape as in the basic model except for the augmented Jacobians:

$$(I - F_{CMP}^*) \cdot dK = F_{CMP}^{R*} \cdot dR \quad (26)$$

where subscript CMP stands for credit market power.

3 Quantitative Analysis

We now turn to a quantitative assessment of the properties of the model.

3.1 Parameterization

We parameterize our model in several steps. All parameter values are listed in Table 1. First, we begin with standard macro parameters. We set the discount factor β to 0.996 to target a steady-state risk-free rate of roughly 1.6% p.a. The capital share α is set to 0.36 and the intertemporal elasticity of substitution is set to unity. Both are standard choices.

We proceed with the banking block. The dividend payout rule σ is set to 0.9, in line with [Gertler et al. \(2020\)](#) and yielding an expected payout of dividends every 2.5 years. The leverage constraint parameter λ is 0.12, which is the average capital requirement ratio across most advanced economies. The asset adjustment cost parameters ζ_1 and ζ_2 are set to 0.001 and 1.5, respectively, following [Jamilov and Monacelli \(2023\)](#). These values help to target an aggregate book leverage of roughly 6.5, which is close to its empirical counterpart of the average ratio of total loans over total equity across U.S. commercial banks.

We now move to the bank heterogeneity block. The persistence of the transitory component of idiosyncratic returns $\kappa_{j,t}$ is set to 0.5 following [Jamilov and Monacelli \(2023\)](#). In the data, when fitting a panel fixed effects model with AR(1) disturbances

Table 1: Model Parameterization

Parameter	Value	Description
Macro		
β	0.996	Discounting
α	0.36	Capital Share
ψ	1	Risk Aversion
Banking		
σ	0.9	Dividend Payout Rule
λ	0.12	Leverage Constraint
ζ_1	0.001	Asset Adjustment Linear
ζ_2	1.5	Asset Adjustment Quadratic
ω	0.025	Minimum Reserve Requirement
R^x	0.00	Steady-state Interest on Reserves
Bank Heterogeneity and Default Risk		
ρ_κ	0.5	Idiosync. Return, Persistence
σ_κ	0.5	Idiosync. Return, st. dev.
ν	5.33	Real Cost of Bank Insolvency
Bank Market Power		
θ_l	2.3	Elasticity of Substitution, Assets
ϵ_l	0.5	mark-up-Size Slope, Kimball
ν_1	3.6	Deposits in utility
ν_2	1	Elasticity of deposit supply
θ_b	2.1	Elasticity of Substitution, Deposits
New Keynesian Block		
γ	10	Elasticity of Substitution, Retail
φ	100	Price Adjustment Cost, Retail
\bar{R}	1.61	Real Rate target (p.a.)

Notes: This table summarizes the parameterization of the baseline model and all of its extensions.

to U.S. commercial bank-level data, one recovers an autoregressive coefficient of about 0.53 with the Durbin-Watson estimator. [Galaasen et al. \(2023\)](#) employ Norwegian bank-firm matched loan-level data and estimate that the persistence of uninsured idiosyncratic borrower-level shocks ranges from 0.1 to 0.32. In this paper we set ρ_κ to 0.5. The volatility of the transitory risk component σ_κ is set to 0.5.

The parameter that governs the real cost of bank insolvency - ν - is calibrated in order to match the empirically documented output losses during systemic banking crises. In particular, we set ν to 5.33, which helps to achieve a steady-state output loss from defaulted banks of roughly 10%, which is a median estimate found in the literature for a sample of developed economies ([Hoggarth et al., 2002](#)).

In the extension of the basic model that introduces liquidity policy we set the minimum reserve requirement ω to 2.5%, which corresponds to the average value across OECD economies, and the interest rate on reserves to zero. The latter is consistent with the recent reserves rate policy of both the Federal Reserve and the European Central Bank.

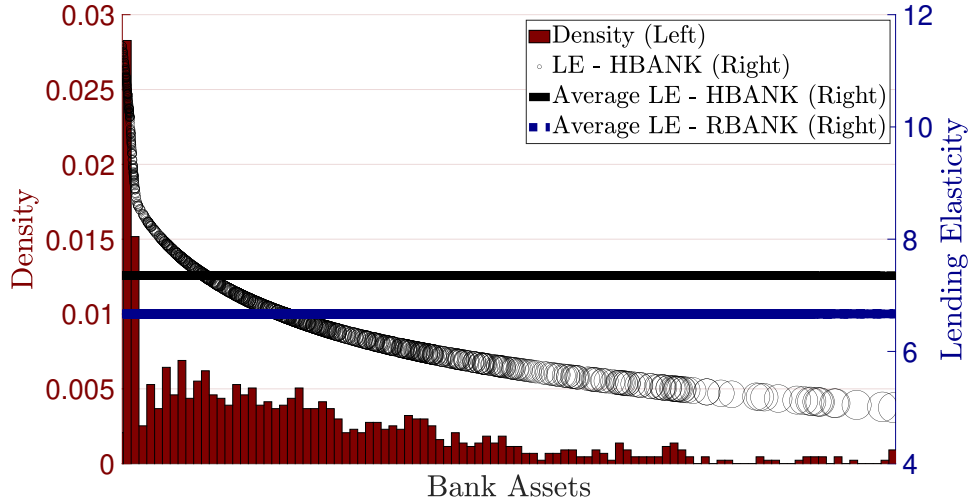
A key extension to the basic model involves imperfect credit and deposit market competition. There are several sets of model parameters that help determine the stationary distribution of credit mark-ups and deposit mark-downs. First, θ_k , which is set to 2.3, governs the implied homogeneous credit mark-up in the case of CES aggregation. Our calibration implies an average mark-up of 1.7, which is close to the empirically observed asset-weighted average credit mark-up (Corbae and D’Erasmus, 2021; Jamilov and Monacelli, 2023). We adopt the Klenow and Willis (2016) parametric specification for the credit market, and the corresponding slope parameter ϵ_l is set to 0.5 which helps to achieve a positive slope in the cross-sectional relationship between size and mark-ups. Second, the duple $\{\nu_1, \nu_2\}$ governs the deposits-in-utility term. We set the two parameters to 3.6 and 1, respectively. The former targets the average mark-down of roughly 0.82, which is in line with the existing empirical estimates in Jamilov and Monacelli (2023). Finally, the elasticity of substitution across deposit franchises θ_b is set to 2.1, which achieves an empirically-consistent mark-down elasticity of bank assets - a moment which is estimated in Jamilov and Monacelli (2023).

We complete the section with the New Keynesian block. The elasticity of substitution in the retail sector γ and the price adjustment cost are set to 10 and 100, respectively. These values are in line with the literature (Kaplan et al., 2018) and deliver a Phillips curve slope of 0.1, which is in the ballpark and slightly on the higher end of the recent micro empirical estimates in Hazell et al. (2021), and an average retail mark-up of 11%. The real rate target loosely corresponds to the average real interest rate in the U.S. over the last three decades.

3.2 Steady-State Properties

We begin the presentation of the results with the analysis of selected steady-state model properties. Figure 1 plots the steady-state distribution of assets in HBANK overlaid with the partial-equilibrium lending elasticity of bank net worth shocks $\frac{\partial k_{jt}}{\partial n_{jt}}$, which loosely corresponds to the slope of the lending function. Two aspects are worth emphasizing. First, the steady-state distribution of assets features pronounced right-skewness, in line with the data. Second, the lending elasticity is monotonically decreasing in net worth. Intuitively, banks with lower ex-ante net worth have steeper lending functions, are more constrained due to either non-interest expenses or the hard leverage cap, and are thus more responsive to sudden net worth windfalls. We also plot average lending elasticities

Figure 1: Lending Elasticities in Stationary Equilibrium



Notes: This figure plots the stationary distribution of bank assets k_j on the left axis, the lending elasticity $\frac{\partial k_{i,t}}{\partial n_{i,t}}$ in the cross section of HBANK on the right axis, and average lending elasticities in HBANK and RBANK on the right axis.

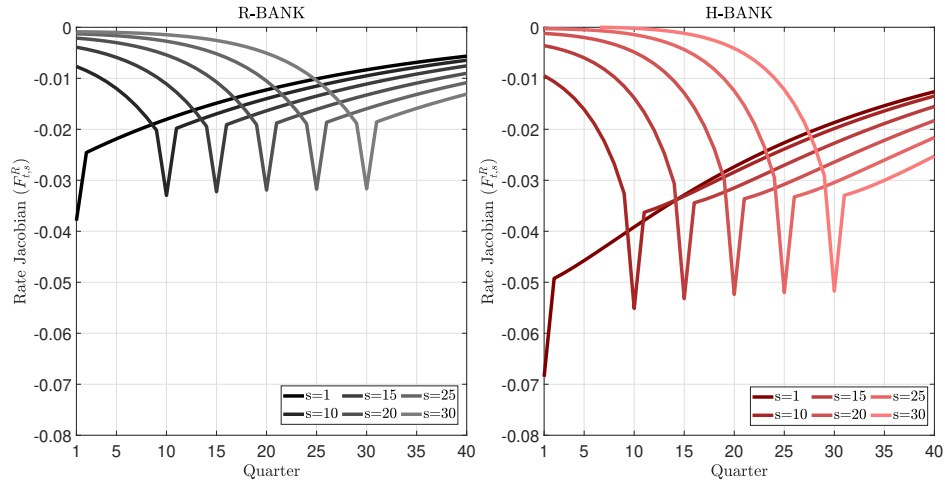
in both HBANK and RBANK. The former is 7.35% and the latter is 6.67%, suggesting that the *average* intermediary in a model with heterogeneity is more elastic than the *representative-agent* counterfactual. This important observation already foreshadows our main quantitative results on the amplifying effects of incomplete markets in the presence of costly default.

3.3 Policy Jacobians and Model iMPLs

As argued above, the model solution delivers two key sufficient statistics: the policy Jacobians and the intertemporal MPLs out of aggregate income. To start with, the F^R matrix determines the first-round direct effect $F^R \cdot dR$ of the monetary impulse. Each row in the matrix stands for a partial-equilibrium response of bank lending at time t to an unexpected shock to the real rate at time $s \geq 0$. We present the F^R matrix across two main model specifications: RBANK with a representative bank and complete markets, and HBANK with heterogeneous banks and incomplete markets. Throughout the rest of the paper we consider Jacobians that are truncated at quarter 40.

Figure 2 shows the columns of F^R for the two models considered. We display six response paths, respectively for real rate shocks at horizons 1, 10, 15, 20, 25, and 30. Notice first that the spikes are all negative, representing contractions in lending following an increase in the cost of funds. For smaller values of s , banks generally do not have

Figure 2: Real Interest Rate Jacobians (F^R)

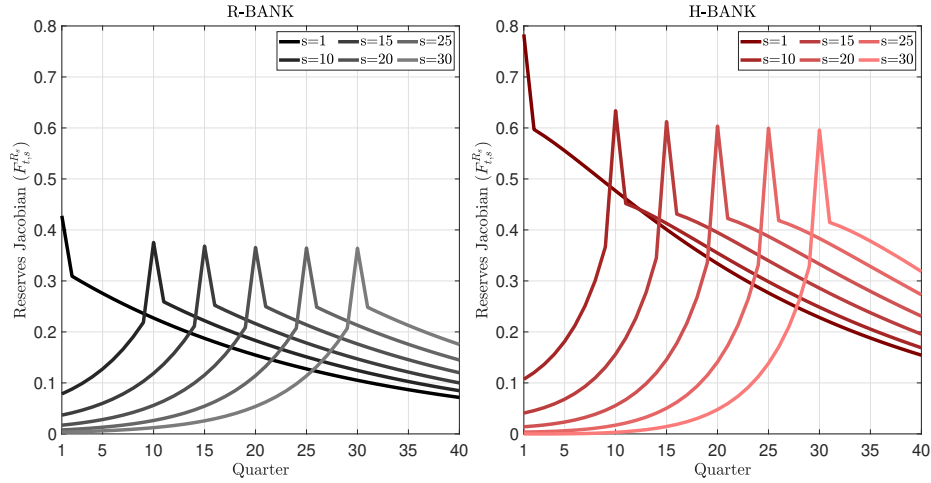


Notes: This figure plots select columns of the real interest rate Jacobian matrix F^R for RBANK (left panel) and HBANK (right panel). The horizontal axes represent rows of the matrices and each shaded line represents a separate column.

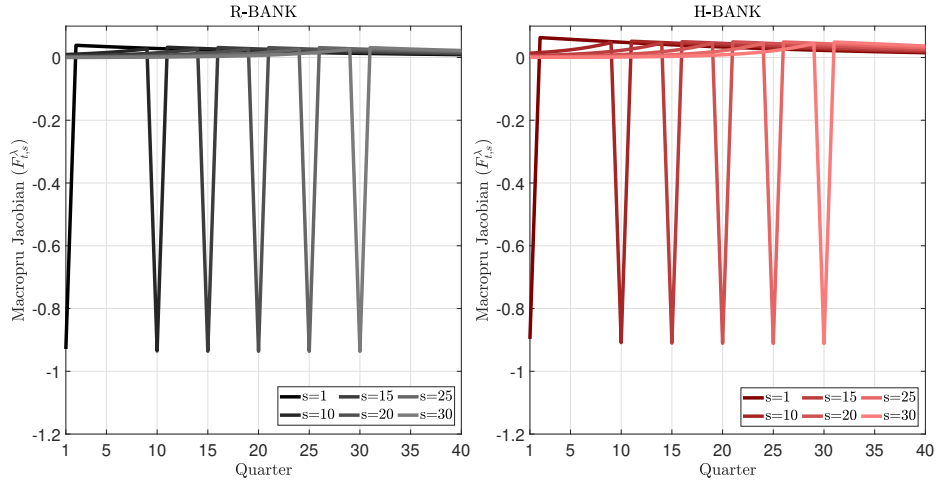
enough time to adjust their balance sheets preemptively, leading to sharper decreases. For higher values of s , the lending response pattern slowly stabilizes in the long run. Most importantly, by comparing the two models we see that the matrices appear very different - the lending responses in HBANK, particularly for lower columns, are significantly larger in absolute value. For instance, the direct effect at time $t=1$ of a real rate shock at horizon $s=1$ is twice as large in HBANK relative to RBANK. In other words, HBANK delivers substantial *amplification* of the direct effect of real interest rate shocks. As we will demonstrate in the next sections, this is due to the presence of a bank insolvency channel which magnifies the initial impulse via the endogenous response of insolvency risk to the policy shock.

Figure 3: Liquidity and Macroprudential Policy Jacobians

(a) Liquidity Policy (F^{R^x})



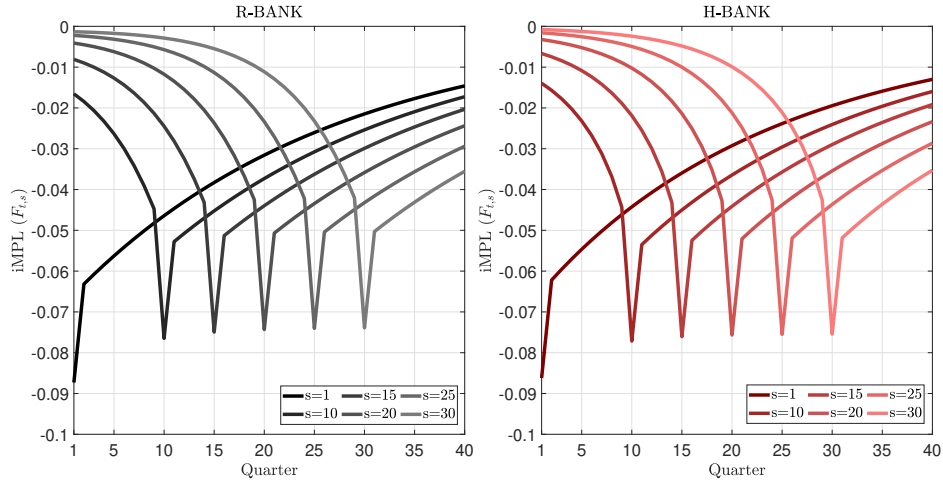
(b) Macroprudential Policy (F^λ)



Notes: In Panel (a), the figure plots select columns of liquidity policy Jacobian matrices F^{R^x} for RBANK (left panel) and HBANK (right panel). In Panel (b), the figure plots select columns of macroprudential policy Jacobian matrices F^λ for RBANK (left panel) and HBANK (right panel). The horizontal axes represent rows of the matrices and each shaded line represents a separate column.

Figure 3 presents columns from the liquidity and macroprudential Jacobian matrices F^{R^x} F^λ , which summarize the partial-equilibrium response of bank lending to shocks to the reserve requirement ω_s and to the capital requirement λ_s at different horizons, respectively. From Panel (a) we observe that, just like with interest rate shocks, HBANK exhibits a two-fold increase in the direct effect of liquidity policies. Panel (b), on the other hand, shows a different pattern: the lending responses to changes in capital requirements are very similar across RBANK and HBANK. This points to the fact that the direct effects

Figure 4: Intertemporal MPLs (F)



Notes: This figure plots select columns of the intertemporal MPL matrix F for RBANK (left panel) and HBANK (right panel). The horizontal axes represent rows of the matrices and each shaded line represents a separate column.

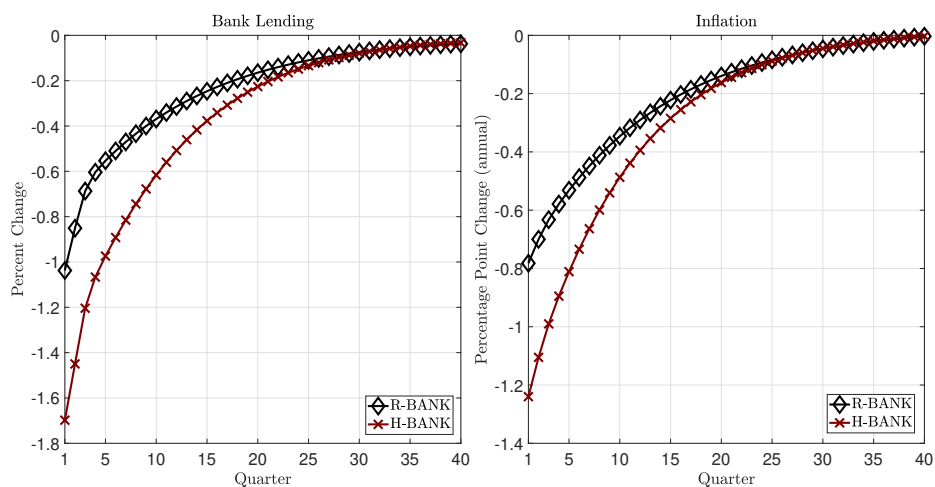
in HBANK depend on the *source* of policy fluctuations. We will return to this point during the discussion of impulse response functions and full general-equilibrium paths.

Figure 4 displays the columns of the iMPL matrix F , a sufficient statistic for the *indirect* effect of policy changes. All response signs are negative, suggesting that the general equilibrium effect is not an amplifier yet a *dampener* of shocks. This will become more apparent in Section 3.8. Note that responses are computed with respect to changes in aggregate capital, and thus reductions in the return on capital R^k . This explains the negative signs: in equilibrium, the fall in aggregate capital leads to an increase in the return on capital, and therefore a conditional increase in lending, which interacts with the direct-effect contraction in lending measured earlier. For small values of s the response is generally weaker (i.e., the general equilibrium effect is more negative) because there is no time for banks to respond to anticipated shocks to K_s . However, as s increases, a stable long-run pattern emerges across all models. This pattern is remarkably similar across the two model economies, suggesting that the general equilibrium feedback is not much affected by bank heterogeneity.

3.4 Monetary Policy Transmission in HBANK

Equipped with the sufficient statistics objects F and F^R , we now compute the impulse response of aggregate lending to a transitory, 1% (p.a.) shock to the real interest rate that mean-reverts at rate 0.5. Recall that obtaining the path of capital dK - which equates

Figure 5: Impulse Response to a Monetary Shock



Notes: Impulse response functions for bank lending dK (left panel) and inflation $d\pi$ (right panel) to an unexpected, transitory 1% (p.a.) real interest rate shock dR .

the path of lending - is sufficient to recover every other endogenous aggregate variable, including inflation.

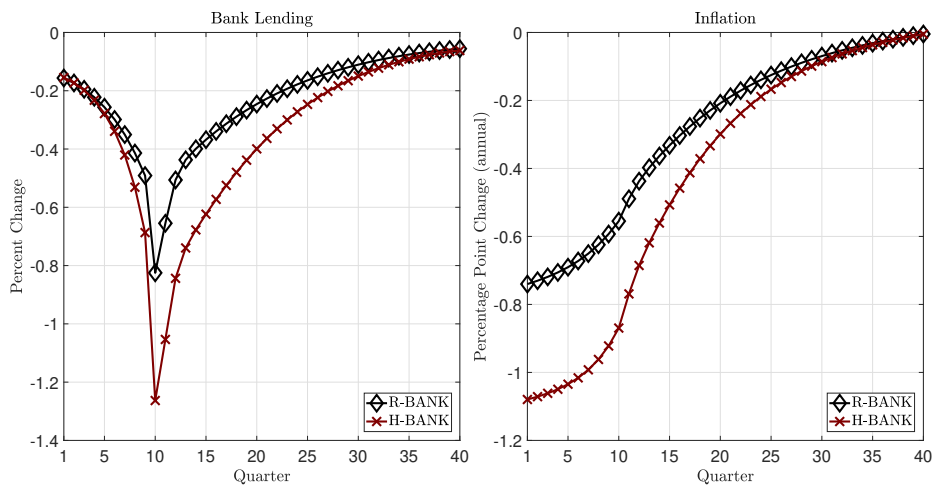
Figure 5 displays a key result of the paper. Aggregate lending quantities (left) and inflation (right) fall by around 70% and 50% more in HBANK than in RBANK, respectively. The macroeconomic response to non-systematic monetary policy shocks is substantially amplified by the presence of bank heterogeneity, incomplete markets, and costly default. Section 3.7 inspects the mechanism behind this result more closely but we already provide a brief explanation here. In RBANK, because of market completeness and perfect insurance, the probability of bank insolvency due to negative idiosyncratic rate of return draws is exactly zero. In HBANK, on the other hand, the presence of incomplete markets generates a non-trivial mass of small intermediaries with a high likelihood of insolvency. This likelihood *rises* in response to a contractionary monetary policy shock because of a higher marginal cost of funds, with banks becoming less profitable, especially if small in size. A higher insolvency risk, in turn, lowers aggregate profitability in the economy even further and causes bank lending to fall by more. A larger decline in bank lending leads to a sharper decline in real activity and, through the New Keynesian aggregate supply relation, to a sharper contraction in final-good inflation.

3.5 Forward Guidance

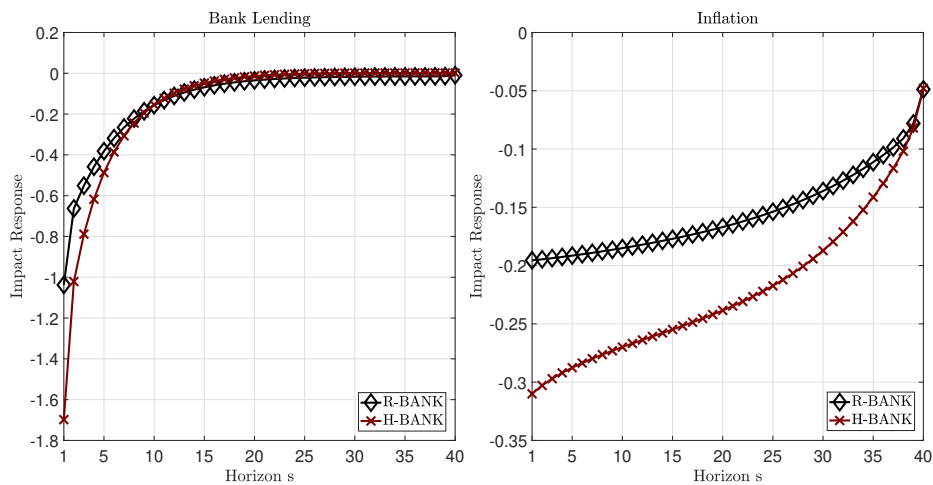
Next, we turn to the analysis of the transmission of forward guidance policy shocks. Forward guidance amounts essentially to an interest rate news shock. The sequence-space approach is ideal for the study of such news shocks because the sufficient-statistic matrices F and F^R are all that is needed to compute general-equilibrium responses to credible interest rate impulses at any horizon $s \geq 0$.

Figure 6: Forward Guidance in HBANK

(a) Impulse Response to a Real Interest Rate Shock at $s = 10$



(b) Impact Responses to Real Interest Rate Shocks at Different Horizons



Notes: Panel (a) of the figure plots impulse response functions for bank lending dK (left panel) and inflation $d\pi$ (right panel) to unexpected, transitory 1% (p.a.) real interest rate shocks dR at a future horizon $s = 10$ (i.e. 10 quarters ahead). Panel (b) of the figure plots impact responses of bank lending dK (left panel) and inflation $d\pi$ (right panel) to unexpected, transitory 1% (p.a.) real interest rate shocks dR at horizons $s = \{1, 2, \dots, 40\}$, which are shown on the horizontal axes.

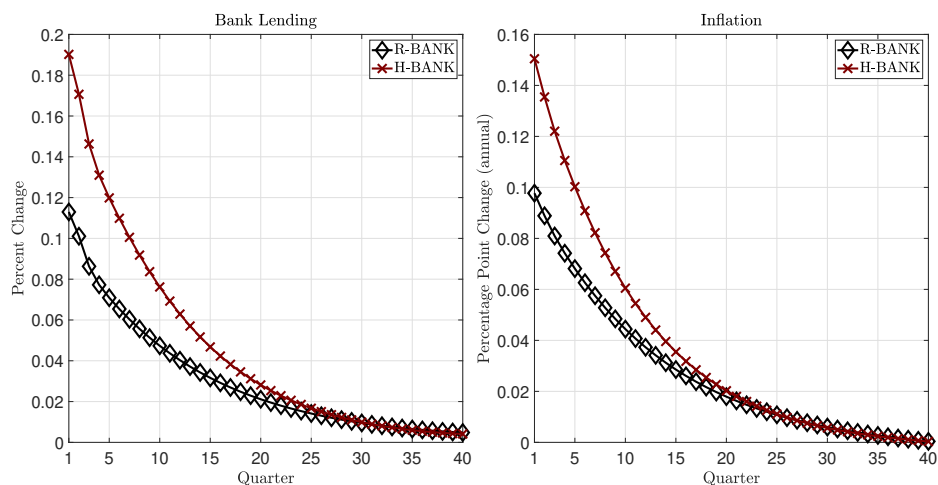
Panel (a) of Figure 6 plots the responses of bank lending and inflation to an interest rate hike that will occur at horizon $s = 10$ but is announced at $t = 1$. We observe that the sharpest spike in both variables is at the implementation date. However, there is a contraction in bank lending even at the time of the announcement, i.e., in the first quarter. The intuition is simple. From equation 11, financial intermediaries' net worth depends on the present discounted value of the excess return on assets, net of non-interest expenses. That value falls with a rise in the expected future safe real interest rate, leading to a current contraction in net worth. Hence banks anticipate a higher cost of funds in the future and begin to de-lever and reduce lending immediately. Notice that this channel is stronger in HBANK because, for a reason similar to the case of conventional monetary policy, bank heterogeneity increases the average elasticity to interest rate shocks at *any* horizon.

An immediate question that follows from the above analysis is whether a forward guidance puzzle holds in HBANK (Del Negro et al., 2023). The puzzle arises whenever the general-equilibrium responsiveness to interest rate shocks at horizon s *increases* with the horizon. Such model behavior is deemed to be unrealistic.

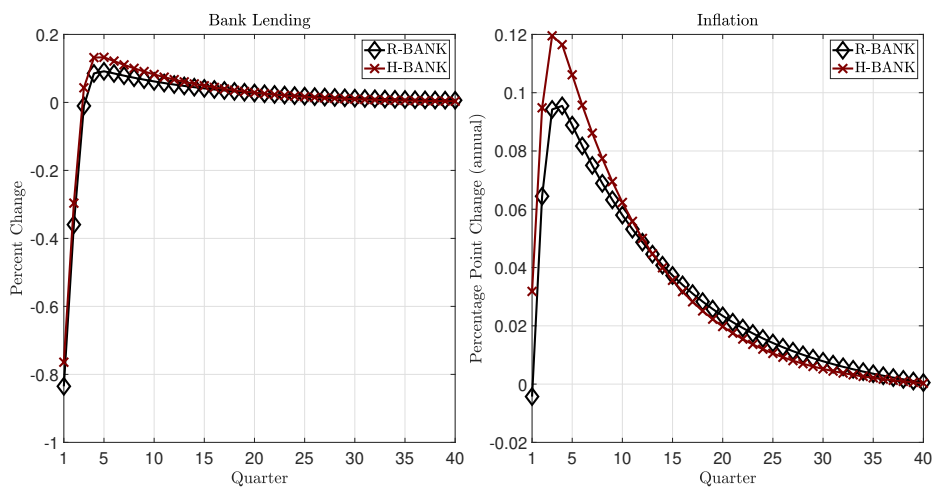
We can test for the presence of a forward guidance puzzle in our framework directly by computing the *impact* response of lending and inflation to (news) shocks to the real interest rate at different *horizons*. Panel (b) of Figure 6 shows the result from this exercise. We note that - for both HBANK and RBANK and for both quantities and prices - the responsiveness *decreases* with the horizon (in absolute value). That is, our model economy does *not* suffer from a forward guidance puzzle. The reason for this is that financial intermediaries - unlike households - do not exhibit a consumption smoothing incentive. In our framework households do not face idiosyncratic risk. They are standard Euler consumers who want to smooth their consumption intertemporally and react to announcements of distant interest rate shocks immediately. Banks, on the other hand, face idiosyncratic return risk in a context of incomplete insurance, and therefore are a lot less likely to substitute intertemporally. This point is argued extensively in Hagedorn et al. (2019) who highlight that one theoretical key for resolving the forward guidance puzzle is precisely to mitigate the intertemporal substitution channel.

Figure 7: Impulse Response to Liquidity and Macroprudential Policy Shocks

(a) Liquidity Policy



(b) Macroprudential Policy



Notes: Panel (a) of the figure plots impulse response functions for bank lending dK (left panel) and inflation $d\pi$ (right panel) to unexpected, transitory 1% (p.a.) shocks to the interest rate on reserves dR^x . Panel (b) of the figure plots impulse response functions for bank lending dK (left panel) and inflation $d\pi$ (right panel) to unexpected, transitory 1% shocks to capital requirements $d\lambda$.

3.6 Liquidity and Macroprudential Policies

We now proceed with the discussion of the effects of alternative economic policies: liquidity and macroprudential. The former is simulated with a transitory, mean-reverting one percentage point (annualized) increase in the interest rate on reserves R^x . The latter is a one-percent, mean-reverting increase in capital requirements λ .

Figure 7 presents the results. Panel (a) shows the outcome of the liquidity policy. Note that an increase in R^x is expansionary since it lowers the cum-reserves interest expense

$\frac{R_{j,t}^b - R_{t+1}^x \omega}{1-\omega}$ for a given minimum reserve requirement $\omega > 0$ and a given deposit interest rate $R_{j,t+1}^b$. Notice that the effect is stronger in HBANK than in RBANK by a factor of two. The intuition is similar to before and rests on the heterogeneity in the lending elasticity across the distribution: the incomplete markets assumption increases the elasticity to all components of interest expenses.

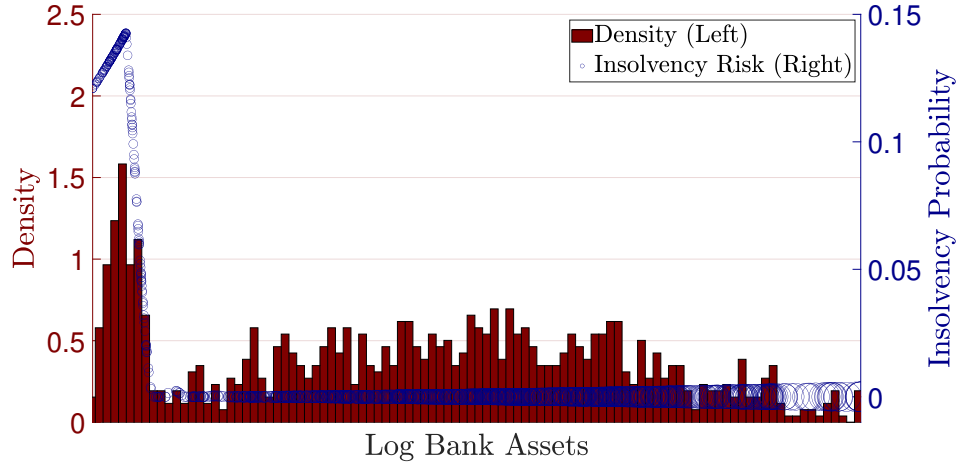
Panel (b) of Figure 7 presents the macroeconomic effects of a macroprudential shock. We observe a decline in bank lending across both models. Inflation, on the other hand, increases along the transition path. This pattern is consistent with the empirical literature on the bank-level effects of capital requirements (Juelsrud and Wold, 2020). Using a 2013 Norwegian policy reform that required banks to carry more capital, Juelsrud and Wold (2020) show that banks responded by cutting loans to non-financial firms and by raising interest rates. The combined negative effect on lending quantities and positive effect on prices identifies a capital requirements shock as a negative “supply-side” disturbance. Our model’s impulse response dynamic in response to tighter macroprudential policy is therefore consistent with the micro evidence. Finally, we observe that the effects are somewhat stronger in HBANK, albeit mildly. This is in part due to the calibration of the steady-state value of λ . The hard cap on leverage binds on a small minority of very low-net worth intermediaries and thus market incompleteness does not add much internal propagation along the leverage constraint dimension.

3.7 Inspecting the Mechanism: the Endogenous Default Channel

Having discussed the macroeconomic implications of interest rate, forward guidance, macroprudential, and liquidity policies we now inspect the mechanism behind the model’s performance. In particular, we highlight a key theoretical channel of the HBANK framework - costly intermediary default.

We begin by plotting the distribution of insolvency probability $\varphi_{j,t}$ in the stationary equilibrium of the economy. Figure 8 shows that $\varphi_{j,t}$ is large in the left tail of the distribution of size, reaching magnitudes of above 10%. For the rest of the banking sector, $\varphi_{j,t}$ is significantly lower. A higher average ex-ante probability of insolvency, everything else equal, immediately implies that the equilibrium mass of insolvent banks s_t is positive, which in turn means that the realized cost of default $S_t = \nu s_t Y_t$ is positive for some $\nu > 0$. Thus, incomplete markets and bank scale-variance deliver an endogenous insolvency risk channel that is concentrated among the low-net worth banks. As the distribution of net worth responds to aggregate shocks, the mass of banks in the left tail increases endogenously due to, e.g., a higher cost of funds. In the second round, the rising likelihood and

Figure 8: Insolvency Risk in the Cross Section



Notes: This figure plots the stationary distribution of bank assets k_j on the left axis and the ex-ante probability of bank insolvency φ_j in the cross section of HBANK on the right axis.

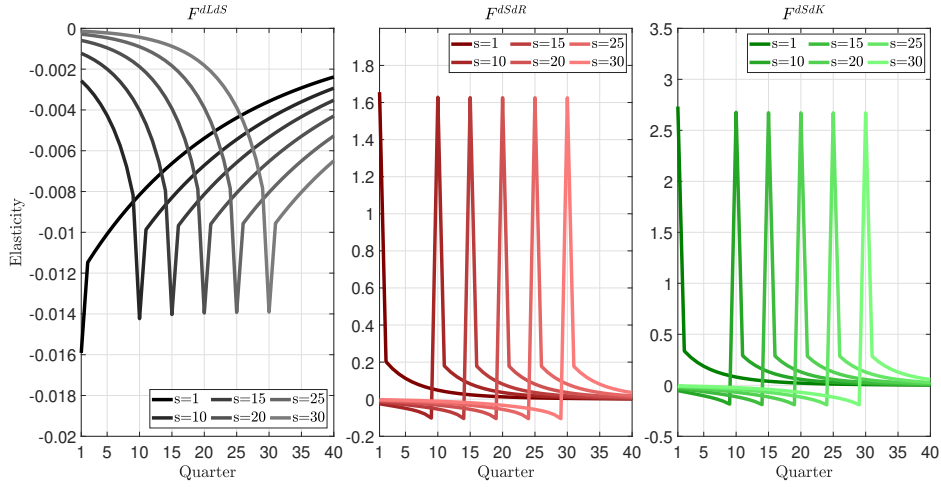
incidence of bank insolvency raises the resource costs of default, lowering the aggregate return on capital R_t^k , and feeding back into the banking sector via a lower loan supply to firms, causing an economic contraction and deflation.

To further shed light on the dynamic behavior of the model, we show the columns of matrices $\mathbf{F}^{d\mathcal{L}dS}$, \mathbf{F}^{dSdK} , and \mathbf{F}^{dSdR} . Recall that $\mathbf{F}^{d\mathcal{L}dS}$ represents the partial-equilibrium elasticity of bank lending to unanticipated shocks to the realized default cost S_t . \mathbf{F}^{dSdK} summarizes the partial-equilibrium elasticity of realized default costs to shocks to aggregate capital K_t . Finally, \mathbf{F}^{dSdR} collects partial-equilibrium elasticities of realized default costs to exogenous shocks to the real interest rate.

Figure 9 plots select columns from the three Jacobians. First, we highlight that the columns of $\mathbf{F}^{d\mathcal{L}dS}$ are negative for every s suggesting that banks lend less whenever the realized cost of default is high and the realized return on capital is low. Second, we see that all columns of \mathbf{F}^{dSdR} are positive, implying that the cost of default *increases* following contractionary monetary policy shocks. Recall that the product of $\mathbf{F}^{d\mathcal{L}dS}$ and \mathbf{F}^{dSdR} constitutes the direct effect of real rate impulses. As we will see in Section 3.8, direct effects account for a dominant share of the total response to aggregate shocks in HBANK. Thus, the sign of the \mathbf{F}^{dSdR} Jacobian is an important testable prediction of our model and a key channel of the mechanism. We will validate this channel explicitly with micro-data on U.S. banks in Section 4.

Finally, columns of \mathbf{F}^{dSdK} are positive which suggests that insolvency risk and costs *rise* following positive changes in aggregate capital. Due to diminishing returns to capital, a

Figure 9: Decomposing the Insolvency Risk Channel



Notes: The left, middle, and right panels plot select columns of the F^{dLdS} , F^{dSdR} , and F^{dSdK} Jacobian matrices for HBANK, respectively. The horizontal axes represent rows of the matrices and each shaded line represents a separate column.

higher K_t induces lower aggregate returns for a given cost of default, and thus incentivizes banks to provide fewer loans to firms. Recall that F^{dSdK} belongs to the indirect effect channel of macroeconomic transmission in HBANK. Thus, because of the positive sign of F^{dSdK} we see that insolvency risk offers a *dampening* effect in general equilibrium in addition to the *amplifying* effects in partial equilibrium as discussed above. Generally, the net impact of the two conflicting forces depends on the relative power of direct and indirect effects which, as Section 3.8 shows explicitly, are heavily tilted towards direct effects. Thus, it is generally the case that the direct effect of endogenous insolvency risk dominates the indirect effect and the aggregate net impact results in amplification of monetary shocks.

There is therefore a *trade-off* between macroeconomic and financial stabilization. Recall from Section 3.4 that HBANK delivers sizeable internal propagation of real interest rate shocks. A monetary authority that wishes to contract and deflate the economy, for instance in response to an expansion in aggregate demand, is successful at achieving the objective but this comes at the cost of raising financial vulnerability in the banking sector as seen directly from the positive signs of the columns in F^{dSdR} . This trade-off is not driven by the calibration or by particular modelling assumptions. It is driven by a simple combination of two well-understood primitives: incomplete markets and costly default.

3.8 Direct and Indirect Effects Decomposition

This section provides an exact quantitative decomposition of the aggregate effects of monetary policy shocks into direct and indirect channels. Our approach follows closely the work of [Kaplan et al. \(2018\)](#), [Auclert \(2019\)](#), and [Auclert et al. \(2023\)](#), among others. Recall that the direct effect in HBANK is a combination of F^R , which is a partial-equilibrium elasticity of bank lending to real rate shocks, and the product of F^{dLdS} and F^{dSdR} , which collectively summarize the endogenous insolvency risk channel. The indirect effect, on the other hand, is comprised of F , which is the iMPL matrix that collects bank lending responses to aggregate capital shocks, and the product of F^{dLdS} and F^{dSdK} , that reflects the general-equilibrium implications of insolvency risk.

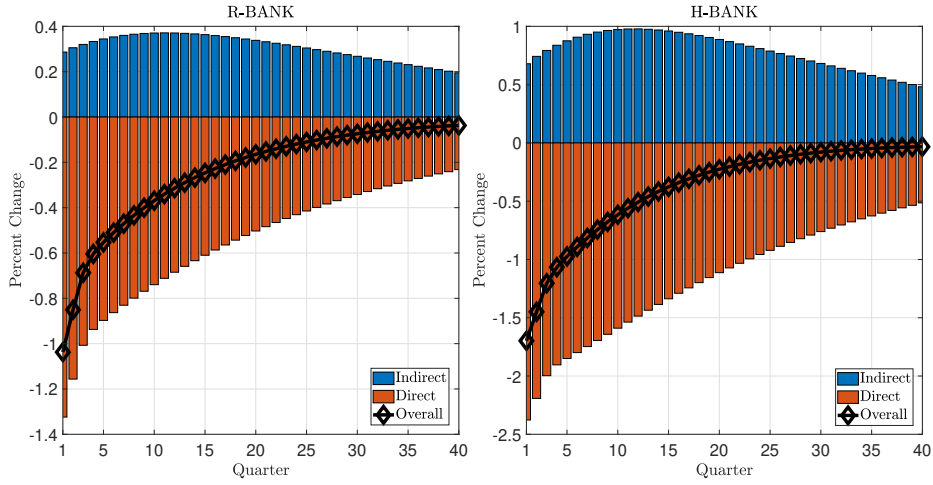
Figure 10 presents the decomposition into overall direct and indirect effect of the total response of bank lending to the same mean-reverting real interest rate shock dR . First, on average, the direct effect explains roughly 70% of the total macroeconomic response to monetary shocks, particularly on impact and at lower horizons. This is true for both RBANK and HBANK, which suggests that the feature is general and not driven by the assumption on market incompleteness. Second, moving from RBANK to HBANK raises the direct effect by almost a full percentage point. This is exclusively due to the presence in HBANK of the endogenous insolvency risk channel. In Section 3.9 below we will reinforce this statement even further. Third and finally, as already implied previously, general-equilibrium effects are *positive*, i.e., they provide a dampening impact on contractionary monetary policy shocks. This is intuitive, as the indirect effect is described by the (positive) equilibrium effect of capital on lending at different horizons. Incomplete markets and insolvency risk roughly double the magnitude of the indirect effect, as can be seen from the comparison of the blue bars in RBANK in HBANK. However, because the indirect effect is still quantitatively mild, it is dwarfed by the growth of the direct effect, resulting in the *overall* contraction of bank lending.

3.9 Illustrating the "As-If" Result

This section complements the previous discussion on the decomposition of aggregate responses to monetary policy shocks into direct and indirect effects. In particular, we study the behavior of the HBANK model *without* insolvency risk. Specifically, we now set the cost of default parameter ν to zero.

Figure 11 presents the result of this exercise. HBANK- ex default risk behaves quantitatively almost identically to the model with a representative intermediary. This finding is consistent with the results in [Gali and Debortoli \(2022\)](#) who argue that a plain-vanilla

Figure 10: Decomposition of the Total Response to Real Interest Rate Shocks



Notes: This figure decomposes the total response of bank lending dK in RBANK (left panel) and HBANK (right panel) to an unexpected, transitory 1% (p.a.) shock to the real interest rate dR into direct and indirect effects.

model with incomplete markets and nominal rigidities yields practically the same impulse responses as the same model with complete markets. In the literature, this is known as the “as-if” result, which holds whenever, generally speaking, a heterogeneous-agent economy has the same aggregate elasticity of exogenous shocks as its representative-agent counterfactual benchmark (Krusell and Smith, 1998; Werning, 2015).

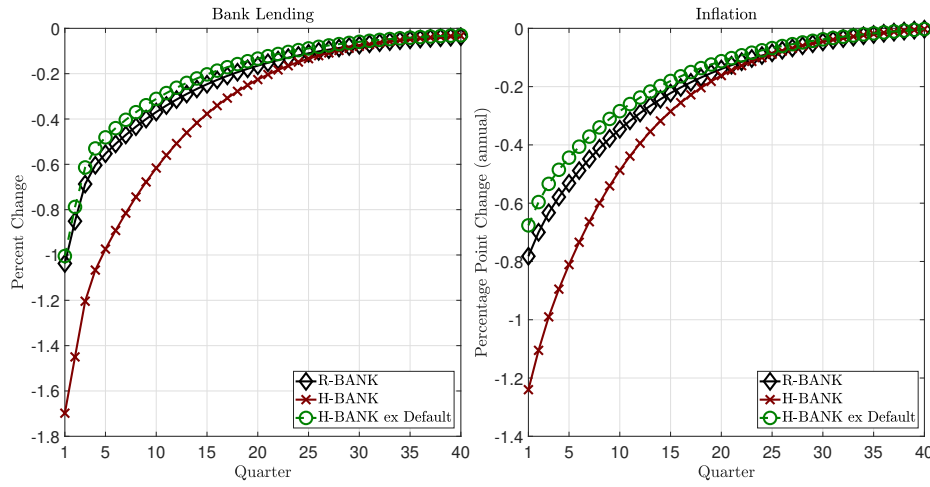
By comparing impulse responses from HBANK and HBANK-ex default we are thus able to directly isolate and identify the general equilibrium role of insolvency risk in the presence of incomplete markets. We conclude that our headline amplification result is entirely due to the insolvency risk channel.

3.10 Monetary Policy with Deposit and Credit Market Power

In this section we study two major extensions of our basic framework: imperfect banking competition in the deposit and credit markets. Figure 12 presents impulse response functions to real interest rate shocks based on three model economies: (i) standard RBANK, (ii) standard HBANK, and (iii) HBANK with either deposit or credit market power.

Panel (a) of Figure 12 shows that deposit market power significantly dampens the macroeconomic effects of monetary shocks. Quantitatively, the impact responses of bank lending and inflation are almost halved. The intuition for this result is straightforward: the *pass-through* from the risk-free real interest rate R_{t+1} to retail deposit rates $R_{j,t+1}^b$ is

Figure 11: HBANK with and without Insolvency Risk

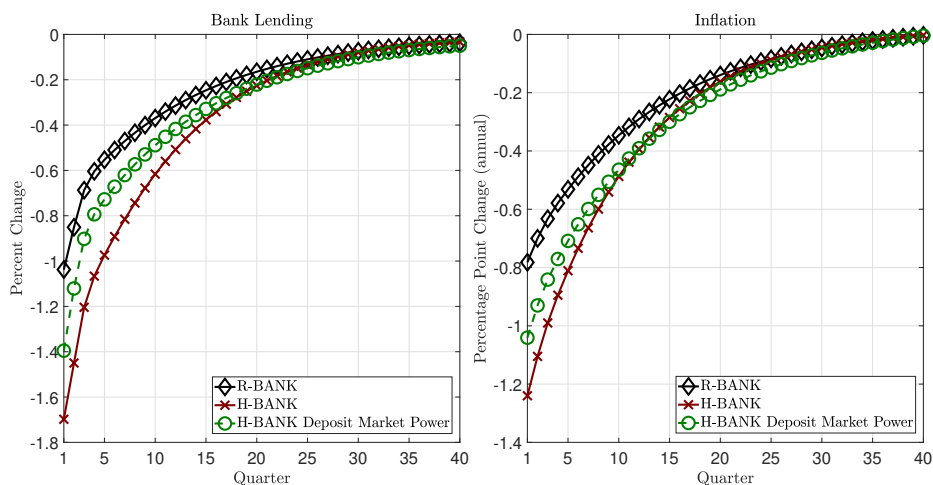


Notes: This figure plots impulse response functions for bank lending dK (left panel) and inflation $d\pi$ (right panel) to unexpected, transitory 1% (p.a.) shocks to the real interest rate dR . Both panels include IRFs for the HBANK economy without costly intermediary default (i.e. ν is set to zero).

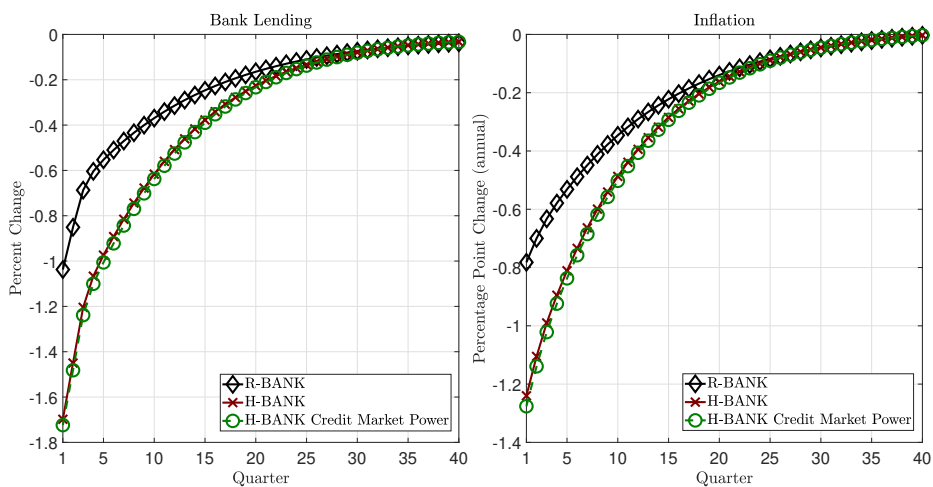
no longer perfect and is governed by the deposit *mark-down* $\left(1 - \frac{U_B(C_t, B_t)}{U_C(C_t, B_t)} \left(\frac{b_{j,t}}{B_t}\right)^{\frac{1}{\theta_b}}\right)$. The deposit rate rises less than proportionally relative to the risk-free rate. This is consistent with households substituting away from deposits to mutual funds assets (whose risk-free return is now relatively higher), thereby raising the marginal utility of deposits relative to consumption, and lowering the markdown. Fluctuations in the risk-free rate are thus passed over to real deposit rates imperfectly, implying that for the same monetary policy contraction the bank marginal cost rises less than proportionally. Imperfect deposit-rate pass-through, driven by the sticky deposit franchise, allows banks to *cut loans by less* than in the perfect-competition benchmark. Bank lending and real activity therefore contract by less than otherwise.

Figure 12: Monetary Policy with Imperfect Banking Competition

(a) Deposit Market Power



(b) Credit Market Power



Notes: Both panels plot impulse response functions for bank lending dK (left panel) and inflation $d\pi$ (right panel) to unexpected, transitory 1% (p.a.) shocks to the real interest rate dR . Panels (a) and (b) include IRFs for the baseline HBANK model that has been extended with deposit and credit market power, respectively.

4 Empirical Evidence on Bank Heterogeneity and Insolvency Risk

In this section we complement our theoretical analysis by providing empirical evidence on the effects of identified monetary policy shocks on the cross section of banks. Our main data source is the Federal Reserve Consolidated Reports of Condition and Income (also known as Call Reports). This dataset includes both income statement and balance sheet

variables for the universe of U.S. FDIC-insured banks at quarterly frequency. Our sample covers the period 1990q1-2019q4. To capture monetary policy surprises, we follow the high-frequency identification approach. Specifically, following [Gurkaynak et al. \(2005\)](#) and [Gertler and Karadi \(2015\)](#) we use the change in the 3-month ahead Fed Funds futures within a 30 minute window around FOMC announcements as our baseline instrument for monetary shocks. Throughout the rest of 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.

4.1 Heterogeneous Lending Responses to Monetary Policy

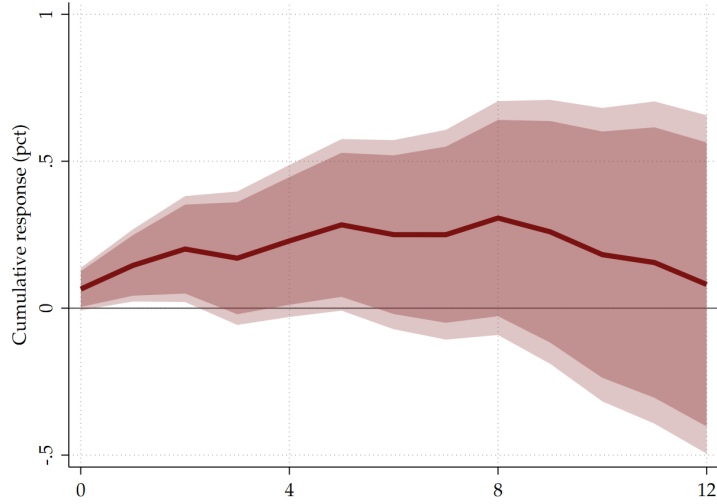
We begin by investigating the cross-sectional effects of U.S. monetary policy shocks on the *size* of U.S. commercial banks. To do so, we run the following lag-augmented panel local projection ([Jordà, 2005](#); [Montiel Olea and Plagborg-Møller, 2021](#)) as in [Bellifemine et al. \(2023\)](#):

$$\Delta \ln(Y_{it+h}) = \underbrace{\alpha_{ih} + \delta_{th}}_{\text{Fixed effects}} + \underbrace{\beta_h \times D_{it} \times \varepsilon_t}_{\text{Size interaction}} + \underbrace{\psi_h D_{it}}_{\text{Interaction controls}} + \underbrace{\sum_{\ell=1}^4 \gamma_{h\ell} \Delta \ln(Y_{it-\ell})}_{\text{Lagged controls}} + u_{iht} \quad (27)$$

where $\Delta \ln(Y_{i,t+h}) = \ln(Y_{i,t+h}) - \ln(Y_{i,t-1})$ represents the h -quarters ahead cumulative change in real *total assets* for bank i , α_{ih} is a bank fixed effect, while δ_{th} denotes a time fixed effect. $\Delta \ln(Y_{it-\ell}) = \ln(Y_{i,t-1}) - \ln(Y_{i,t-\ell-1})$ denotes past bank-level asset growth, while ε_t is the monetary surprise. Finally, D_{it} is a dummy variable which is equal to one only for those banks that were in the top 20% of the bank size distribution in the quarter preceding the shock. We use two-way clustered standard errors at the bank and quarter level.

Figure 13 plots the estimated $\hat{\beta}_h$ coefficient. Notice that, while the time fixed effect δ_{th} absorbs the average effect of the monetary shock, what we are interested in is the *differential* response to the shock across banks. This differential response is exactly what the coefficient β_h is capturing. In particular, β_h can be interpreted as the differential real asset response of banks in the top 20% of the asset distribution compared to the baseline group, which is represented by those banks which are in the bottom 80% of the asset distribution. As already emphasized in [Kashyap and Stein \(1995, 2000\)](#), Figure 13 shows that large banks tend to contract real assets by less following a contractionary monetary policy shock, compared to small ones. More specifically, following a 1 standard deviation contractionary monetary policy shock, we find that banks in the top quintile of the assets distribution experience a contraction in the size of their balance sheet which is up to 0.3%

Figure 13: Heterogeneous size response in the cross-section of banks



Notes: estimates of β_n from (27) to a 1 standard deviation contractionary monetary shock. Errors are two-way clustered at the time and bank level. Lightly shaded areas represent 95% confidence intervals. Darkly shaded areas are 90% confidence intervals. The y-axis represents the cumulative percentage change in real assets. The x-axis represents quarters elapsed since the shock.

smaller compared to banks in the bottom 80% of the size distribution.

This differential elasticity in bank-level responses to monetary policy shocks is one of the key factors that motivates our heterogeneous-bank modeling approach. This is closely related to Figure 1 and the discussion of heterogeneous lending elasticities in the stationary equilibrium of HBANK. Large banks are less sensitive to net-worth fluctuations and to the extent that monetary policy affects bank net worth they are therefore also less responsive to monetary policy surprises.

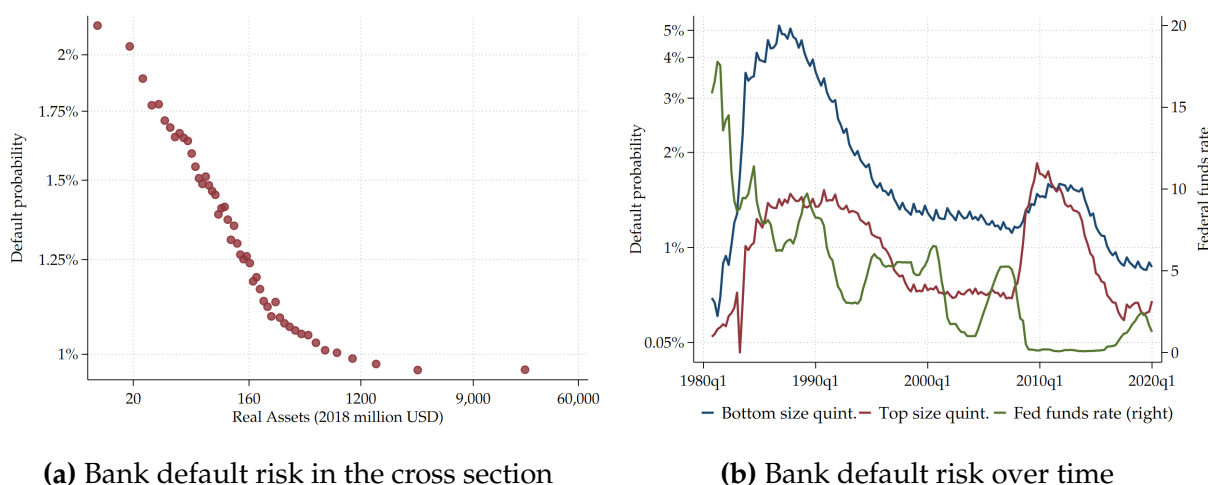
4.2 Bank Default Risk: Cross Section and Time Series

We now turn to bank default risk, which is a key channel of policy transmission in our model. Following Laeven and Levine (2009), we use the *z-score* as our main measure for default risk. 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:

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

where RoA is defined as net income over total assets, leverage is defined as book assets over book equity, and $\text{SD}(\text{RoA})$ is a moving average of the bank-level standard deviation

Figure 14: Banks' default risk over time and in the cross section



of RoA over some time window.¹ Under the assumption of normally distributed profits, the z-score equals the inverse probability of insolvency of a given bank. Thus, in the rest of our analysis we consider the inverse z-score, which is proportional to a bank's default risk. In our empirical setting the z-score is a convenient proxy for default risk as it is constructed from income statement and balance sheet items only and is thus readily available for all banks and time periods in our sample. Market-based measures of default risk, on the other hand, are only available for publicly traded banks, which constitute a small minority of the universe of U.S. commercial banks.²

Panel (a) of Figure 14 shows a binned scatter plot of bank size - defined as total book assets - against default probability as proxied by the inverse z-score. We first split our sample into 50 equally-sized bins based on real assets, each including roughly 18,000 observations. We then residualize both axes from a time fixed effect. Finally, for each bin we display average assets against the average default probability. There is a robust negative cross-sectional relationship between bank size and default risk, a pattern that also maps directly into our model. In particular, HBANK strongly predicts that default risk is decreasing with bank size, as was demonstrated in Figure 8. Thus, we have verified a key testable prediction of our model with U.S. micro data.

Panel (b) of Figure 14 plots the evolution of bank default risk across time. It displays the time-series behavior of average default risk for banks in the bottom and top quintiles of

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

²We have checked and confirmed that our z-score measure is very highly correlated to market-based measures of default risk as in Nagel and Purmanandam (2019).

assets, together with the federal funds rate. Default risk tends to be lower for larger banks also along the time dimension. 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. However, we are interested in the *conditional* response of bank default risk to monetary shocks, which we are now going to document.

4.3 Bank Default Responses to Monetary Policy

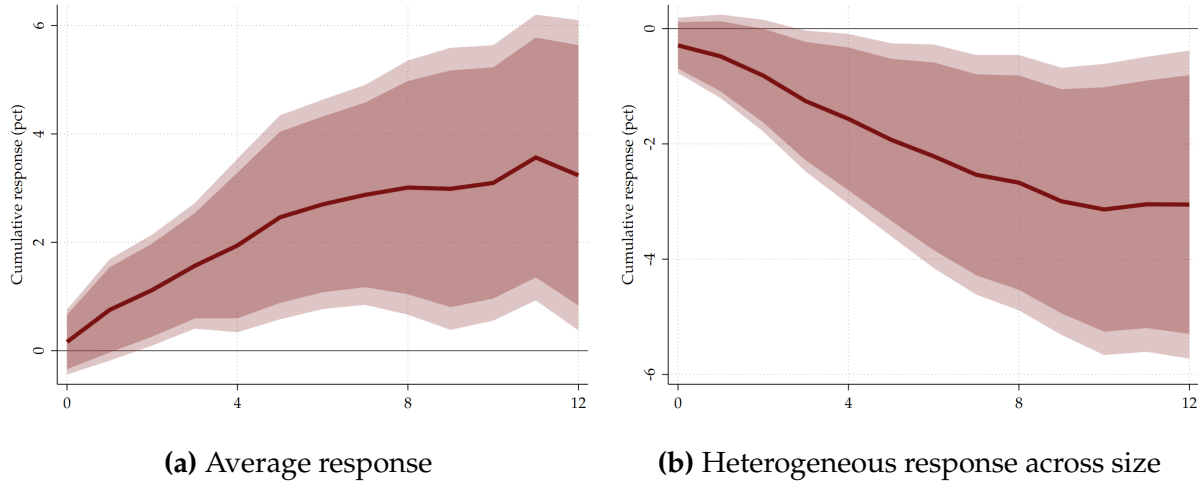
We now investigate the response of banks' default risk to monetary policy shocks, both in the aggregate and in the cross section. To estimate the average default-risk response to monetary shocks, we rely on the following panel local projection:

$$\Delta \ln(Y_{it+h}) = \alpha_{ih} + \psi_h \varepsilon_t + \sum_{\ell=1}^4 \gamma_{h\ell} \Delta \ln(Y_{it-\ell}) + \sum_{\ell=1}^4 \phi_{h\ell} X_{t-\ell} + u_{iht} \quad (28)$$

where $\Delta \ln(Y_{it+h})$ is the cumulative change in the inverse z-score and $X_{t-\ell}$ is a vector of controls which includes the CPI, real GDP, the return on the S&P 500 index, the federal funds rate and the excess bond premium. We use two-way clustered standard errors by bank and time. To analyze the heterogeneous effects of monetary policy on default risk in the cross-section of banks we use the same specification as in Equation (27), but use the inverse z-score as our dependent variable Y_{it} , as opposed to total assets.

Figure 15 reports the response of bank default risk to monetary policy shocks. Panel (a) shows that the default probability for the average bank increases by around 3% following a one standard deviation contractionary monetary shock. Thus, monetary tightenings increase default risk of the average bank in the economy. This identified moment is highly relevant for our model, as it puts structure on the Jacobian F^{dSdR} , which summarizes the partial equilibrium response of default risk to changes in the real interest rate. Furthermore, Panel (b) of Figure 15 shows that, compared to banks in the bottom 80% of the size distribution, default risk increases by less for banks in the top quintile of assets, following a monetary contraction. This result validates one of the main features of our model, namely that insolvency risk increases after a monetary contraction and especially so for small banks, whose distance to default is smaller. In general, this once again points to the importance of taking into account bank heterogeneity in assessing the transmission of monetary policy shocks.

Figure 15: Default Risk and Monetary Shocks



Notes: panel (a) plots estimates of ψ_h from (28) to a 1 standard deviation contractionary monetary shock. Panel (b) plots estimates of β_h from (27) to the same shock, plugging default risk as the outcome variable. Errors are two-way clustered at the time and bank level. Lightly shaded areas represent 95% confidence intervals. Darkly shaded areas are 90% confidence intervals. The y-axis represents the cumulative percentage change in the default probability. The x-axis represents quarters elapsed since the shock.

5 Conclusion

This paper develops a tractable macroeconomic model with heterogeneous financial intermediaries for the general equilibrium analysis of monetary and financial policies. A small number of sufficient statistics - the policy Jacobian matrices and the intertemporal marginal propensity to lend (iMPL) - completely summarize the full path of the response of macroeconomic variables to policy shocks. We solve the model in the sequence space, adding transparency to the distinction between direct (partial equilibrium) and indirect (general equilibrium) responses to alternative policy shocks. Our key result is twofold. First, the transmission of contractionary monetary policy shocks is amplified in a model with bank heterogeneity and incomplete markets, relative to a representative-bank perfect-insurance benchmark. This is due to the magnification effect working through the increase of insolvency risk in the bottom quintiles of the asset size distribution. Second, there is a trade-off between macroeconomic and financial stability. That trade-off is driven by the interaction of market incompleteness and costly bank default. Monetary tightenings simultaneously tame inflation and increase the likelihood of bank insolvency, which endogenously feeds back into the banking sector causing an amplified contraction in bank lending. We extend our baseline framework to the cases of imperfect credit and deposit market competition, apply it to the study of monetary, forward guidance, liquidity, and macroprudential policies, and offer empirical support for all of the model's key

mechanisms and testable predictions.

References

- Acharya, S., E. Challe, and K. Dogra**, “Optimal monetary policy according to HANK,” *American Economic Review*, 2023, 113.
- Adrian, T. and H. S. Shin**, “Liquidity and Leverage,” *Journal of Financial Intermediation*, 2010, 19(3), 418–437.
- and **N. Boyarchenko**, “Intermediary leverage cycles and financial stability,” *Staff Reports, Federal Reserve Bank of New York*, 2015, 567.
- Auclert, A.**, “Monetary Policy and the Redistribution Channel,” *American Economic Review*, 2019, 109(6).
- , **B. Bardoczy, M. Rognlie, and L. Straub**, “Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models,” *Econometrica*, 2021, 89(5).
- , **M. Rognlie, and L. Straub**, “Micro Jumps, Macro Humps: Monetary Policy and Business Cycles in an Estimated HANK Model,” *NBER Working Paper*, 2020, 26647.
- , – , and – , “The Intertemporal Keynesian Cross,” *NBER Working Paper*, 2023.
- , – , **M. Souchier, and L. Straub**, “Exchange Rates and Monetary Policy with Heterogeneous Agents: Sizing up the Real Income Channel,” *NBER Working Paper 28873*, 2021.
- Baqae, D., E. Farhi, and K. Sangani**, “The Supply-Side Effects of Monetary Policy,” *Journal of Political Economy*, 2023.
- Bayer, C., R. Luetticke, L. Pham-Dao, and V. Tjaden**, “Precautionary Savings, Illiquid Assets, and the Aggregate Consequences of Shocks to Household Income Risk,” *Econometrica*, 2019, 87.
- Begenau, J. and T. Landvoigt**, “Financial Regulation in a Quantitative Model of the Modern Banking System,” *Review of Economic Studies*, 2022, 89.
- , **S. Bigio, J. Majerovitz, and M. Vieyra**, “A Q-Theory of Banks,” *NBER Working Paper*, 2021, 27935.
- Bellifemine, M., A. Couturier, and R. Jamilov**, “The Regional Keynesian Cross,” *Working Paper*, 2023.
- Benhabib, B., A. Bisin, and M. Luo**, “Wealth distribution and social mobility in the US: A quantitative approach,” *American Economic Review*, 2019, 109(5).
- and – , “Skewed Wealth Distributions: Theory and Empirics,” *Journal of Economic Literature*, 2018, Forthcoming.
- Bernanke, B. and M. Gertler**, “Inside the Black Box: The Credit Channel of Monetary Policy Transmission,” *American Economic Review*, 1995, 9.
- Bianchi, J. and S. Bigio**, “Banks, Liquidity Management and Monetary Policy,” *Econometrica*, 2022, 90(1).
- Bigio, S. and Y. Sannikov**, “A Model of Credit, Money, Interest, and Prices,” *NBER Working Paper 28540*, 2021.
- Bilbiie, F.**, “Limited asset markets participation, monetary policy and (inverted) aggregate demand logic,” *Journal of Economic Theory*, 2008, 140 (1), 162–196.
- Bocola, L.**, “The Pass-Through of Sovereign Risk,” *Journal of Political Economy*, 2016, 124.
- Boppart, T., P. Krusell, and K. Mitman**, “Exploiting MIT shocks in heterogeneous-agent economies: the impulse response as a numerical derivative,” *Journal of Economic Dynamics and Control*, 2018, 89.

- Brunnermeier, M. and L. Pedersen**, “Market Liquidity and Funding Liquidity,” *Review of Financial Studies*, 2009, 22, 2201–2238.
- **and Y. Sannikov**, “A Macroeconomic Model with a Financial Sector,” *American Economic Review*, 2014, 104(2), 379–421.
- Coimbra, N. and H. Rey**, “Financial Cycles with Heterogeneous Intermediaries,” *Review of Economic Studies*, 2023.
- Corbae, D. and P. D’Erasmus**, “Capital Buffers in a Quantitative Model of Banking Industry Dynamics,” *Econometrica*, 2021, 89 (6), 2975–3023.
- Davila, E. and A. Schaab**, “Optimal Monetary Policy with Heterogeneous Agents: Discretion, Commitment, and Timeless Policy,” *NBER Working Paper*, 2023, 30961.
- Del Negro, M., M. Giannoni, and C. Patterson**, “The forward guidance puzzle,” Technical Report 2023.
- Drechsler, I., A. Savov, and P. Schnabl**, “The deposits channel of monetary policy,” *Quarterly Journal of Economics*, 2017, 132 (4), 1819–1876.
- , – , **and –**, “Banking on Deposits: Maturity Transformation without Interest Rate Risk,” *Journal of Finance*, 2021, 76.
- Elenev, V., T. Landvoigt, and S. Van Nieuwerburgh**, “A Macroeconomic Model With Financially Constrained Producers and Intermediaries,” *Econometrica*, 2021, 89.
- Galaasen, S., R. Jamilov, R. Juelsrud, and H. Rey**, “Granular Credit Risk,” *NBER Working Paper 27994*, 2023.
- Gali, J.**, “Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications,” *Princeton University Press*, 2008.
- **and D. Debortoli**, “Idiosyncratic Income Risk and Aggregate Fluctuations,” *NBER Working Paper*, 2022, 29704.
- Galí, J., D. López-Salido, and J. Vallés**, “Understanding the Effects of Government Spending on Consumption,” *Journal of the European Economic Association*, 2007, 5(1).
- Gertler, M. and N. Kiyotaki**, “Financial Intermediation and Credit Policy in Business Cycle Analysis,” *Handbook of Monetary Economics*, 2010, 3, 547–599.
- **and P. Karadi**, “A Model of Unconventional Monetary Policy,” *Journal of Monetary Economics*, 2011, 58(1), 17–34.
- , **N. Kiyotaki, and A. Prestipino**, “Wholesale Banking and Bank Runs in Macroeconomic Modelling of Financial Crises,” *Handbook of Macroeconomics*, 2016, 2.
- , – , **and –**, “A Macroeconomic Model with Financial Panics,” *Review of Economic Studies*, 2020, 87(1).
- Gertler, Mark and Peter Karadi**, “Monetary Policy Surprises, Credit Costs, and Economic Activity,” *American Economic Journal: Macroeconomics*, January 2015, 7 (1), 44–76.
- 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).
- Hagedorn, M., J. Luo, I. Manovskii, and K. Mitman**, “Forward Guidance,” *Journal of Monetary Economics*, 2019, 102.
- Hazell, J., J. Herreno, E. Nakamura, and J. Steinsson**, “The Slope of the Phillips Curve: Evidence from U.S. States,” *Quarterly Journal of Economics*, 2021, Forthcoming.
- He, Z. and A. Krishnamurthy**, “Intermediary Asset Pricing,” *Journal of Financial Economics*, 2013, 103(2), 732–770.

- Hoggarth, G., R. Reis, and V. Saporta**, “Costs of banking system instability: some empirical evidence,” *Journal of Banking Finance*, 2002, 26.
- Jamilov, R. and T. Monacelli**, “Bewley Banks,” *CEPR Discussion Paper No. DP15428*, 2023, p. Available at SSRN: <https://ssrn.com/abstract=3737561>.
- Jermann, U. and V. Quadrini**, “Macroeconomic Effects of Financial Shocks,” *American Economic Review*, 2013, 102(1), 238–271.
- Jordà, O.**, “Estimation and Inference of Impulse Responses by Local Projections,” *American Economic Review*, March 2005, 95 (1), 161–182.
- Juelsrud, J. and E. Wold**, “Risk-weighted capital requirements and portfolio rebalancing,” *Journal of Financial Intermediation*, 2020, 41.
- Kaplan, G., B. Moll, and G. Violante**, “Monetary Policy According to HANK,” *American Economic Review*, 2018, 108(3).
- , **K. Mitman, and G. Violante**, “The Housing Boom and Bust: Model Meets Evidence,” *Journal of Political Economy*, 2020, 128 (9).
- Kashyap, A. and J. Stein**, “The impact of monetary policy on bank balance sheets,” *Carnegie-Rochester Conference Series on Public Policy*, 1995, 42.
- and – , “What Do a Million Observations on Banks Say about the Transmission of Monetary Policy?,” *American Economic Review*, 2000, 90(3).
- Kimball, M.**, “The quantitative analytics of the basic neomonetarist model,” *Journal of Money, Credit and Banking*, 1995, 27(4).
- Klenow, P. and J. Willis**, “Real Rigidities and Nominal Price Changes,” *Economica*, 2016, 83.
- Krueger, D., K. Mitman, and F. Perri**, “Macroeconomics and Household Heterogeneity,” *Handbook of Macroeconomics*, 2016, 2.
- Krusell, P. and A. Smith**, “Income and Wealth Heterogeneity in the Macroeconomy,” *Journal of Political Economy*, 1998, 106, 867–896.
- Laeven, L. and F. Valencia**, “Systemic Banking Crises Database: An Update,” *IMF Working Paper*, 2012, 12/163).
- and **R. Levine**, “Bank governance, regulation and risk taking,” *Journal of Financial Economics*, 2009, 93.
- Lee, S., R. Luetticke, and M. Ravn**, “Financial Frictions: Macro vs Micro Volatility,” *CEPR DP*, 2020, 15133.
- Lenel, M. and R. Kekre**, “Monetary Policy, Redistribution, and Risk Premia,” *Econometrica*, 2022, 90.
- Mendicino, C., K. Nikolov, J. Rubio-Ramirez, J. Suarez, and D. Supera**, “Twin Default Crises,” *ECB Working Paper 2414*, 2020.
- Nagel, S. and A. Purnanandam**, “Banks’ Risk Dynamics and Distance to Default,” *Review of Financial Studies*, 2019, 33.
- Nuno, G. and C. Thomas**, “Bank Leverage Cycles,” *American Economic Journal: Macroeconomics*, 2017, 9(2).
- Olea, J. L. Montiel and M. Plagborg-Møller**, “Local Projection Inference Is Simpler and More Robust Than You Think,” *Econometrica*, 2021, 89 (4), 1789–1823.
- Ottobello, P. and T. Winberry**, “Financial Heterogeneity and the Investment Channel of Monetary Policy,” *Econometrica*, 2020, 88(6).
- Ravn, M. and V. Sterk**, “Macroeconomic Fluctuations with HANK SAM: an Analytical

- Approach," *Journal of the European Economic Association*, 2020, 19(2).
- Rull, V. Rios, T. Takamura, and Y. Terajima**, "Banking Dynamics, Market Discipline and Capital Regulations," *Manuscript*, 2020.
- Schaab, A. and S. Y. Tan**, "Monetary and Fiscal Policy According to HANK-IO," *Working Paper*, 2022.
- Scharfstein, D. and A. Sunderam**, "Market Power in Mortgage Lending and the Transmission of Monetary Policy," *Working Paper*, 2016.
- Sidrauski, M.**, "Inflation and Economic Growth," *Journal of Political Economy*, 1967, 75.
- Werning, I.**, "Incomplete Markets and Aggregate Demand," *NBER Working Paper*, 2015, 21448.
- Woodford, M.**, "Interest and Prices: Foundations of a Theory of Monetary Policy," *Princeton University Press*, 2003.