A Quantitative Theory of Relationship Lending[*]

Kyle Dempsey
Ohio State University

Miguel Faria-e-Castro
FRB St. Louis

This version: January 12, 2024

Abstract

Borrower-lender relationships tend to be long-lasting, and lender switching is infrequent. What are the aggregate consequences of these facts? We address this question in a model of heterogeneous banks subject to financial frictions. We incorporate lending relationships using loan portfolio adjustment costs for borrowers and accumulation of "relationship capital" for lenders. The model's implied loan demand system is directly estimated on administrative loan-level micro data to recover the key novel parameters governing the strength and persistence of lending relationships. We find that financial and relationship capital are complements, and so banks constrained with respect to one tend to be constrained with respect to the other. Relationship lending generates endogenous persistence in the economy's response to financial crises, with recoveries becoming more sluggish, and it also affects the interplay between monetary policy and financial stability.

JEL Classification: E4, G2

[*]The views expressed here are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of St. Louis or the Federal Reserve System. We thank Asha Bharadwaj for research assistance with the data, and Mike Carter and Jeongwoo Moon for research assistance with computation. We thank seminar participants at the Federal Reserve Board, the Federal Reserve Bank of Philadelphia, the Federal Reserve Bank of Richmond, the Bank of Canada, the Bank of England, Queen Mary University of London, and London Business School, as well as conference participants at the 2021 Society for Economic Dynamics, 2021 CEBRA, 2021 Econometric Society Winter Meetings for helpful comments. First version: February 2020. Contact: `dempsey.164@osu.edu`.

1 Introduction

Banks operate in imperfectly competitive markets in their core business activities: deposit taking and loan making.[1] Market power generates economic profits for banks by enabling them to lend at interest rates above the fair (risk-adjusted) cost of capital and borrow at interest rates below the prevailing risk-free rate. Such economic profits have long been considered a "feature, not a bug", as they generate franchise value that curbs risk-taking by banks, thereby promoting financial stability (Demsetz et al., 1996).

This paper studies one source of banks' lending market power: long-lasting lending relationships between borrowers and lenders. We present a theory in which these relationships introduce dynamic considerations into banks' loan pricing and financing decisions. Specifically, banks internalize the fact that while higher interest rates may generate larger profits today, they may erode relationships and thereby worsen the bank's lending prospects tomorrow. The central focus of this paper is to quantify the *aggregate* consequences of lending relationships by showing how they interact with financial frictions at the individual bank and industry levels.

We study lending relationships in the context of a dynamic equilibrium model with heterogeneous banks who are subject to financial constraints. We model lending relationships using two key features. First, borrowers may borrow from many banks, but face costs of adjusting the shares of their total lending sourced from each bank. These adjustment costs endow banks with market power in lending. The adjustment is relative to a single metric – which we term "relationship capital" – which summarizes borrowers' prior loan sourcing decisions in a manner akin to a "deep habit" (e.g. Ravn et al. (2006)). In this setup, banks face loan demand curves which depend on both their interest rates and their relationship capital; stronger relationships mean higher levels and lower price elasticities of loan demand. Likewise, aggregate loan demand depends not only on the interest rates offered, but on the joint distribution of interest rates and relationships. Second, whether a borrower's relationship with a given bank strengthens or weakens period-to-period depends on the bank's pricing decision. Thus loan market power has a dynamic component: like standard monopolists, banks extract rents commensurate with the static inverse elasticity of loan demand, but these rents are extracted over the (potentially infinite) life of the relationship.

This paper makes four main contributions.

First, our model is the first to our knowledge which can be used to evaluate how lending relationships shape industry-level and aggregate outcomes in the presence of financial constraints. Importantly, our framework nests a simpler competitive benchmark in which borrowers do not face loan portfolio adjustment costs. In this case, there is no meaningful notion of relationships, and banks take prices as given and choose how much to lend, how many deposits to issue, and how large a dividend to pay out. This nesting allows us to neatly detail how relationships alter bank behavior in our quantitative analysis.

[1] See, for example, Berger and Hannan (1998). Banking industry concentration and market power is well documented both internationally (Fernández de Guevara et al., 2005) and in the U.S., which has experienced a stark secular decline in the number of banks (Prescott and Janicki, 2006) over the last several decades.

Second, we establish that the three key parameters which govern lending relationships in the model can be directly estimated using loan-level micro-data. This proceeds in two steps. First, we apply the method of Amiti and Weinstein (2018) to estimate the model's bank-level demand curve. This yields an estimate of the *static* interest rate elasticity of loan demand, and this pins down the intensity of the loan portfolio adjustment costs. Second, we use the residuals of this demand curve estimation to estimate the law of motion for relationships at the bank level. This informs the *dynamic* components of relationship capital, providing estimates of its persistence and responsiveness to loan prices. This step uses the fact that, through the lens of our model, the key shifter of loan demand beyond price is relationship capital. Beyond facilitating our quantitative analysis, the ability to estimate the demand system directly helps justify how we specify relationships in the model. In particular, one might consider two alternative "aggregators" which might be able to encode relationships: constant elasticity of substitution (CES) and Kimball (1995). The former is rejected by the fact that banks demonstrably face different loan demand elasticities in the data; the latter allows for larger banks to face lower elasticities, but does not allow for us to cleanly recover the parameters determining the dynamics of relationships.

We show that, given our calibration and structural estimation, financial and relationship capital emerge as complements. Relationship capital governs the speed at which banks recapitalize in the wake of adverse financial shocks. When a bank receives a negative shock and has high customer capital, it can "expend" that capital by charging high interest rates in order to weather and adverse shock. Banks with more customer capital therefore recapitalize more quickly than other banks. Customer capital therefore functions as an extra buffer, on top of conventional equity.

Finally, we show that relationship lending plays an important role in determining how the economy responds to aggregate shocks. We compare the responses to aggregate shocks of our baseline economy, with relationship lending, to those of a perfectly competitive banking economy. In response to a large aggregate financial shock, where banks lose a fraction of their equity, the relationship lending economy generates more muted effects on impact in credit market variables. These more muted effects are also more persistent: relationship lending introduces endogenous persistence that make bank recapitalizations "slower", as banks manage the complementarity of financial and customer capital. We also show that both types of economies feature strong pass-through of monetary policy shocks to credit market variables, but that the effects on bank net worth are very different. A monetary policy tightening causes net worth to fall in the competitive case, and to rise in the relationship lending case. In a competitive economy, higher costs of lending induces banks to deleverage and thus deplete their net worth. In the relationship economy, banks have incentives to sustain their lending, and a rise in deposits costs induces them to switch their funding structure towards retained earnings. The nature of bank relationships can therefore have important implications for how monetary policy interacts with financial stability. The third type of shock we consider is a shock to credit demand, and we show that the relationship lending economy exhibits larger fluctuations in interest rates, but smaller fluctuations in the quantity of credit.

The rest of the paper is structured as follows. The remainder of this section discusses our paper's context in the relevant literatures. Section 2 establishes several empirical facts which motivate our approach to modeling lending relationships. Section 3 presents our model environment. Section 4 discusses how we take our model to the data. Sections 5 and 6 present the main results from our quantitative model, with the former focusing on the cross-section and the latter focusing on aggregate dynamics. Section 7 concludes and describes some promising areas for future research related to this paper.

Related Literature This paper contributes to three distinct literatures in macroeconomics and finance: customer capital in macroeconomic models, structural models of banking, and empirical studies of the effects of bank market power.

While the dynamics of customer capital can be related to an older literature on consumption habits, a seminal formalization of customer capital in a macroeconomic model is the work of Gourio and Rudanko (2014). In the context of nonfinancial firms, Gilchrist et al. (2017) argue that the interaction between customer capital dynamics and financial constraints was key to explain the dynamics of inflation in the U.S. during the Great Recession. Our modeling of customer capital dynamics is reminiscent of theirs, also in the context of a model of heterogeneous firms. We argue that customer capital interacts with capital constraints that are specific to the banking industry. this is critical to understanding dynamics around recent recessions, since the aggregate capitalization of the banking sector has been argued to be a relevant state variable for macroeconomic performance (Adrian and Boyarchenko, 2012).

We study the effects of bank customer capital from a positive perspective in the context of a dynamic equilibrium model of heterogeneous banks that take deposits, make loans, and face constraints that depend on their net worth. We therefore contribute to an emerging literature that employs the tools of heterogeneous agent macroeconomic models to study questions that are related to the banking industry. Bigio and Bianchi (2014) use a quantitative model with heterogeneous banks and liquidity frictions in the interbank market to study monetary policy implementation. Corbae and D'Erasmo (2019) use a quantitative model of heterogeneous banks where size is correlated with market power to study the effects of capital requirements. We take these requirements as given, and study how their interaction with customer capital affects the overall stability of the banking system. (Neri et al., 2010) introduce a monopolistically competitive banking sector in an otherwise standard monetary DSGE model, and study how this affects the transmission of standard shocks.

Finally, our paper relates to a broader empirical literature that studies the efficiency and stability consequences of banking market power and concentration. Recent work on this topic has been focused on market power on the deposits market. Egan et al. (2017) use detailed branch-level data on deposit quantities and prices to estimate a demand system for secured and unsecured deposits at large U.S. banks. These estimates are then combined with a dynamic model of bank runs, which allows them to study the probabilities of counterfactual runs on these large banks during the financial crisis. Drechsler et al. (2017) argue that bank market power in the deposit markets gives rise to a new channel of transmission for monetary policy in the U.S.

We focus instead on customer capital on the loan side of the balance sheet. Our interpretation of bank customer capital is closely linked to the notion of relationship lending, the fact that both banks and borrowers find it worthwhile to maintain long-standing relationships. A long-standing literature has found that banks smooth loan rates when faced with adverse cost of funding shocks (Berger and Udell, 1995; Berlin and Mester, 1998). There is also an extensive theoretical literature that derives conditions under which the optimal contract between a lender and a borrower shares some of those features under a variety of frictions, such as asymmetric information, search frictions, or switching costs. We take a different approach, by taking the lending contract and the process for customer capital dynamics as given, and instead study their macroeconomic implications.

2 Empirical Motivation

In this section, we present some facts on bank loan markets for the US that serve as motivation for our model analysis. In particular, we use loan-level micro data for the U.S. to document three main facts regarding bank loans: (i) loan markets are highly concentrated, (ii) switching between banks is relatively infrequent, and (iii) there exists an interest rate lifecycle for new lending relationships, featuring low interest rates in the beginning that rise over the length of the relationship.

2.1 Data

Our main source of data is the Commercial & Industrial loan schedule H.1 of the Federal Reserve's FR Y-14Q dataset (Y-14 for short). This is a quarterly panel of individual loan facilities held in the books of the largest bank holding companies (BHCs) in the US.[2] The Y-14 includes all loan facilities held in the books of covered BHCs with commitments larger than $1 million. It contains detailed information about the characteristics of each loan, such as the identity of the borrower, the type of loan, interest rate, purpose of loan, etc.

We restrict our loan sample along several dimensions. First, we exclude loans to non-US addresses, loans in currencies other than the US dollar, and loans to firms without a US Tax Identification Number (TIN, our main firm identifier). We also exclude loans to the financial and public administration sectors, that is, to any entity classified as a bank or with NAICS code 52 or 92.[3] Due to their different nature and imperfect coverage, we also drop syndicated loans.

2.2 Facts on US Bank Loan Markets

2.2.1 Loan switching

Figure 1 presents time series plots on the percentage of loans that correspond to "switches", as a percentage of total outstanding loans. Our definition of "switch" is adapted from Ioannidou and Ongena (2010): a loan is considered a switch if it is a new loan, and if it originates from a bank with

[2] Until 2019, the dataset includes all BHCs with more than $50 billion in assets. From 2019 onwards, only BHCs with more than $100 billion in assets are included.

[3] We also exclude loans made to companies with NAICS codes 5312 (Offices of Real Estate Agents and Brokers) or 551111 (Offices of Bank Holding Companies).

whom the firm has had no (observable) relationship in the past year. The time series plots show that both in terms of value and loan counts, switches are between 2 and 3.5% of total loans. Thus switching is relatively infrequent.

It is worth noting that due to the characteristics of the Y-14 dataset, we are likely to be overestimating the frequency of switching. First, loan observations may enter and/or leave our panel for many reasons other than origination or maturity. A loan may have been originated with a committed exposure of under \$1 million, with a credit limit increase above \$ 1 million being renegotiated at a later date. In that case, we only observe the loan after the credit limit has increased. Additionally, banks may not necessarily keep originated loans on their portfolio, and may dispose of them by selling them to other financial institutions, for example. Second, since we only observe credit facilities above \$1 million dollars, we do not observe small firms that borrow lower amounts. It is well documented that large firms tend to have more relationships and switch more often than smaller ones (Petersen and Rajan, 1994). Among studies that use more comprehensive loan-level datasets, Ioannidou and Ongena (2010) find that 3% of all originations are classified as switching loans for Bolivia, while Farinha and Santos (2002) find that on average 4% of all yearly originations involve switching, using data for Portugal.

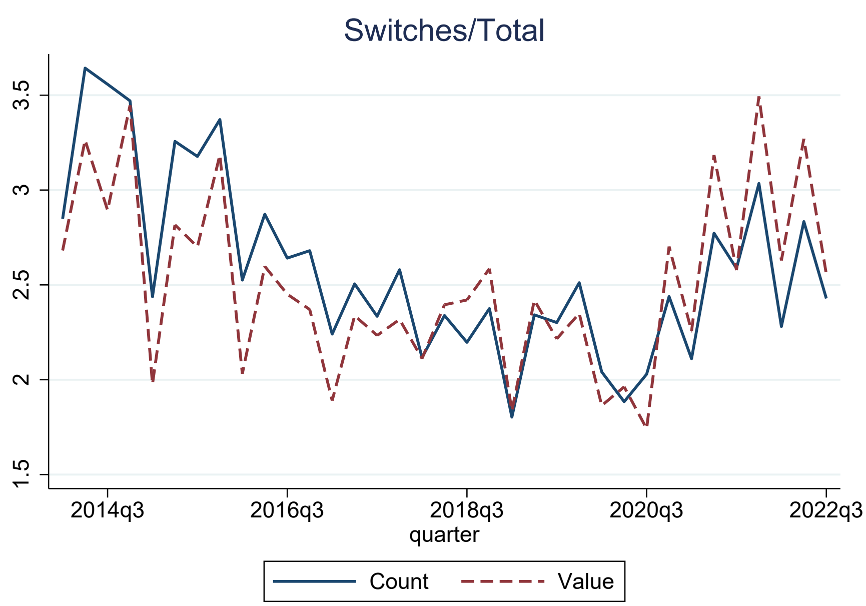


Figure 1: **Switches as a percentage of total outstanding loans**

Notes: See text for details. A loan is classified as a switch if it is (i) a new loan, and (ii) from a bank with which the firm has had no relationship in the past year.

2.2.2 Life cycle of relationships

We next investigate how interest rates evolve over the lifecycle of firm-bank relationships. To study this, we follow an approach inspired by the methodology in Ioannidou and Ongena (2010): we identify loan originations that correspond to new relationships (again, defined, as the absence of an observable relationship between a firm and a bank in the previous 4 quarters) and match those with loan originations that correspond to existing relationships and that have similar observed characteristics.

More specifically, we take all loans that correspond to new relationships between a firm and a bank, and find similar loans that do not correspond to new relationships. We match on the following observable characteristics: (i) same origination date for the loan, (ii) same maturity (in years), (iii) same originating bank, (iv) same percentile of loan size, (v) same loan type (term loan, credit line, or other), (vi) same interest rate variability, and (vii) same percentile of default probability. Since there are more nonswitching loans than switching loans in our data, we match each unique nonswitching loan with a similar switching loan, meaning that some switching loans may appear multiple times in the dataset. This procedure generates 20,155 matched loan pairs.

For each pair p, we compute the spread between the switching and nonswitching loans and denote it by $y_{p,t}$, where t is the quarter in which the spread is computed. We then run regressions of the following type:

$$y_{p,t} = \sum_{i=1}^{13} \gamma_i \mathbf{1}[\tau_{p,t} = i] + \epsilon_{p,t} \tag{1}$$

where $\tau_{p,t}$ is time since origination for the matched loan pair p at time t, measured in quarters. Time since origination can alternatively be interpreted as the length of the relationship for the switching loan (by construction, as a switching loan is such that no relationship existed). We interpret the estimated coefficients γ_i as the average discount or premium that a switching loan obtains relative to a non-switching loan over its lifecycle.[4]

The estimation results for regression (1) are summarized in Figure 2, which plots the estimated marginal effects along with 90% confidence intervals based on robust standard errors. The figure shows that the average spread between switching and nonswitching loans is negative for most of the first year, meaning that switching loans pay, on average, lower interest rates. After one year, the spread becomes positive, meaning that switching loans start paying interest rates that are, on average, higher than those of nonswitchers. This positive spread seems to persist for the remainder of the relationship, even though the spread becomes statistically indistinguishable from zero more or less after three years. This lifecycle pattern is consistent with the one detected by Ioannidou and Ongena (2010) for Bolivia: they also find that switchers start out by paying lower interest rates than nonswitchers, but that these interest rates increase over time and eventually exceed those of nonswitchers. The authors use these findings to discriminate between alternative theories of firm-bank relationships, as these results suggest that relationships are influenced by switching costs that give rise to a hold-up problem: the bank initially attracts the borrower using a "teaser rate", and

[4] We consider $i = 1, \ldots, 13$, where each i is a quarter since origination and $i = 13$ stands for twelve plus quarters.

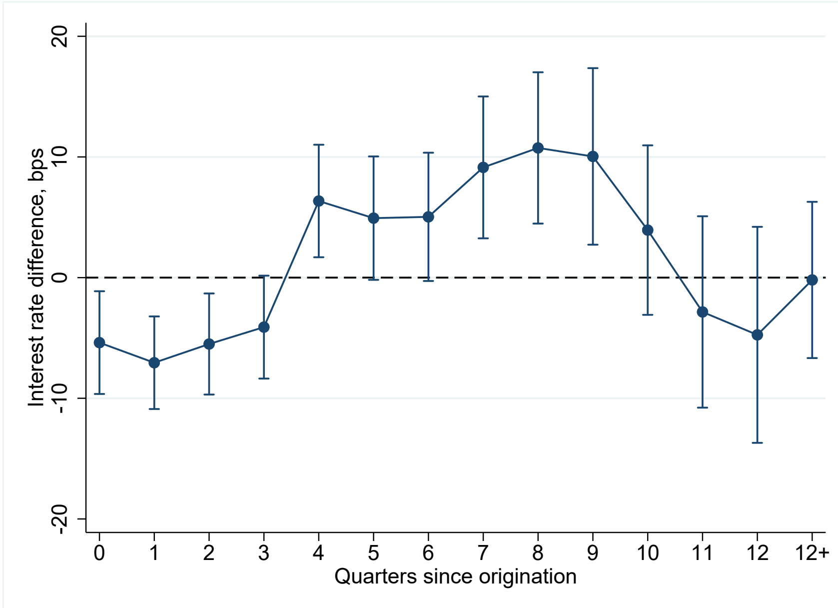


Figure 2: **Average spread between loans for new and existing relationships**

Notes: See text for details. At each time since loan origination, the dot represents the point estimate of γ_i from (1), and the bars represent the associated 90% confidence interval.

then exploits the fact that switching is costly in order to extract surplus.

It is also worth pointing out that we find spreads of smaller magnitudes than those found by Ioannidou and Ongena (2010). This can be explained by many factors, including lower average interest rates in the US versus Bolivia (which results in interest rate compression), as well as the fact that the firms in our sample tend to be larger than those in Bolivia and potentially safer, which again results in less variability in interest rates. The \$1 million dollar loan cutoff effectively excludes most small enterprises from our sample.

3 A Model of Relationship Lending

Motivated by the empirical evidence in the previous section, we develop a model of relationship lending that endogenously generates some of these facts. We consider a stationary economy populated by a unit continuum of monopolistically competitive banks $j \in [0, 1]$ and a continuum of identical firms who borrow from them. Time is discrete and infinite, and there is a single good. The risk-free rate is $\overline{r}$, which defines a risk-free discount price of $\overline{q} = (1 + \overline{r})^{-1}$, the wage rate is $\overline{w}$, and the user cost of capital is $\overline{uc}$. All these prices are exogenously specified. While the model's focus is on bank behavior, described in Section 3.2, we present first the firms' problem in Section 3.1 since it delivers the demand system banks face and helps introduce notation. Section 3.3 defines equilibrium. Finally, Section 3.4 discusses the motivation for and implications of the main assumptions in our framework. Proofs of all propositions are contained in Appendix A.

3.1 The firm: defining bank-specific and aggregate loan demand

There is a continuum of firms indexed by $i \in [0, 1]$. All firms are identical, and so we focus on a representative firm. The representative firm operates a decreasing returns production technology using labor n and capital k, producing $y = Ak^\alpha n^\eta$ units of output for $\alpha + \eta \in (0, 1)$ given total factor productivity A. Each period, the firm chooses: (i) how much labor and capital to hire; (ii) how much to borrow; and (iii) the *sourcing* of its borrowing across banks j. The firm is subject to a working capital constraint as in Christiano et al. (2005): total lending must be at least a fraction $\kappa \geq 0$ of its total costs, which include the wage bill and the costs of renting capital. The firm's total loan demand today is L', and the distribution of borrowing across banks is $\mathcal{L}' = \{\ell'_j\}$, where ℓ'_j is the face value of this period's loan from bank j. The discount price of a loan from bank j is q_j, and we denote the set of loan prices across banks by $\mathcal{Q} = \{q_j\}$.

We model lending relationships as follows. For each bank j, we summarize the intensity of a firm's relationship with that bank by s_j; the set of relationships across all banks is $\mathcal{S} = \{s_j\}$. We assume it is costly for a firm to source its loans in a way that deviates from the distribution of relationships. We implement this with a quadratic cost function with scale parameter $\phi \geq 0$, which penalizes deviations in the share of total lending sourced from bank j from the (relative) intensity of the firm's relationship with bank j.[5] For tractability, we assume that borrowers take current relationships as given and do not internalize how current loan sourcing decisions affect future relationships, in the spirit of "external" habits in the literature (e.g. Ravn et al. (2006)).

Under this formulation, a firm does not directly care about the "identity" j of any bank from which it borrows; rather, it cares only about the intensity of its relationship with the bank, s, and the loan price the bank offers, q. Therefore, the decision-relevant object which defines borrowing opportunities for the firm is the joint density of prices and relationships across banks, $\mu(q, s)$, which summarizes $\{\mathcal{Q}, \mathcal{S}\}$. The firm's dynamic optimization problem may then be written recursively as:

$$
\begin{aligned}
W(\mathcal{L};\mu) \quad = \quad & \max_{n,k,L',\{\ell'(q,s)\}} \underbrace{Ak^\alpha n^\eta - \overline{w}n - \overline{uc}k}_{\text{net operating income}} + \underbrace{L' - \int \ell(q,s)\mathrm{d}\mu(q,s)}_{\text{borrowing net of repayments}} \qquad (2)\\
& - \underbrace{\frac{\phi}{2}L'\int\left(\frac{q\ell'(q,s)}{L'} - 1 - (s-S)\right)^2 \mathrm{d}\mu(q,s)}_{\text{sourcing adjustment costs}} + \underbrace{\overline{q}\mathbb{E}\left[W(\mathcal{L}';\mu)\right]}_{\text{continuation value}}
\end{aligned}
$$

$$
\text{subject to [working capital]} \qquad \kappa(\overline{w}n + \overline{uc}k) \leq L' \qquad (3)
$$

$$
\text{[loan sourcing]} \qquad L' \leq \int q\ell'(q,s)\mathrm{d}\mu(q,s) \qquad (4)
$$

The firm's flow profits in (2) sum net operating income $Ak^\alpha n^\eta - \overline{w}n - \overline{uc}k$ and net borrowing (new loans less repayments), less adjustment costs. $L' \equiv \int q\ell'(q,s)\mathrm{d}\mu(q,s)$, defined in the loan sourcing constraint (4), are total funds borrowed today, and $S \equiv \int s\mathrm{d}\mu(q,s)$ is the average relationship intensity. The firm discounts at the risk free rate and recognizes that, in a stationary equilibrium,

[5] We need not restrict the adjustment costs to be quadratic; this assumption yields intuitive closed forms that facilitate exposition and estimation and make computation more efficient. We discuss this later in more detail.

the joint distribution of prices and relationship intensities will be the same tomorrow as today, even if specific banks shift around in the distribution. Constraint (3) is the working capital constraint. Note again that the borrower does not take into account its choice of loan portfolio today on habits tomorrow: hence there is no "law of motion" for s here. This reflects the externality of habits: each individual firm is infinitesimally small and does not internalize the impact of its actions on relationship intensities.

Intuitively, the adjustment cost function induces firms to choose a distribution of loan sourcing that aligns with the distribution of relationship intensities. All else equal, firms would like to choose their borrowing shares at each bank in line with the relative intensity of their relationship with that bank, since this implies no adjustment costs.[6] The quadratic functional form is not essential to our results (see Appendix A.3), but gives rise to a linear demand system that is amenable to estimation. The following proposition summarizes the loan demand system that arises from the firm's problem:

Proposition 1. ***(Loan demand system)*** *Given a joint distribution of prices and relationship intensities $\mu(q, s)$, bank-specific loan-demand $\ell'(q, s)$ and aggregate loan demand L' satisfy*

$$\frac{q\ell'(q, s; \mu)}{L'(\mu)} = 1 + (s - S) - \frac{\overline{q}}{\phi}\left[r(q) - R(\mu)\right] \text{ for all } q, s \tag{5}$$

$$L'(\mu) = \kappa(\alpha + \eta)\left[\frac{A\left(\frac{\alpha}{\overline{u}\overline{c}}\right)^{\alpha}\left(\frac{\eta}{\overline{w}}\right)^{\eta}}{1 + \kappa(\overline{q}\tilde{R}(\mu) - 1)}\right]^{\frac{1}{1-\alpha-\eta}} \tag{6}$$

where $r(q) = q^{-1}$ is the interest rate implied by the bank's choice of loan price q, $S = \frac{\pi\rho_q}{1-\pi\rho_s}$ is the average relationship intensity, $R(\mu) = \mathbb{E}_{\mu}[r(q)]$ is the average interest rate, and $\tilde{R}(\mu)$ is the effective interest rate, defined as:

$$\tilde{R}(\mu) = R(\mu) + \mathbb{C}_{\mu}[r(q), s] - \frac{\overline{q}}{2\phi}\mathbb{V}_{\mu}[r(q)] \tag{7}$$

which adjusts the average interest rate for the covariance of interest rates and relationship intensities $\mathbb{C}_{\mu}(r, s)$, and the overall variance of interest rates $\mathbb{V}_{\mu}(r)$.

Equation (5) defines the demand curve faced by a bank with relationship intensity s charging price q as a function of aggregate loan demand, the average interest rate, and the average relationship intensity. The loan demand at a given bank is decreasing in the loan rate spread over the benchmark $r(q)-R(\mu)$, with elasticity governed by the risk free rate and the adjustment cost. This is a standard price effect: when a given bank's loans are cheap relative to its competition, that bank will capture a higher share of total lending, all else equal. Steeper adjustment costs (higher ϕ) imply a lower elasticity of loan demand with respect to price. In addition, bank-level loan demand increases in the strength of the firm's relationship with that bank s. Thus, stronger existing lending relationships simultaneously increase the level and lower the price elasticity of loan demand, endowing these banks with more effective market power.

Equation (6) determines aggregate loan demand. Conveniently, the entire joint distribution of loan prices and relationship intensities may be summarized by a single statistic: the effective interest

[6] "Relative" relationship intensities are simply deviations between s and S.

rate $\tilde{R}(\mu)$ from equation (7). This term has three components. First, the average interest rate term $R(\mu)$ conveys that when interest rates are higher on average, aggregate loan demand is lower. Second, loan demand is dampened further when the banks with whom the firm has the strongest relationships charge the highest spreads, as indicated by the covariance term $\mathbb{C}_\mu(r,s)$. Third, holding fixed the previous two terms, greater cross-sectional interest rate variance, $\mathbb{V}_\mu(r)$, burnishes loan demand by creating scope for the firm to gravitate towards cheaper banks.

3.2 Banks: dynamic pricing in the presence of relationships

Each bank uses retained earnings (its net worth), newly issued equity $e<0$, and deposits $d' \geq 0$ (investment in riskless securities if $d'<0$) to make loans ℓ' at discount price q. Deposits are risk-free (insured) and issued at exogenous price $\overline{q}^d$ that is the same for all banks. Banks value dividends $e \geq 0$ and face costs of issuing equity: we follow the dynamic corporate finance literature and model bank preference for dividends via the increasing function $\psi(e)$. The value of positive dividends is simply $\psi(e)=e, e\geq 0$, but equity issuance (i.e. negative dividends) is costly, with $\psi'(e)>1$ for $e<0$. Banks' net worth can be shifted by an idiosyncratic shock z which proxies excess returns on unmodeled sectors of banks' loan portfolios and is drawn from a distribution $\Gamma(z,z')$. Finally, we assume that banks exit with exogenous iid probability $1-\pi$ for $\pi \in [0,1]$ each period. Exiting banks are replaced with banks with no net worth and no relationships at the start of the next period.

Banks face a regulatory capital constraint that specifies that total lending, scaled by a factor χ, may not exceed the total value of equity, reflecting the current period's lending and financing decisions. We assume that each bank is monopolistically competitive, setting its loan price taking as given the bank-specific loan demand function (5), as well as the level of aggregate demand and the key moments of the distribution $\mu(q,s)$ described in Proposition 1. Crucially, individual banks do take into account the impact of their lending choices today on their relationship intensities tomorrow. We assume that relationships build up over time as a convex combination of the current relationship intensity (coefficient ρ_s) and the share of total loans issued today (coefficient ρ_q).

At the beginning of the period, a bank's state can be summarized by its net worth, n, its relationship intensity, s, and its realization of the idiosyncratic shock, z. We can write the problem of an individual bank recursively as:

$$\begin{aligned}
V\left(n,s,z;\mu\right) \quad &= \max_{q,e,\ell'\geq 0,d',s',n'} \psi(e) + \overline{q}\pi\mathbb{E}\left[V\left(n',s',z';\mu\right)\right] && (8)\\
\text{subject to [budget constraint]} \quad & q\ell' + e \leq n + z + \overline{q}^d d' && (9)\\
\text{[capital requirement]} \quad & \chi q\ell' \leq q\ell' - \overline{q}^d d' && (10)\\
\text{[relationship building]} \quad & s' = \rho_q \frac{q\ell'}{L'(\mu)} + \rho_s s && (11)\\
\text{[market power]} \quad & \ell' = \ell(q,s;\mu) && (12)\\
\text{[net worth accumulation]} \quad & n' = \ell' - d' && (13)
\end{aligned}$$

The optimal policies for the control variables associated with solving this problem are denoted $g_y(x)$

for $y \in \{q, e, \ell', d', s'\}$, where $x = (n, s, z)$ summarizes banks' state variables.

The bank's objective function (8) reflects its valuation of the present value of dividends net of issuance costs, discounted at factor $\overline{q}$ and adjusting for the exit probability $1-\pi$.[7] Constraint (9) is the bank's flow budget constraint: loan issuances and dividends must be financed from net worth, adjusted for the realization of the shock, deposits. Constraint (10) is the capital requirement: the value of the bank's equity (loans less deposits) must exceed a pre-specified fraction χ of the value of its assets (loans). Equation (11) is the law of motion for the intensity of the firm's relationship with the bank, which the bank internalizes. Equation (12) imposes the relationship between loan demand, price, and relationship intensity implied by (5). Finally, equation (13) shows how the bank's net worth evolves as a function of its lending and financing policies.

We can establish the following result about the bank's problem:

Proposition 2. ***(Optimal lending policies)*** *Is $\psi(e)$ is twice continuously differentiable, banks' optimal loan prices satisfy the Euler equation*

$$\frac{\Pi_t + \overline{q}\pi\rho_q \mathbb{E}_t\left[\sum_{i=0}^{\infty}\left(\overline{q}\pi(\rho_q+\rho_s)\right)^i \frac{L_{t+2+i}}{L_{t+1}}\Pi_{t+1+i}\right]}{\frac{\overline{q}}{q_t}\pi\mathbb{E}\left[\psi'(e_{t+1})\right]} = \epsilon^{-1}(q\ell, q) \tag{14}$$

where Π_t is the bank's net rate of return in period t per unit of loan and $\epsilon^{-1}(q\ell, q)$ is the inverse price elasticity of loan demand, given respectively by

$$\Pi_t = \frac{\overline{q}}{q_t}\pi\mathbb{E}_t\left[\psi'(e_{t+1})\right] - \psi'(e_t) + \lambda_t(1-\chi) \tag{15}$$

$$\epsilon^{-1}(q\ell, q) = \phi\frac{q}{\overline{q}}\frac{q\ell}{L'} \tag{16}$$

and where $\lambda_t = \psi'(e_t) - \frac{\overline{q}}{\overline{q}^d}\pi\mathbb{E}_t\left[\psi'(e_{t+1})\right] \geq 0$ is the Lagrange multiplier on the capital requirement in period t.

Equation (14) has an intuitive interpretation. The left hand side represents the sum of the bank's discounted marginal net profits associated with increasing its loan price. The choice of loan price today affects not only today's profits (Π_t), but also profits in all future periods (summation term).[8] The weight on this second term increases with the loading on current period lending in the law of motion for relationship intensity ρ_q, since this indicates a stronger dynamic pricing effect. The effective discount rate for future profits is $\overline{q}\pi(\rho_q+\rho_s)$: the first two terms reflect the equilibrium discount factor and the probability of bank survival, while the latter term reflects the *overall* persistence of relationships.[9] The profits in each period (15) reflect the return on loans, less their financing cost, plus the marginal benefit of easing the capital requirement.

This discounted profit stream in (14) must equal the inverse price elasticity of loan demand,

[7] Note that for the individual bank problem, expectations are taken with respect to the idiosyncratic shock z; hence why we make explicit the expectation with respect to μ in the firm problem.

[8] Note that in the stationary equilibrium, $L_t = L'(\mu)$ for all t, and so the term $\frac{L_{t+2+i}}{L_{t+1}}$ cancels out.

[9] If $\rho_q + \rho_s = 1$, then there is no depreciation in relationships and this term gets its maximal weight.

$\epsilon^{-1}(q\ell', q)$, which measures the bank's effective market power. As shown in equation (16), this term is only positive due to the relationship adjustment costs ($\phi > 0$), and increases with the bank's relative loan share. It is instructive to consider two extreme cases. First, when the bank's discount factor is zero, expression (14) resembles the classical static monopolist pricing condition, where the optimal price is set so that the markup is equal to the inverse elasticity of demand. The same also holds if ρ_q is zero, i.e. if there is no dynamic effect of today's loan price choice on tomorrow's demand. Second, in the competitive limit as $\phi \to 0$, the price elasticity of loan demand becomes infinite, eliminating the term on the right hand side of (14). Moreover, in this limit case, there is no notion of relationships, which eliminates the second term on the left hand side. Thus, in the competitive case, we recover the standard pricing condition $\Pi_t = 0$.

Evolution of bank distribution Given a current distribution of banks over states $m(x)$, the mass of banks next period with a particular x' is

$$\begin{aligned} m'(x';\mu) \;=\; \pi\Bigg[\int \mathbf{1}\left[n(x') = g_\ell(x;\mu) - g_d(x;\mu), s(x') = \rho_q \frac{g_q(x;\mu)g_\ell(x;\mu)}{L'(\mu)} + \rho_s s(x)\right] \\ \times\Gamma(z(x), z(x'))\mathrm{d}m(x;\mu)\Bigg] + (1-\pi)\mathbf{1}\left[n(x') = 0, s(x') = 0\right]\overline{\Gamma}(z(x')) \end{aligned} \tag{17}$$

The term in brackets in equation (17) describes state transitions for incumbent banks. For these banks, we require that next period's net worth and relationship intensity be consistent with the policies chosen this period, and that the evolution of the idiosyncratic shocks be consistent with Γ. The second term captures entrant banks, who begin with no net worth, no lending relationships, and idiosyncratic shocks drawn from the ergodic distribution $\overline{\Gamma}(z)$ implied by $\Gamma(z, z')$

3.3 Definition of equilibrium

Definition 1. A **stationary recursive competitive equilibrium** consists of: (i) bank-specific and aggregate loan demand functions, $\ell(q, s; \mu)$ for all (q, s) and $L(\mu)$; (ii) bank policy functions $g_q(n, s, z; \mu)$ and $g_d(n, s, z; \mu)$; (iii) a stationary joint distribution of prices and relationships $\mu(q, s)$; and (iv) a stationary joint distribution of banks over idiosyncratic states $m(n, s, z; \mu)$ which satisfy:

1. **borrower optimality**: bank-specific and aggregate loan demand satisfy (5) and (6);

2. **bank optimality**: banks' optimal policy functions solve the bank problem (8) – (13);

3. **stationarity of bank distribution**: the distribution of banks over idiosyncratic states is a fixed point of the operator defined in (17); and

4. **consistency of distributions**: the joint distribution of prices and relationships is consistent with the bank state distribution and with banks' optimal policies:

$$\mu(q, s) = \int \mathbf{1}\left[q = g_q(n, s, z; \mu)\right] m(\mathrm{d}n, s, \mathrm{d}z) \text{ for all } q, s \tag{18}$$

3.4 Discussion of assumptions

Implementation of lending relationships Two key elements of the structure of lending relationships in our model which bear further comment. First, we assume the representative firm maintains relationships with all banks, and that costs come not only from relationship *formation*, but more generally from relationship *adjustment*.[10] This specification embodies two simple assumptions: (i) all else equal, borrowers want to borrow more from banks with whom they have stronger relationships; and (ii) firm-bank relationships strengthen through exposure. Our specification of adjustment costs in (2) and the evolution of relationships (11) are exactly consistent with these assumptions. Exposures – and therefore relationships – shift through time for two reasons in our model. First, idiosyncratic shocks render some banks financially constrained, which leads them to charge different prices and lend different amounts than other banks with the same s. Second, the exogenous exit of banks and replacement with new banks yields a natural "life cycle" structure. As banks optimally respond to their financial conditions, they may either build up or expend relationships as a form of "customer capital," as in Gourio and Rudanko (2014).

Second, we assume that the firm does not internalize the formation of lending relationships, while banks do. The former assumption is made purely for tractability, as it is of course reasonable to expect borrowers to respond to developments in their banks' financial conditions by altering exposures to these banks. While possible in principle to allow for the firm to internalize relationship formation, it would require costly iteration between the firm and bank problems in the solution algorithm. By contrast, the demand system in the current framework allows us to solve the model with sole focus on the heterogeneous bank block. The fact that banks do internalize relationship formation shapes their optimal pricing policies, as highlighted in Proposition 2. As will be shown in the quantitative analysis below, this has important implications for how banks respond to financial shocks at both the individual level and in the aggregate.

Specification of relationship adjustment costs We assume quadratic adjustment costs in loan shares in our baseline model. This specification is attractive for two primary reasons. First, it delivers a simple closed form for bank-specific loan demand (5). Not only does this facilitate computation (see Appendix B), but – more importantly – it also yields a simple structural equation which we can map to the data in order to obtain an empirical estimate of the critical relationship parameter ϕ (see Section 4.2.2). Second, it delivers a single sufficient statistic – the effective interest rate $\tilde{R}(\mu)$ from equation (7) – which summarizes the key economic forces driving *aggregate* outcomes in the model. As we show in Appendix A.3, though, the same central economic forces still hold under a more general specification of adjustment costs.

Macroeconomic models of customer capital typically feature constant elasticity of substitution (CES) preferences that feature the relationship intensity or level of customer capital as a preference shifter within the CES aggregator, e.g. Gilchrist et al. (2017). While feasible, a CES specification

[10]Of course, relationship formation is also costly in our model; the specification of adjustment costs in (2) implies that the firm incurs costs for borrowing any positive amount from a bank with no relationships ($s = 0$).

raises some issues in our framework. First there is a matter of interpretation, as what is being aggregated is not utility over consumption of goods and services but rather loan dollars. Second, the CES with customer capital as a preference shifter still features a constant price-elasticity of demand, which does not vary with the intensity of the relationships. We derive the demand system under CES preferences in Appendix A.4.

One way to address the second concern is to aggregate loans across banks using the more general Kimball (1995) aggregator, which allows for a price-elasticity of demand that varies both with price and relationship intensity. As we show in Appendix A.5, the main drawback of this specification is that the resulting bank-specific demand is no longer linear or log-linear and therefore not amenable to direct estimation, which is one of the main advantages of our framework.[11]

Credit risk In our model, all loans are risk-less. We abstract from borrower credit risk for two main reasons. First, most firms in the sample that we use to calibrate the model have very low default risk (the median 1-year probability of default in our sample is of 0.73%). Second, default risk would complicate the model substantially by making it harder to aggregate outcomes for the borrower across banks. This assumption is not innocuous, as default risk would significantly affect banks' pricing decisions, interacting with their own state-dependent discount factors arising from equity issuance costs. In particular, it has been shown that ongoing relationships between banks and firms may distort pricing incentives and generate instances of overlending or insurance provision by the bank to the firm (see, for example, Faria-e-Castro et al. (Forthcoming)). To account for the fact that the model does not feature credit risk, all of our estimation exercises either include explicit controls for default risk, or factors that subsume this risk (such as firm-time fixed effects).

Customer capital in bank liabilities Our model also abstracts from broader definitions of bank customer capital, particularly its accumulation on the liability side of the balance sheet through deposit relationships. For example, Drechsler et al. (2017) argue that imperfect competition in deposit markets is a key factor that modulates the transmission of monetary policy. Polo (2021) expands on this idea and develops a quantitative macroeconomic model where banks accumulate customer capital in deposit markets, showing that this amplifies monetary policy shocks. An interesting extension of our model would feature customer capital accumulation on both sides of the balance sheet, and how the two relate to each other (i.e., whether they are substitutes or complements).

4 Mapping the Model to the Data

We parameterize our model in three steps. First, we assign values externally (i.e. outside the solution of the model) to standard parameters in the macroeconomics and banking literature. Second, we directly estimate our model's unique relationship lending parameters – the adjustment cost ϕ and the persistence parameters ρ_q and ρ_s – from the micro-data using a semi-structural approach. Third,

[11] Additionally, the Kimball (1995) specification does not solve the conceptual issue of aggregation of dollar values.

we jointly estimate the remaining parameters so that the model's stationary equilibrium matches a series of relevant banking industry moments. We now describe each of these steps in detail. The full parameterization of the model is summarized in Table 3.

4.1 Externally set parameters

We set nine parameters externally. The risk-free quarterly discount price $\overline{q}$ implies an annualized risk free rate of $\overline{r}_{\text{ann}} = 2\%$, in line with recent macroeconomic data. We set the interest rate on deposits $\overline{q}^d$ to be consistent with this risk-free rate and an annualized liquidity premium of 17 bps (van Binsbergen et al., 2022). The capital requirement is $\chi = 8\%$, in line with current capital requirements for large bank holding companies in the US. Since all exit is exogenous in the model, we set the bank exit rate equal to the historical average quarterly bank exit frequency, $1-\pi = 0.72\%$. We set total returns to scale to be consistent with a profit share of 5%, a capital share of 0.4, and a labor share of 0.6. The user cost of capital is set to be consistent with an annual interest rate of 2% and depreciation rate of 7%. Finally, we normalize the wage rate $\overline{w}$ to imply a marginal factor cost of one and the steady state level of aggregate TFP $\overline{A} = 1$.

4.2 Directly estimated parameters

Our model features three parameters that are not standard in models of banking and financial frictions: the cost of adjusting relationships, ϕ, and the parameters governing the persistence of relationships at the bank-level, ρ_q and ρ_s. We directly estimate these parameters on micro data using the relevant model demand equations. In particular, we use loan-level data from the Federal Reserve's FR Y-14Q data to estimate the equation for bank-specific loan demand in (5) with an instrumental variables approach and obtain an estimate for ϕ. We then infer series for bank-level relationships by aggregating the residuals of that estimated equation at the bank level, and use these series to obtain estimates for ρ_q and ρ_s. We now describe the data and the procedure in detail.

4.2.1 Data and sample selection

We use the cleaned Y-14 loan-level dataset described in Section 2 as the starting point to construct a "relationship panel" at the firm-BHC-quarter level, where the quantity of credit ℓ_{fbt} is defined as the total value of loans outstanding of firm f owed to BHC b at quarter t, and the interest rate r_{fbt} is the average rate on those loans, weighted by utilized loan value.[12] After all sample restrictions, our final panel runs from 2013Q1 to 2022Q2 and includes 3.361 million observations, for 242,568 distinct firms and 41 distinct BHCs.

[12] We also impose some additional restrictions that are aimed at eliminating observations that are likely errors: we drop all loans with interest rates equal to zero or above 50%, as well as loans for which the size of the commitment is non-positive, or the utilized quantity is larger than the commitment.

4.2.2 Estimating adjustment costs using bank-level loan demand

To estimate ϕ, we take advantage of the fact that this parameter appears in the bank-specific demand curve (5). Given data on loan quantities and interest rates, we treat the unobservable relationship intensity as a residual, and estimate this equation using linear regression. In particular, we estimate a specification of the following type:

$$\frac{\ell_{fbt}}{L_{ft}} = \beta(r_{fbt} - r_{ft}) + \alpha_{ft} + \alpha_b + u_{fbt} \tag{19}$$

where $L_{ft} \equiv \sum_b \ell_{fbt}$ is total borrowing by a particular firm across all banks, and $r_{ft} \equiv \sum_b \frac{\ell_{fbt}}{L_{ft}} r_{fbt}$ is the average interest rate paid by a particular firm across all banks, weighted by the borrowing amount. The goal therefore is to regress the loan share of each bank within a given firm on the spread between the interest rate charged by the bank to that firm and the average rate paid by the firm. We include firm-time fixed effects α_{ft} to capture fluctuations in overall credit demand by the firm that are unrelated to the characteristics of its relationship with different banks, and we use bank fixed effects α_b to control for time-invariant bank-level characteristics, such as different business models. An estimate $\hat{\phi}$ can then be retrieved from the estimated slope coefficient, given an externally set value for $\overline{q}$, i.e. $\hat{\phi} = -\overline{q}/\hat{\beta}$.

The main challenge to estimating equation (19) directly is that it is a demand curve, and thus OLS estimates suffer from the classical problem of simultaneity bias. We address this issue by constructing an instrument for bank-specific credit supply shocks following Amiti and Weinstein (2018). Specifically, we first estimate the following regression

$$r_{fbt} - r_{ft} = \gamma_{ft} + \gamma_{bt} + v_{fbt} \tag{20}$$

where γ_{ft} is a firm-time fixed effect and γ_{bt} is a bank-time fixed effect. The idea is that the firm-time fixed effect controls for any factor that is related to the demand for credit, while the bank-time fixed effect absorbs all variation that is related to the supply of credit, and is by construction orthogonal to demand. We therefore use $\hat{\gamma}_{bt}$ as a valid instrument for the credit spread $r_{fbt} - r_{ft}$ in (19).

Estimation results are reported in Table 1. The relatively low number of observations (compared to the size of the full sample) is due to two factors: first, our identification strategy relies on multi-bank firms, that is, firms that borrow from multiple banks, while the vast majority of firms in our data borrow from one bank only. We elaborate on and address this issue below. Second, we estimate (19) only on loans originated in the last 4 quarters. The model features one-period debt, and so the firm can effectively adjust its demand for debt across banks every period. In reality, firms borrow at many different maturities, and it is not clear that a firm will find it advantageous or even feasible to constantly prepay debt that was contracted in the past. Older loans are likely to be priced at rates that no longer reflect aggregate credit market conditions, and so including them could bias our estimates in the direction of estimating a lower credit demand elasticity.

The first column reports the simple OLS results, while column (2) reports the estimation results

	(1)	(2)	(3)	(4)
$r_{fbt} - r_{ft}$	-13.850***	-29.976***	-11.924***	-25.346***
	(4.015)	(3.694)	(1.651)	(7.851)
Firm identifier	TIN	TIN	ISL cell	ISL cell
Observations	57,833	57,731	221,674	221,637
Firm-Quarter FE	✓	✓	✓	✓
Bank FE	✓	✓	✓	✓
Model	OLS	IV	OLS	IV

Standard errors in parentheses, clustered at the BHC level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1: **Estimating ϕ using firm- and cell-level data**

using the credit supply shock instrument. Both specifications include the full set of fixed effects. In the instrumental variables specification the point estimate $\hat{\beta} = -29.976$ implies $\hat{\phi} = 0.0332$.

One issue with our estimation method that also applies more broadly to the identification approach of Amiti and Weinstein (2018) is that it relies on firms that borrow from multiple banks in order to isolate demand from supply effects. It is well documented across space and time that the vast majority of firms borrow from a single lender; in our sample over 80% of firms maintain a relationship with a single lender. This means that our reliance on multi-lender firms for identification precludes the use of the majority of our data. Degryse et al. (2019) address this issue by defining borrowers at the industry-size-location level, instead of at the firm level. The identification assumption is that the demand for credit should be relatively stable among firms of the same industry, size, and location (I, S, and L). We apply their methodology and define "ISL cells" where industry is the 3-digit NAICS, location is the CBSA of the borrower's address, and size is the borrower's decile in terms of total assets. This generates a total of 82,377 unique cells in our sample.

The results for this alternative estimation procedure are reported in columns (3) and (4) of Table 1. This sample – now almost four times larger – yields a slightly larger coefficient (in absolute value): the IV estimate for the slope parameter is $\hat{\beta} = -25.346$, which implies $\hat{\phi} = 0.0393$.

4.2.3 Estimating the law of motion for relationships

Our strategy to estimate ρ_s and ρ_q, the coefficients of the law of motion for relationship intensity, follows from the estimation of ϕ. Recall that we treat the relationship intensity term in the bank-specific demand s_{fbt} as a residual when estimating (19). The idea is to treat that residual as a measure of relationship intensity and directly estimate the law of motion in (11) using OLS. We are consistent with our procedure for estimating ϕ, and map the model to the data at the representative firm-level. That is, defining $\hat{\alpha}_b + \hat{u}_{fbt} \equiv s_{fbt}$, we use the residuals from (19) to directly estimate:

$$\hat{u}_{fbt} = \alpha_t + \alpha_b + \alpha_f + \rho_q \frac{\ell_{fbt}}{L_{ft}} + \rho_s \hat{u}_{fbt-1} + \nu_{fbt}$$

	(1)	(2)
$\frac{\ell_{fbt}}{L_{ft}}$	0.773***	0.791***
	(0.012)	(0.005)
s_{fbt-1}	0.176***	0.141***
	(0.011)	(0.004)
Firm identifier	TIN	ISL cell
Observations	36,694	134,274
R-squared	0.91	0.89
Quarter FE	✓	✓
Bank FE	✓	✓
Firm FE	✓	✓

Bootstrapped standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: **Estimating ρ using firm- and cell-level data**

where $\alpha_t, \alpha_b, \alpha_f$ are time-, bank-, and firm- fixed effects, respectively. The results for the estimation are reported in column (1) of Table 2, while the results for the estimation using industry-size-location cells are reported in column (2). We obtain $\hat{\rho}_s = 0.178, \hat{\rho}_q = 0.771$ when using firm-level data and $\hat{\rho}_s = 0.141, \hat{\rho}_q = 0.791$ when using ISL cells. The standard errors reported in these tables are bootstrapped, to correct for the fact that these specifications include generated regressors.

4.3 Jointly estimated

4.3.1 Parameters and targets

The remaining parameters are jointly estimated so that the model matches a series of targets from the data, given the externally set and directly estimated parameters. These parameters and target moments are summarized in Panel C of Table 3. The working capital parameter κ determines the level of overall loan demand given the firm's production parameters and output. Therefore, this parameter is informed by the level of business debt relative to total output, which averages 71.5% for the U.S. economy. The intensity of loan demand also helps determine how much market power banks have. Since this is a key element of our model, we also target banks' average net interest margin. Because our model has no default, and risk premia are an important component of loan spreads and therefore of net interest margins, we target a "no-default" net interest margin which filters out default risk premia from loan spreads. To construct this measure, we use Y-14 data to regress interest rates on new loans (originated in the last 4 quarters) on the originating bank's reported estimate for the 1-year probability of default. We then sum the constant and the residual, and call this the "zero-default interest rate". We compute the average zero-default interest rate for each bank in our sample, weighted by loan size, and subtract the average interest expense on deposits computed from the Call Reports. This measure averages 1.8% in our sample period.

	Description	Value	Target / Reason	Data	Model
Panel A: Externally Assigned Parameters					
$\overline{r}_{\text{ann}}$	Annualized risk-free rate	2%	Quarterly discount price $\overline{q} = (1+\overline{r}_{\text{ann}})^{-\frac{1}{4}}$		
ν_{ann}	Deposit liquidity premium	0.17%	Quarterly deposit price $\overline{q}^d = (1+\overline{r}_{\text{ann}} - \nu_{\text{ann}})^{-\frac{1}{4}}$		
χ	Capital requirement	8%	Current US bank regulation		
π	Bank survival rate	0.9928	Quarterly bank exit rate of 0.72%		
α	Capital share	0.38	Profit share of 5%, capital share of 0.4		
η	Labor share	0.57	Profit share of 5%, labor share of 0.6		
$\overline{w}$	Wage rate	4.41	Normalization		
$\overline{uc}$	Ann. user cost of capital	9%	2% interest plus 7% depreciation rate		
$\overline{A}$	Aggregate TFP	1	Normalization		
Panel B: Directly Estimated Parameters					
ϕ	Lending share adj. costs	0.0362	Average of estimates, Section 4.2.2		
ρ_q	Mkt. share impact on rels.	0.782	Average of estimates, Section 4.2.3		
ρ_s	Persistence, relationships	0.159	Average of estimates, Section 4.2.3		
Panel C: Internally Calibrated Parameters					
κ	Working capital constraint	0.755	Business debt to GDP ratio	71.5%	71.6%
$\overline{\psi}$	Equity issuance cost curvature	0.11	Gross equity issuance / NW	1.1%	1.1%
ρ_z	Persistence of net worth shocks	0.262	Net dividend payouts / NW	5.8%	3.7%
σ_z	Variance of net worth shocks	0.00264	Average net interest margin	1.8%	1.5%
			Average bank leverage	92.0%	91.5%

Table 3: **Summary of calibration**

Notes: Firm leverage and business debt to GDP are sourced from the Flow of Funds. The leverage moment corresponds to corporate firms. Gross equity issuance and net dividend payout rates are computed following Baron (2020). The net interest margin is computed using Y-14Q interest rate on new loans (originated in the last four quarters), residualized from firm 1-year probability default, and deposit expense data from the Call Reports. All moments are averaged between 2009Q1 and 2020Q3.

The remaining three parameters describe the idiosyncratic shocks to bank net worth and the costs of equity financing. We assume that the shocks to bank net worth follow an AR(1) process with mean $\overline{z} = 0$, persistence ρ_z, and standard deviation of innovations σ_z.[13] We model smooth but convex costs of issuing equity by using a piece-wise linear cost function

$$\psi(e) = \begin{cases} e(1+\overline{\psi}) & \text{if } e < 0 \\ e & \text{if } e \geq 0 \end{cases}$$

The parameters ρ_z, σ_z, and $\overline{\psi}$, then, are closely related to the financing choices banks make, and so we discipline them with moments of the data describing these choices. Given the costs of issuing equity and the relative cheapness of deposits, banks generally prefer to finance using deposits, and so our model replicates the high average leverage of 92% we observe in the banking sector. Since the

[13]That is, we assume $z' = \rho_z z + (1-\rho_z)\overline{z} + \varepsilon_z$, where $\varepsilon_z \sim \mathcal{N}(0, \sigma_z)$. This shock process is discretized over a grid of size $N_z = 21$ using the Adda-Cooper method.

capital requirement is close to binding for most banks, then, they generally respond to idiosyncratic shocks either by retaining earnings or issuing new equity. We ensure realistic behavior along the former dimension by targeting the net dividend payout rate of 5.8%, and the latter by targeting the gross equity issuance rate of 1.1% (both scaled by total net worth).

4.3.2 Solving the model

Since internally calibrating the parameters described above requires iteratively solving the model for a range of potential parameter values, and since computing our model requires several non-standard steps, we describe our solution algorithm at a high level before proceeding. Appendix B contains a more detailed, formal description of our computational algorithm.

The main complication in solving for a stationary equilibrium is that equilibrium is described not by a small vector of aggregate prices, but by the entire joint distribution of prices and relationship. However, given guesses of the distribution of banks over idiosyncratic states, $m(x)$, and bank pricing policy functions, $g_q(x)$, we can use the consistency condition (18) to infer the implied joint distribution of prices and relationships $\mu(q, s)$. Given this distribution, we can compute the demand-relevant summary statistics $R(\mu)$ and $\tilde{R}(\mu)$, which are the necessary inputs to bank-specific and aggregate loan demand according to equations (5) and (6). Finally, we can use these implied demand curves to solve for updates of banks' optimal policies, which in turn deliver an implied update to the initial guess of the distribution of banks. This procedure can be repeated until convergence on both policy functions and the distribution in order to obtain a stationary equilibrium.

5 Model Mechanics and the Role of Relationships

In this section, we use the cross-section of our baseline economy and several variants to explain the key mechanisms in our model of relationship lending. The analysis in this section provides the underpinnings for understanding how relationship lending alters aggregate dynamics, which is the focus of the next section.

Model variants Throughout our quantitative analysis, we focus on two main versions of the model: (i) the **baseline**, whose calibration was described in the previous section; (ii) a **competitive** version of the model where banks take market interest rates as given and choose how much to lend. Details of this second economy are presented in Appendix A.6. In this model, the lack of adjustment costs in the borrower's problem removes any meaningful notion of relationships. This implies a single equilibrium lending rate is taken as given by all banks. Banks then choose ℓ' directly, and a bank's state is fully described by (n, z). The competitive version of the model feature banks with much lower *static* market power since they face an infinite price elasticity of loan demand at each date.

In order to understand the economic forces at play, we also report results for two other variants of the model: (iii) a "low elasticity" version, where the loan share adjustment cost ϕ is greater than in the baseline, and so the elasticity of demand with respect to the spread is lower and banks have more market power; and (iv) a **"low punishment"** version in which $\rho_q \to 0$ so that banks face the same static loan demand elasticity implied by the estimated ϕ, but do not sever relationships by increasing prices as much as in the baseline model.[14] Table 4 summarizes key cross-sectional statistics across these different specifications of the model.

5.1 How do lending relationships shape industry-level and aggregate outcomes?

Loan rates decrease as competition increases. Panel A of Table 4 presents statistics on interest rates and loan quantities. At the highest level, this panel confirms unsurprising results about competition in our model environment. The effective interest rate $\tilde{R}$ varies sharply with the degree of competition, dropping around 34.4% in the competitive version of the model. Lower effective interest rates raise loan volumes by 4.3% relative to the baseline model. Raising banks' static market power in the low elasticity economy increases effective interest rates by 37.4%. When banks' static market power remains unchanged but becomes less sensitive to pricing decisions in the low punishment economy, effective interest rates increase by 15.6%. These movements along the demand schedule result in commensurate decreases in total loan volumes in these two economies.

The next three rows decompose these differences using the three components of $\tilde{R}$ from equation (7). The bulk of the difference arises from higher average interest rates: banks in the baseline economy exercise their market power by charging higher rates across the board. The positive covariance between relationship intensity and interest rates behaves similarly, but with a smaller magnitude: banks with stronger relationships can afford to charge higher rates for a given loan volume.[15] The covariance term is notably large (49 bps, or 12.8% of the total effective interest rate) for the low punishment model: in this case, not only are banks with stronger relationships inclined to charge higher rates, but the greater persistence in relationships induces a more unequal bank distribution. Finally, there is a small attenuation effect arising from interest rate dispersion: greater variance in loan rates provides more scope for borrowers to substitute into cheaper borrowing. This effect, however, is quantitatively small across model specifications.

Relationships compress the distribution of bank net worth. Panel B of Table 4 focuses on moments related to the distribution of financial capital (net worth) and relationships in the banking industry. Net worth is both slightly higher on average and less dispersed in the baseline economy than in the competitive economy. Figure 3 plots the partial distributions of net worth and relationships, and the joint distribution of net worth and relationships in the baseline model. Note that the latter two objects are not defined for the competitive economy. In particular, Figure 3(a) shows that this compression of the net worth distribution combines two main forces. First, bank market power generates a lower optimal lending scale: banks ration quantities to keep markups

[14] For completeness, the low elasticity version of the model has $\phi = 0.0724$, double the baseline value for this parameter, and the low punishment version of the model has $\hat{\rho}_q = 0.078$, one-tenth the baseline value. When ρ_q is changed to $\hat{\rho}_q$, ρ_s is also changed to $\hat{\rho}_s$ so that S has the same value as in the baseline economy; that is, $\hat{\rho}_s = \frac{S - \pi\hat{\rho}_q}{\pi S}$, and reducing ρ_q implies increasing ρ_s.

[15] Of course, this effect is absent in the perfectly competitive economy, which has no notion of relationships.

		level				% diff rel to baseline		
		baseline (i)	comp. (ii)	low elas. (iii)	low pun. (iv)	comp. (ii)	low elas. (iii)	low pun. (iv)
Panel A: pricing and lending								
effective IR (pp, ann.)	$\tilde{R}(\mu)$	3.29	2.16	4.52	3.81	-34.36	37.43	15.64
= average rate	$R(\mu)$	3.26	2.16	4.44	3.36	-34.36	36.36	3.20
+ covariance term	$\mathbb{C}_\mu(r, s)$	0.05	-	0.10	0.49	-	102.1	910.8
+ variance term	$\mathbb{V}_\mu(r)$	-0.01	-	-0.02	-0.05	-	-23.14	-280.8
loan-weighted avg. IR		3.28	2.15	4.51	3.76	-34.27	37.51	14.56
loan volume	$L'(\mu)$	0.26	0.27	0.25	0.25	4.25	-4.38	-1.86
Panel B: banking industry moments								
average net worth		0.023	0.022	0.022	0.023	-4.84	-1.50	0.72
std dev, net worth		0.005	0.010	0.004	0.008	105.0	-10.13	63.55
std dev, relationships		0.143	-	0.128	0.412	-	-10.59	187.0
share of switches (pp)		1.34	4.15	0.86	2.96	209.4	-36.00	121.0
corr, net worth and relationships		0.795	-	0.765	0.894	-	-3.83	12.42
corr, net worth and spread		0.002	-	0.068	0.306	-	3286	15192
corr, relationships and spread		0.123	-	0.191	0.391	-	55.90	218.5

Table 4: **Cross-sectional and aggregate results across model variants**

Notes: In Panel A, all pricing moments are expressed in annualized net percentage points. In Panel B, all net worth objects are computed using total beginning-of-period net worth, $n + z$. See Appendix A.7 for a description of the “share of switches” metric.

high, as is standard in models of imperfect competition. Thus banks tend to cluster at a *higher* level of net worth in the more competitive economies, i.e. the biggest concentration of bank mass in the competitive model is to the right of its analog for the baseline model. Counteracting this first effect, though, is the fact that it takes longer for banks to accumulate net worth in the competitive economy. Since each unit of lending is less profitable, the “life cycle” of net worth tends to have a much flatter profile in the more competitive economies (shown in Panel (c) of Figure 6 below and discussed later in this section). Correspondingly, there are far more banks with net worth below the “long run” level in competitive economy.

The low static elasticity economy features slightly lower average, even more compressed (lower standard deviation) net worth than the baseline. This show that the effects of competition on the overall level of net worth are not monotonic. This is because there are two main effects at play: on one hand, higher profitability per unit of lending induces higher net worth. On the other hand, more profitable lending means banks are better able to smooth dividends in the face of idiosyncratic uncertainty. Better insurance allows them to operate with lower levels of net worth, and this latter

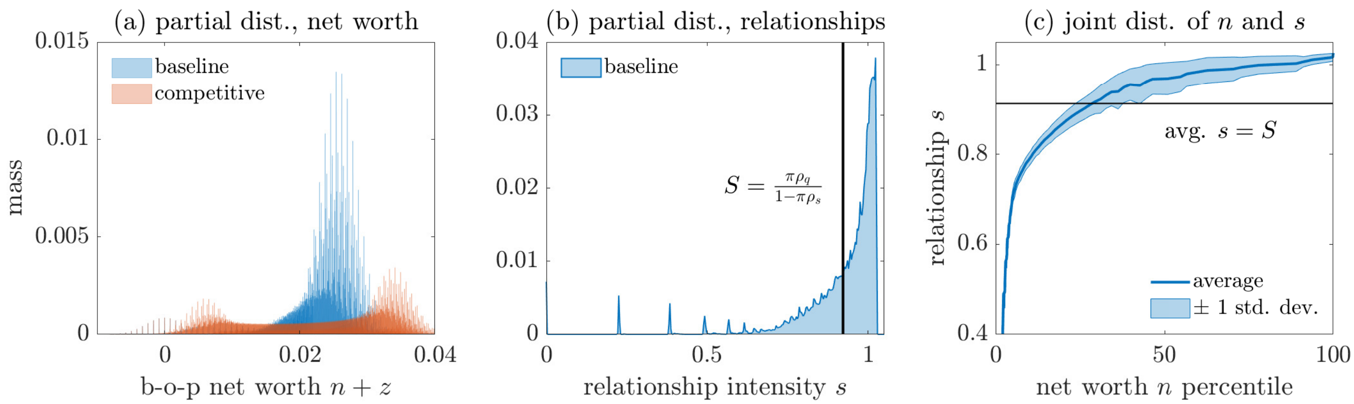


Figure 3: **Distributions of the endogenous state variables across models**

Notes: Panels (a) and (b) plot the partial cumulative distribution functions for total beginning-of-period net worth ($n + z$) and relationships (s), respectively. Average net worth for each model is presented in Table 4. The net worth distributions for the competitive and nearly competitive models effectively lie on top of each other. The average relationship intensity, represented by the black line in panel (b), is common across models by construction. Panel (c) depicts the joint distribution of net worth and relationships for the baseline model by plotting the average relationship intensity at each percentile of the net worth distribution, plus or minus one standard deviation of the mean at that net worth percentile.

effect dominates as ϕ becomes large enough. Varying the degree of *dynamic* market power instead, as in the low punishment economy, yields slightly larger dispersion, reversing which of these two potentially offsetting effects dominates. On the one hand, greater persistence of lending relationships means that banks have incentives to accumulate a lot of net worth to lend large quantities at high spreads. On the other hand, the dominance of the banks with strong relationships in this model makes it very hard for smaller banks to compete and grow, and so there is a long left tail of smaller banks in this case. The first effect dominates, though, and net worth is 0.72% higher on average in the low punishment model, relative to the baseline economy.

Table 4.B also contains measures of the extent to which the borrower switches between lenders in the model. Since the firm borrows from all banks, we define "switching" to account for adjustment of the loan portfolio along the intensive margin as well as the extensive. That is, we define a switch as an instance where a firm's borrowing from a given bank in a given bank exceeds the firm's borrowing from that bank in the prior period. We then scale the total volume of switching loans by total loan volume to obtain the percentage metric reported in the table (see Appendix A.7 for the explicit formulation of this metric). The share of total lending from switches in the baseline model is 1.34%. The competitive economy features much more switching. Constrained lenders in this economy cut loan supply sharply because they cannot raise interest rates and rely on their relationships to prop up lending. Since aggregate lending remains constant in steady state, this implies greater switching. Unsurprisingly, the low elasticity economy features less switching than the baseline, as switching is now more costly. Interestingly, the low punishment economy features almost twice as much switching as the baseline economy. This stems from the fact that lenders with strong relationships charge higher rates, inducing more borrower substitution. Due to the increased

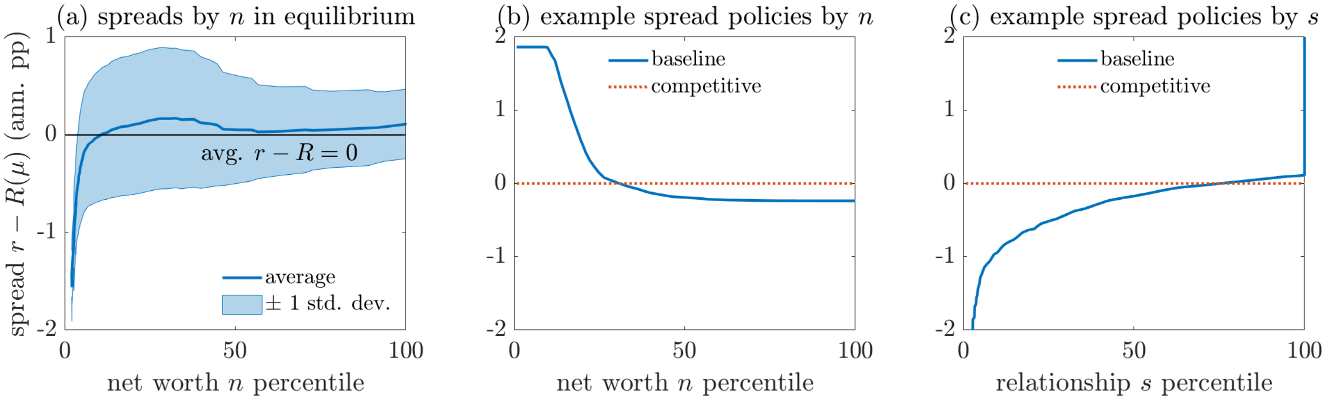


Figure 4: **Steady state loan pricing policies**

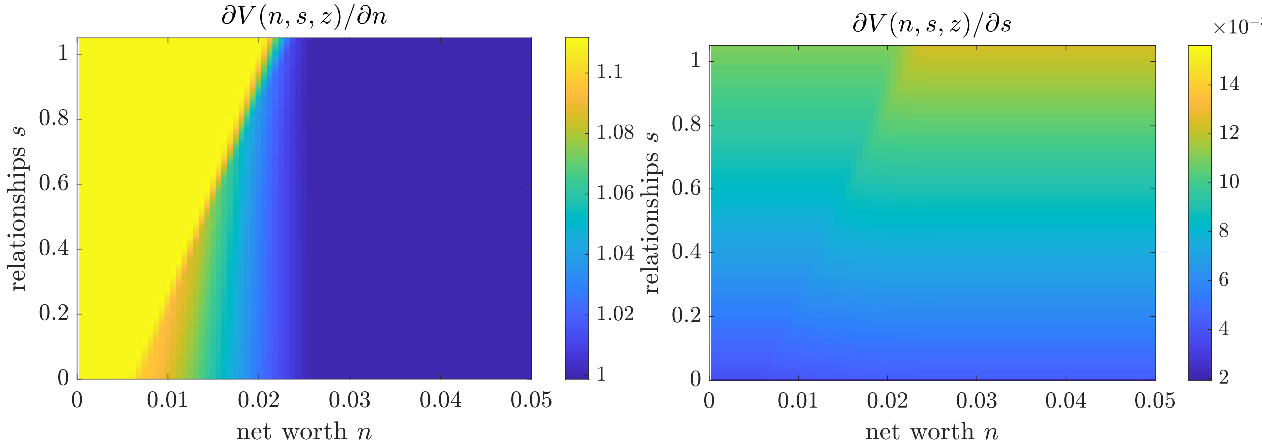
Notes: Panel (a) plots the first and second moments of the pricing policies (expressed as annualized percentage point spreads over the average interest rate) conditional on a given level of net worth over the equilibrium distribution of banks for the baseline model. Variation in pricing at each level of n comes from dispersion in s and z. Panels (b) and (c) plot sample pricing policy functions (expressed in the same units as in panel (a)) over the equilibrium distribution of net worth and relationships for the baseline and competitive economies. Percentiles for each line are from the respective equilibrium distribution for each model variant. Panels (b) and (c) each fix $z = 0$, the median level, and s or n at the median level from the baseline economy.

persistence of relationships, however, this increased substitution has only a small impact on banks' ability to maintain simultaneously high spreads and loan volumes.

Financial and relationship capital are complements. The last rows in Table 4.B present correlations between the key state variables and interest rate spreads. In each of the less competitive models, there is a strong positive correlation between net worth and relationships (79.5% in the baseline). This is highlighted further in Figure 3(c), which depicts the joint distribution of net worth and relationships in the baseline economy. Overwhelmingly, small banks in terms of financial capital have weaker relationships. While relationship strength increases across the entire net worth distribution, this rise is especially sharp over the bottom quartile of the distribution.

In our baseline model, the correlation between net worth and spreads is close to zero, while the correlation between relationships and spreads is positive. These results combine two effects. First, as shown in Figure 4(b), less capitalized banks charge higher spreads in order to escape their financial constraints by lending small amounts very profitably. In the face of costly equity issuance and a capital requirement which constrains deposit financing, banks with low net worth cut lending by raising rates. Second, banks with weak relationships tend to price extremely competitively – even below market, in the sense that the spread $r - R < 0$ – in order to build up relationships for the future, as shown in Figure 4(c). As relationships strengthen, banks can sustain lending above market interest rates.

In isolation, these forces would lead the correlation between spreads and net worth (relationships) to be negative (positive). These forces are tempered, however, by the strong positive correlation between net worth and relationships described above. That is, the banks who are financially con-



(a) **Marginal value of net worth** (b) **Marginal value of relationships**

Figure 5: **Complementarity of financial and relationship capital**

Notes: Panels (a) and (b) plot numerical approximations to the slope of the bank value function with respect to net worth and relationships, respectively, for z fixed to the median level in the baseline model. The ranges for n and s are chosen to be those with positive mass of banks according to the distributions in Figure 3.

strained and would like to charge high spreads tend to be the very same banks who have weak relationships and therefore would like to charge low spreads. These two effects roughly offset each other across the distribution, as highlighted in Figure 4(a), which plots the joint distribution of net worth and spreads in the baseline model.

How do changes in the competitive landscape alter these effects? With more market power (static or dynamic), constrained banks are able to charge higher interest rates while sustaining similar levels of lending. This naturally strenghtens the correlation between relationships and spreads, but also strengthens the correlation between net worth and spreads.

Our model, then, sheds additional light on the nature of financial constraints in banking. Measuring banks' net worth provides information on banks' pricing and lending decisions, but the degree to which these policy functions are actually elastic with respect to net worth, and the levels of net worth at which this elasticity manifests, can vary considerably across relationship intensities and with the competitive landscape.

Ultimately, we find that financial and relationship capital are complements. These two types of capital are complements if more net worth delivers more value to a bank with stronger relationships. The most direct measure of this is the cross-partial of the value function, $\frac{\partial^2 V}{\partial n \partial s}$, which should be positive. Figure 5 provides two lenses into this across the range of n and s for our baseline model: panel (a) shows the marginal value of net worth, while panel (b) shows the marginal value of stronger relationships across the state space. Complementarity implies that the former is increasing in relationship intensity, while the latter is increasing in net worth; both are confirmed in the figure.

Relationships shape banks' life cycles. Figure 6 investigates the life cycle of bank beginning from entry across all versions of our model. In all cases, new banks start out with no net worth

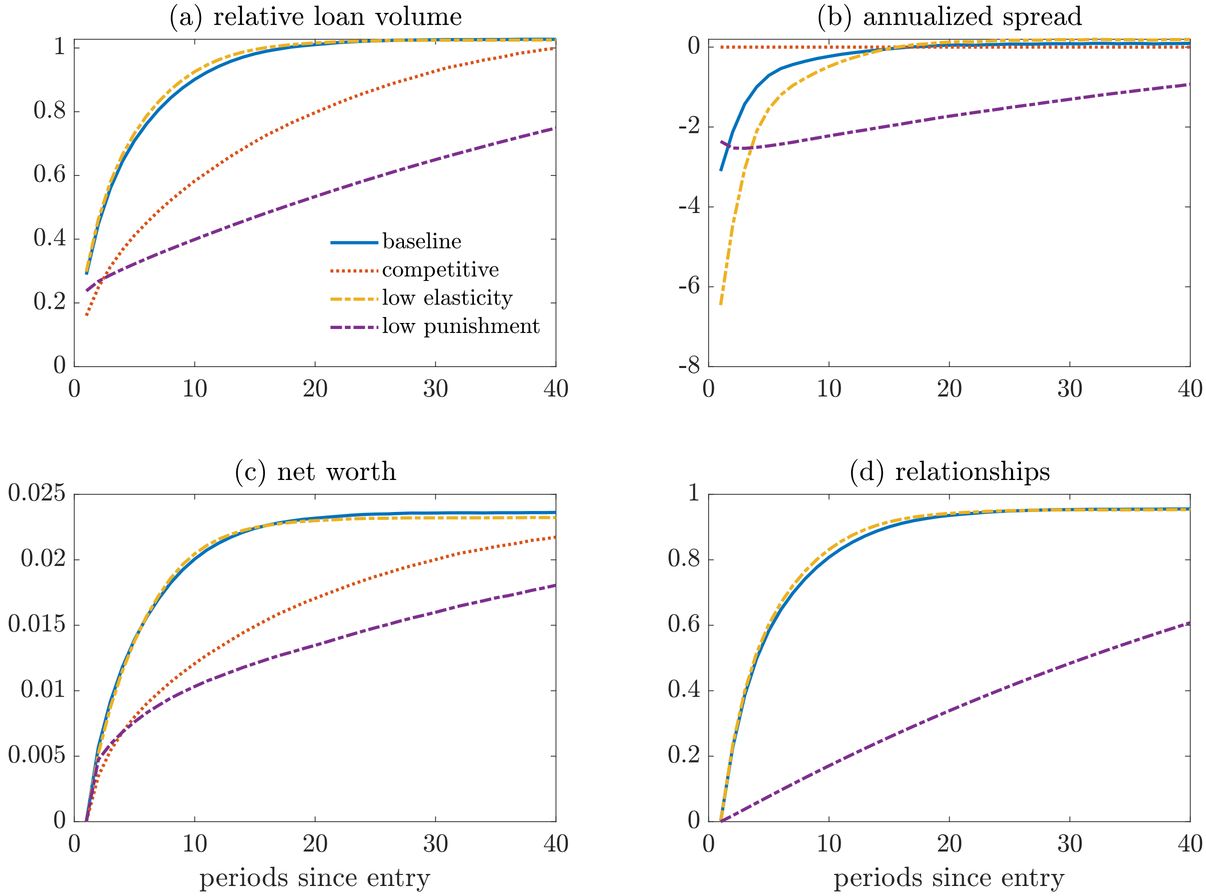


Figure 6: **Average bank life cycle across model variants**

Notes: For each version of the model, we simulate paths of $T = 40$ periods for $N = 10,000$ entrant banks with initial states $n = 0$, $s = 0$, and z drawn from the ergodic distribution $\overline{\Gamma}(z)$. Each plot presents averages of the indicated metrics across the N banks for each date $t = 1, ..., T$.

(panel (c)), which does not allow for much lending (panel (a)). In the less competitive baseline and low punishment economies, banks optimally price below market (panel (b)) in order to build relationships (panel (d)). Given the persistence of relationships in the low punishment case, the required period of pricing below market is both longer and less profitable (except in the very early dates). By contrast, in the competitive model, banks simply price at the market rate and increase loan volume gradually alongside net worth. Notably, the results in Figure 6 suggest that there is a "sweet spot" with regard to the accumulation of net worth: in the competitive economies, low profit margins across all banks make it hard for banks to accumulate financial capital, while in the less competitive low punishment economy the intensity with which small, weak relationship banks must compete against more established banks renders profits similarly low. As banks accumulate financial and relationship capital, they lend more while increasing their spreads.

5.2 Empirical validation: the evolution of spreads over a relationship

The key novelty in our framework is the link between persistent lending relationships and banks' pricing decisions. In order to have confidence in our model's predictions about how relationships affect aggregate outcomes, then, we must first provide evidence that our model delivers pricing implications given changes in relationships that are in line with the data.

Recall that Figure 2 shows the average difference in interest rates for loans from new banks ("switches") compared to loans from banks with existing relationships. The key insights from this figure are that: (i) switching loans are priced relatively favorably immediately upon and in the first year following the switch (negative relative spreads); and (ii) this pattern reverses in the second year after the switch (positive relative spreads).

Does our model deliver similar patterns? To address this question, we simulate a panel of banks drawn randomly from the stationary distribution and study how they behave after a share δ_s of their relationship capital s is destroyed. Holding the rest of the banks' states fixed, this reduction in relationship capital accounts for differences in lending practices solely attributable to differences in relationships, as in our empirical analysis. We choose δ_s to match the average initial drop in spreads immediately upon switching of 5.38 bps.[16]

Figure 7 plots the difference between switching and non-switching loan spreads for our baseline model and several variants, with the empirical relationship reproduced as the black line and shaded region. Our baseline model matches the dynamics of spreads in the data quite closely along several dimensions. In both model and data, banks price below market upon switching, but then steadily charge slightly above-market spreads thereafter. Our model correctly captures – both qualitatively and in terms of magnitudes – the gradual increase in spreads after the "honeymoon phase" following the switch; over the entire three-year span depicted in Figure 7, the relative increase in hovers between 5 and 10 bps as in data.

When we repeat this exercise for the low elasticity and low punishment models, the comparison with the data is far less favorable. While both match the initial drop in spreads by construction, they fail to match the data in the subsequent periods in different ways. In the low elasticity model, banks leverage high switching costs to take immediate advantage of their newly captive borrower, charging much higher spreads than we observe in the data. Spreads remain persistently above what we observe in the data over the life of the relationship. In the low punishment case, the initial increase in spreads is lower, but the profile does not level off over the horizon of the figure. This is because need more time to accumulate sufficient relationship capital to viably increase rates without inducing substitution. Once it is high enough, charging higher spreads does not erode the relationship nearly as much as it does in our baseline model.

[16]It is perhaps natural to think of a true "switch" as the case $\delta_s = 100\%$, i.e. eliminating any past relationships, in this experiment. This implementation, however, is far too strong given our representative borrower structure: a bank with $s = 0$ has faces extremely small loan demand under equation (5), and so will price aggressively to build it up (recall the life cycle analysis above. Empirically, both the incumbent and switching lenders have a portfolio of borrowers, and so the drop in relationship intensity we implement must be more marginal.

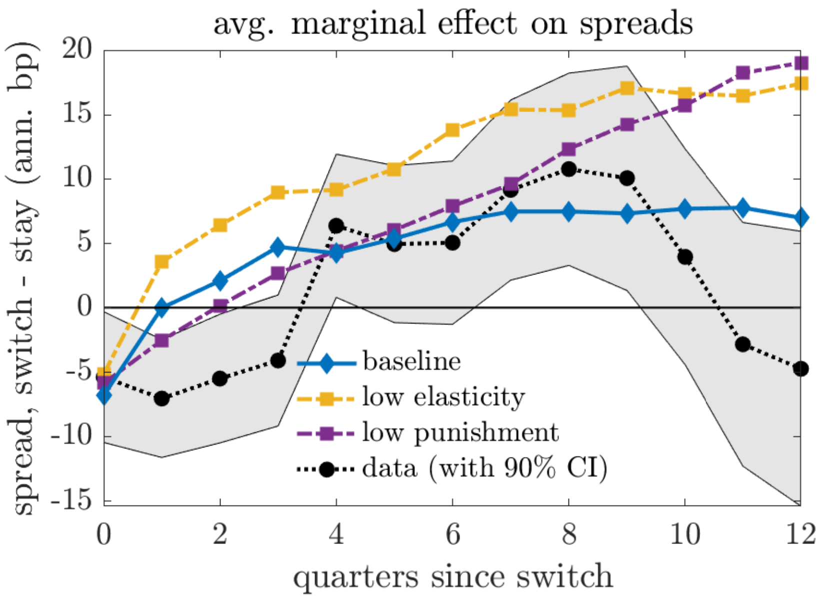


Figure 7: **Differences in spreads from switching relative to non-switching loans**

Notes: This figure plots the differences in spreads between switching and non-switching loans for several model variants and in the data. For detailed descriptions of the construction of this figure, see Section 2.2.2 for the data and Section 5.2 for the model.

6 Aggregate Dynamics

We now analyze how lending relationships shape how the economy responds to aggregate shocks. In this section, we consider three types of aggregate shocks: (i) a "financial crisis", where every bank experiences a proportional decline in their own net worth; (ii) a negative credit supply shock that affects the deposit funding cost of banks, $\bar{q}^d$, and (iii) a negative shock to loan demand. We assume that these are one-time, unanticipated shocks.

6.1 Net worth shock

Figure 8 plots the response of aggregate variables to a negative aggregate shock to bank net worth. We assume that the net worth of every bank unexpectedly declines by 5%: we motivate this shock as an unexpected loss in profitability arising from other business lines within the bank, such as mortgage lending and/or MBS holdings during the 2007-08 financial crisis. Panels (a) through (e) present the effects of the shock on the effective interest rate, total lending, total net worth, total deposits, and net dividend payouts, respectively, for the baseline and competitive economies.[17]

While qualitatively the responses are similar between the two economies (in terms of the signs), there are important quantitative differences between the two.

Panels (a) and (b) show that the increase in the effective interest rate and corresponding decrease

[17] Appendix C.4 contains the same results for the low elasticity and low punishment economies. These results are qualitatively similar to the baseline, for the most part.

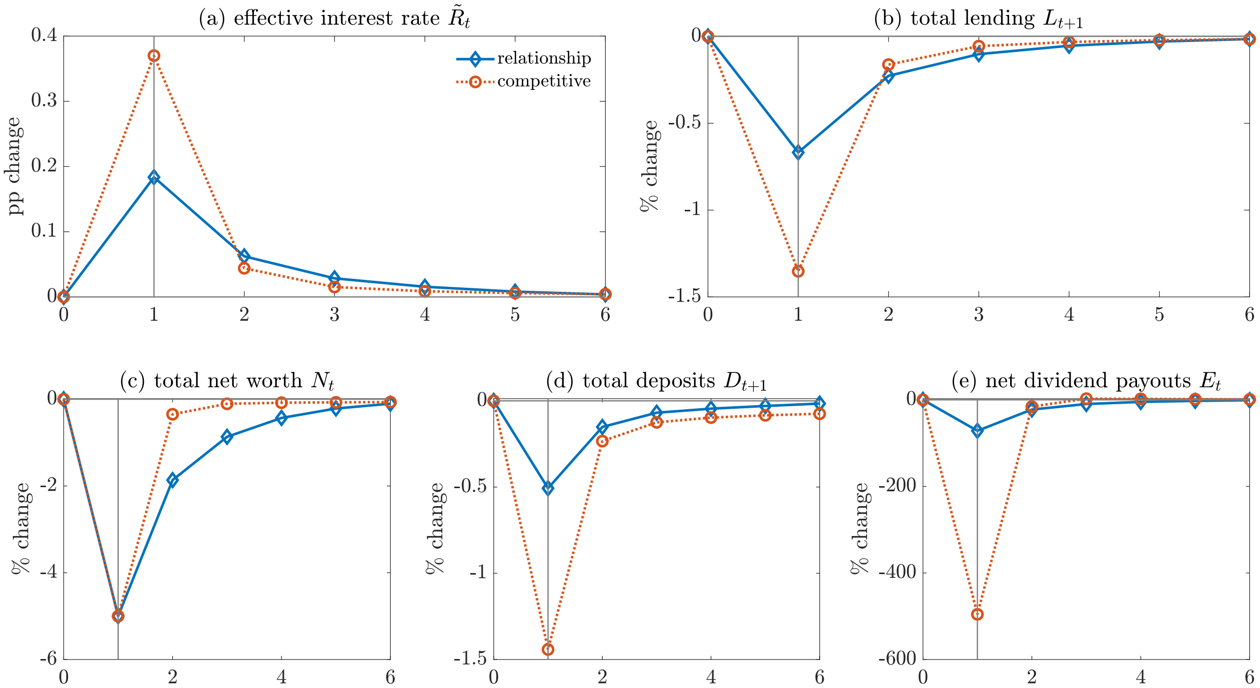


Figure 8: **Aggregate shock to bank net worth**

Notes: This figure plots the paths of key aggregate variables in units of percent deviations from steady state, or date 0 (except for the effective interest rate, which is units of percentage point differences). The aggregate shock considered reduces each bank's net worth by 5% at date 1.

in lending are larger on impact in the competitive economy, but also less persistent. Panel (c) shows that this has implications for the behavior of aggregate net worth: while the fall is quantitatively similar across models, the recovery is considerably slower for the baseline economy.

In the baseline economy, banks exploit the complementarity between customer and financial capital: faced with binding constraints, they expend customer capital and raise interest rates. Banks, however, internalize that interest rates cannot be raised substantially above what other banks are charging, as otherwise this will lead to a faster depletion of customer capital. Since interest rates rise by relatively less, lending also falls by relatively less, but this results in more protracted recapitalization dynamics.

In the competitive economy, the shock triggers a standard financial accelerator effect: banks become constrained and cut back lending, and interest rates rise substantially in equilibrium. This is the mechanism underlying "fast recapitalizations" and the lack of endogenous persistence in many macro-finance models with lenders that are subject to a collateral constraint. The significant rise in interest rates and/or drop in asset prices at the time of the shock generates large ex-post returns that allow intermediaries to recapitalize themselves extremely fast. The sharp decrease in lending on impact induces a sudden deleveraging in the competitive model that is promptly reversed, as shown in panel (d).

Since the recapitalization is slower in the baseline model, the effects of the shock on interest rates

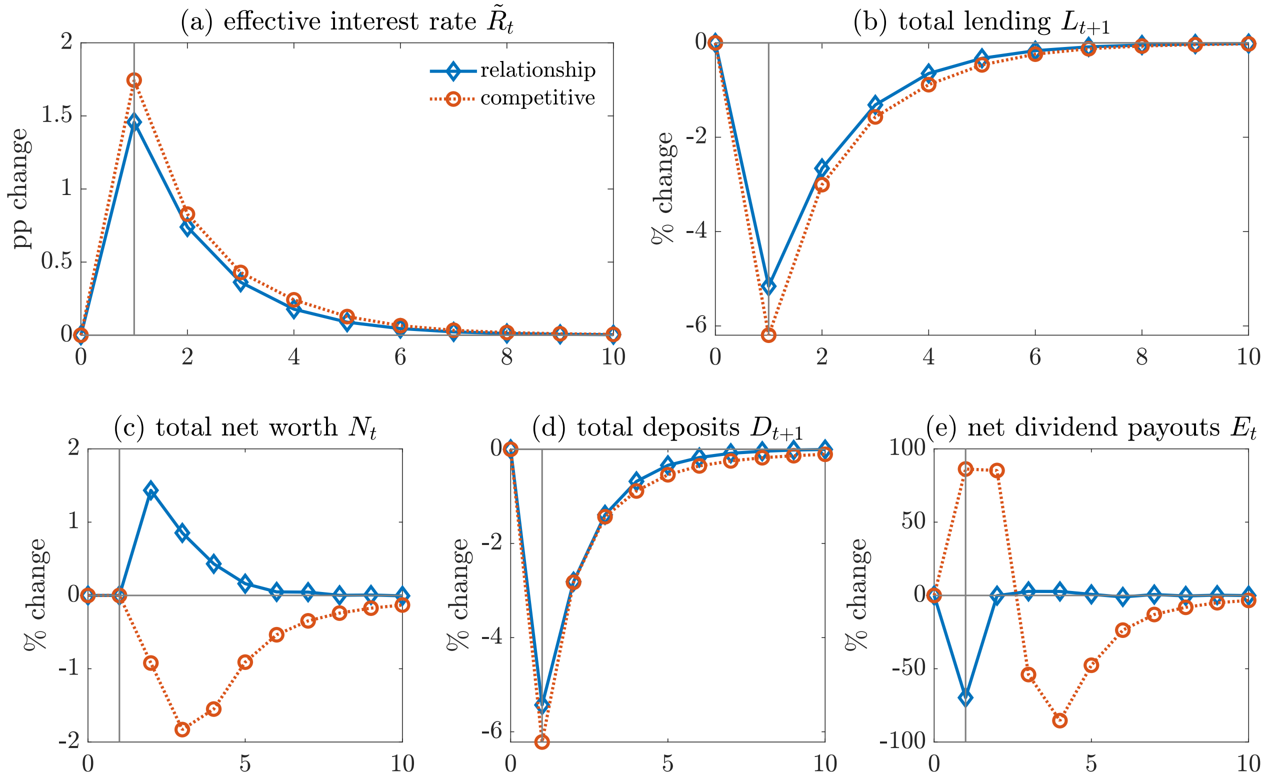


Figure 9: **Aggregate shock to cost of funding**

Notes: This figure plots the paths of key aggregate variables in units of percent deviations from steady state, or date 0 (except for the effective interest rate, which is units of percentage point differences). The aggregate shock considered raises the cost of funding from 2% to 4% at date 1, with the funding costs recovering with a persistence of 0.5.

and lending are also felt for longer, even if the response on impact is more moderate than that in the competitive lending economy.

6.2 Interest rate shock

Figure 9 considers a negative shock to $\bar{q}^d$, such that the bank cost of funding increases from 2% (annualized) to 4%. Unlike the one-time net worth shock, this is persistent with $\rho = 0.5$. This is one way of modeling a monetary policy shock and studying its transmission in the context of our model. In this case, both economies feature similar responses in terms of quantities and price of lending, with effective interest rates rising in both cases, and aggregate quantities of credit falling. Thus both models feature similar pass-through of funding costs to lending rates, and generate a standard monetary policy transmission mechanism to quantities of credit, which in turn directly affects real activity via the binding working capital constraint.

The two models generate very different responses of net worth, however, with bank net worth falling in the competitive model but rising in the relationship lending model. In the relationship lending model, banks do not want to contract their lending by as much so as to preserve relationship capital. However, one of the two sources of funding for lending has become more expensive, so lending falls by relatively less and banks switch to using relatively more retained earnings than deposits to fund that lending. This induces an increase in net worth.

In the competitive case, the mechanism is more standard: since relationship capital is not a concern, banks do not need to keep lending as elevated as in the baseline case, and so they simply deleverage, cutting both net worth and deposits (and paying out dividends, as shown in panel (e)). This is also why lending falls by relatively more than in the baseline, and correspondingly, interest rates rise by more.

This allows us to conclude that while the pass-through of monetary policy to credit market variables is similar in the two economies, the degree of competition in the banking industry can generate stark differences in how aggregate bank net worth reacts to a tightening of monetary policy, which has natural implications for financial stability.

6.3 Loan Demand Shock

Finally, Figure 10 plots the effects of a contraction in loan demand, which we model as a 1% drop in TFP with a persistence of 0.5. While total lending responds similarly across models, the effective interest rate dynamics are quite different. In the competitive model, the mechanism is similar to the one described above: faced with less loan demand, banks delever, reducing both net worth and deposits while paying out dividends. With lending relationships, banks actively try to maintain their relationship capital, which induces them to cut interest rates more. Consequently the quantity of lending falls relatively less. This cut in interest rates has implications for net worth, which falls less than in the competitive case. In the case of a credit demand shock, the degree of competition in the banking system has clear implications for the aggregate effects of shocks on variables such as interest rates and quantities of credit, as well as on the extent to which bank capitalization reacts to the shock, which in turn matters for financial stability considerations.

7 Conclusion and Directions for Future Research

This paper presents a quantifiable, estimable framework with which to evaluate the aggregate consequences of lending relationships. Our model environment combines standard features from the literature on heterogeneous banks subject to financial constraints with two novel elements: (i) loan sourcing adjustment costs on the part of borrowers; and (ii) a fully internalized law of motion for relationships on the part of banks. These novel elements yield a tractable but rich notion of relationships. Importantly, the way we specify our model simultaneously facilitates both direct estimation of the novel relationship parameters of interest and efficient computation, despite the richness of heterogeneity and financing choices within the banking sector.

Quantitatively, we present three primary results. First, we show that our baseline model matches the profile of interest rate spreads over the life of a bank-borrower relationship that we observe in the data. Notably, our model appears to get both the static and dynamic components of the market

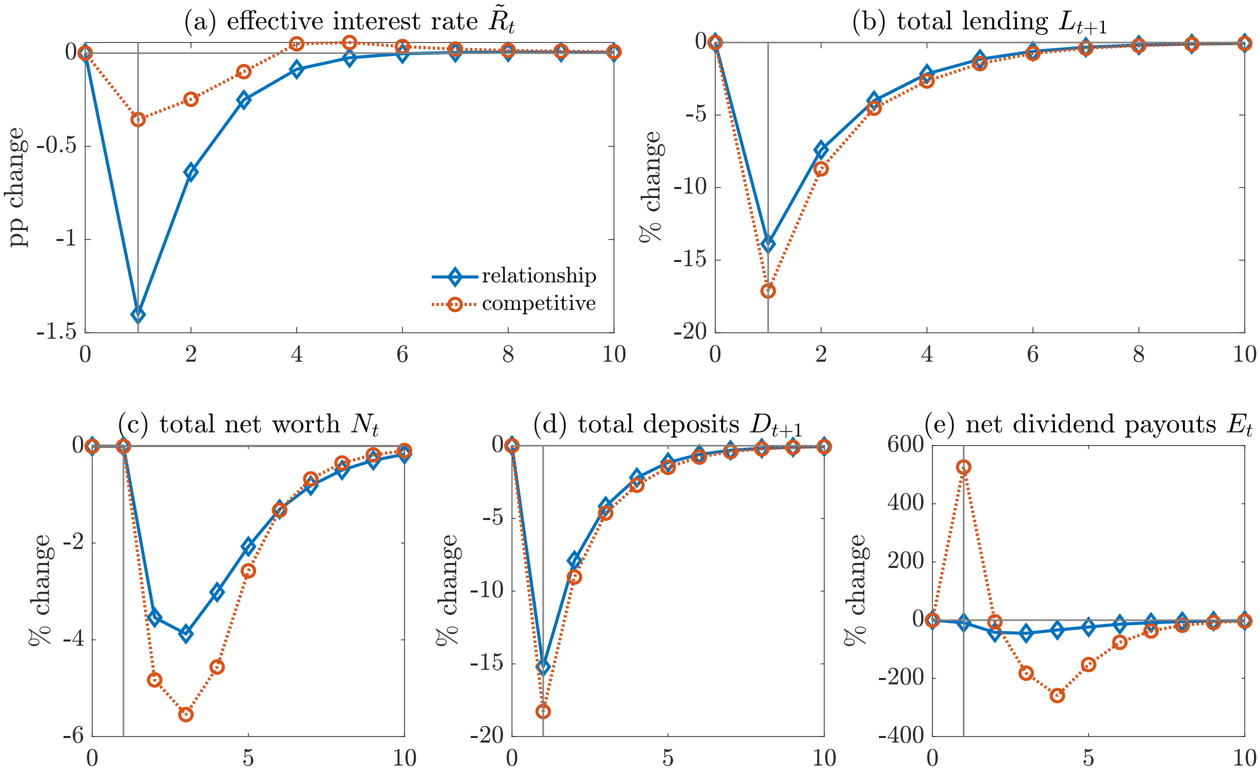


Figure 10: **Aggregate shock to loan demand**

Notes: This figure plots the paths of key aggregate variables in units of percent deviations from steady state, or date 0 (except for the effective interest rate, which is units of percentage point differences). The aggregate shock considered lowers TFP by 1% at date 1, recovering with a persistence of 0.5.

power which arises from relationships right, as model variants which vary either element struggle to match this empirical profile. Second, we show that financial and relationship capital are complements at the bank level, with banks expending one of the types of capital to recover from negative shocks to the other. Third, we show that the degree of bank competition matters for the transmission of different types of shocks to the financial system. Relationship capital slows down the recovery of bank net worth after a financial crisis, generating endogenous persistence in variables such as interest rates and quantities of lending. It also generates opposite implications for how the capitalization of the banking system reacts to a monetary policy tightening: while bank net worth falls in response to an increase in bank funding costs in a competitive system, it rises in the presence of relationship capital. Finally, the presence of relationship capital amplifies the response of interest rates to credit demand shocks, but attenuates that of quantities.

Our analysis suggests several promising directions for future research, and we provide some examples of these directions here. First, two trends have played out simultaneously in the U.S. financial sector over the past several decades: the consolidation of the commercial banking industry and the rise of alternative financial intermediaries such shadow banks and online lenders ("fintech"). On the one hand, consolidation would suggest that relationships matter even more, while the competition from outside the traditional banking sector could plausibly counteract this effect. Our framework is well-suited to measuring the aggregate impacts of these trends along this transition path. Second, the fact that different countries' banking sectors are structured in ways very different from the U.S. might imply different interactions between relationship capital and financial frictions than those described in the present paper, which was intended to model the U.S. banking sector. For example, it is well-documented that the Canadian banking sector is considerably more consolidated than the U.S. financial sector. A "discrete bank" version of our framework could be used to examine how relationships interact not only with financial constraints, but also with standard sources of market power arising from industry concentration. We leave these and other avenues for future research.

References

Adrian, Tobias and Nina Boyarchenko, “Intermediary leverage cycles and financial stability,” Staff Reports 567, Federal Reserve Bank of New York 2012.

Amiti, Mary and David E. Weinstein, “How Much Do Idiosyncratic Bank Shocks Affect Investment? Evidence from Matched Bank-Firm Loan Data,” *Journal of Political Economy*, 2018, *126* (2), 525–587.

Berger, Allen N. and Gregory F. Udell, “Relationship Lending and Lines of Credit in Small Firm Finance,” *The Journal of Business*, 1995, *68* (3), 351–381.

_ **and Timothy H. Hannan**, “The Efficiency Cost of Market Power in the Banking Industry: A Test of the Quiet Life and Related Hypotheses,” *The Review of Economics and Statistics*, 1998, *80* (3), 454–465.

Berlin, Mitchell and Loretta J Mester, “On the profitability and cost of relationship lending,” *Journal of Banking & Finance*, 1998, *22* (6), 873 – 897.

Bigio, Saki and Javier Bianchi, “Banks, Liquidity Management, and Monetary Policy,” Staff Report 503, Federal Reserve Bank of Minneapolis September 2014.

Christiano, LawrenceÂ J., Martin Eichenbaum, and CharlesÂ L. Evans, “Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy,” *Journal of Political Economy*, 2005, *113* (1), 1–45.

Corbae, Dean and Pablo D’Erasmo, “Capital Requirements in a Quantitative Model of Banking Industry Dynamics,” NBER Working Papers 25424, National Bureau of Economic Research, Inc January 2019.

de Guevara, Juan Fernández, Joaquín Maudos, and Francisco Pérez, “Market Power in European Banking Sectors,” *Journal of Financial Services Research*, Apr 2005, *27* (2), 109–137.

Degryse, Hans, Olivier De Jonghe, Sanja Jakovljević, Klaas Mulier, and Glenn Scheppens, “Identifying credit supply shocks with bank-firm data: Methods and applications,” *Journal of Financial Intermediation*, 2019, *40* (C).

Demsetz, Rebecca, Philip E. Strahan, and Marc R. Saidenberg, “Banks with something to lose: the disciplinary role of franchise value,” *Economic Policy Review*, 1996, (oct), 1–14.

Dotsey, Michael and Robert G. King, “Implications of state-dependent pricing for dynamic macroeconomic models,” *Journal of Monetary Economics*, January 2005, *52* (1), 213–242.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl, “The Deposits Channel of Monetary Policy,” *The Quarterly Journal of Economics*, 2017, *132* (4), 1819–1876.

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

Faria-e-Castro, Miguel, Pascal Paul, and Juan M. Sanchez, “Evergreening,” *Journal of Financial Economics*, Forthcoming.

Farinha, Luísa A. and Joao Santos, "Switching from Single to Multiple Bank Lending Relationships: Determinants and Implications," *Journal of Financial Intermediation*, 2002, *11* (2), 124–151.

Gilchrist, Simon, Raphael Schoenle, Jae Sim, and Egon Zakrajsek, "Inflation Dynamics during the Financial Crisis," *American Economic Review*, March 2017, *107* (3), 785–823.

Gourio, François and Leena Rudanko, "Customer Capital," *Review of Economic Studies*, 2014, *81* (3), 1102–1136.

Ioannidou, Vasso and Steven Ongena, ""Time for a Change": Loan Conditions and Bank Behavior when Firms Switch Banks," *The Journal of Finance*, 2010, *65* (5), 1847–1877.

Kimball, Miles S., "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit and Banking*, 1995, *27* (4), 1241–1277.

Neri, Andrea Geraliand Stefano, Luca Sessa, and Federico M. Signoretti, "Credit and Banking in a DSGE Model of the Euro Area," *Journal of Money, Credit and Banking*, 2010, *42*, 107–141.

Petersen, Mitchell A.. and Raghuram G. Rajan, "The Benefits of Lending Relationships: Evidence from Small Business Data," *The Journal of Finance*, 1994, *49* (1), 3–37.

Polo, Alberto, "Imperfect pass-through to deposit rates and monetary policy transmission," Bank of England working papers 933, Bank of England July 2021.

Prescott, Edward Simpson and Hubert P. Janicki, "Changes in the size distribution of U.S. banks: 1960-2005," *Economic Quarterly*, 2006, (fall), 291–316.

Ravn, Morten, Stephanie Schmitt-Grohé, and Martín Uribe, "Deep Habits," *Review of Economic Studies*, 2006, *73* (1), 195–218.

van Binsbergen, Jules H., William F. Diamond, and Marco Grotteria, "Risk-free interest rates," *Journal of Financial Economics*, 2022, *143* (1), 1–29.

Appendix for "A Quantitative Theory of Relationship Lending"

A Model Appendix

A.1 Proof of Proposition 1: Loan Demand System

First note that cost-minimization implies an optimal capital-labor ratio that allows us to express optimal labor as a function of the choice of capital

$$n = \frac{\overline{uc}}{\overline{w}}\frac{\eta}{\alpha}k \tag{A.1}$$

This implies that total costs can be written as $\overline{uc}k + \overline{w}k = \overline{uc}k\frac{\alpha+\eta}{\alpha}$. Begin by placing multipliers $\lambda \geq 0$ on constraint (3) and $\zeta \geq 0$ on constraint (4) and taking first order conditions in the borrower's problem (2) – (4):

$$[k]\ A\left(\frac{\alpha}{\overline{uc}}\right)^{1-\eta}\left(\frac{\eta}{\overline{w}}\right)^{\eta}k^{\alpha+\eta-1} = 1+\lambda\kappa$$

$$[L']\ 1-\frac{\phi}{2}\int\left(\frac{q\ell'(q,s)}{L'}-s+S-1\right)^2 \mathrm{d}\mu(q,s) - \phi L'\int\left(-\frac{q\ell'(q,s)}{(L')^2}\right)\left(\frac{q\ell'(q,s)}{L'}-s+S-1\right)\mathrm{d}\mu(q,s)$$
$$+\lambda-\zeta = 0$$

$$[\ell'(q,s)]\ \ -\phi L'\frac{q}{L'}\left(\frac{q\ell'(q,s)}{L'}-s+S-1\right)\mathrm{d}\mu(q,s) + \overline{q}\mathbb{E}\left[V_{\mathcal{L}'}(\ell';\mu)\right] + \zeta q \mathrm{d}\mu(q,s) = 0$$

With the envelope condition $W_\ell(\mathcal{L};\mu) = -\mathrm{d}\mu(q,s)$, we obtain the optimality conditions:

- for capital demand:

$$k = \left[\frac{A\left(\frac{\alpha}{\overline{uc}}\right)^{1-\eta}\left(\frac{\eta}{\overline{w}}\right)^{\eta}}{1+\lambda\kappa}\right]^{\frac{1}{1-\alpha-\eta}} \tag{A.2}$$

  Applying the binding working capital constraint (3) again, using (A.1) and (A.2) gives aggregate loan demand:

$$L' = \kappa(\alpha+\eta)\left[\frac{A\left(\frac{\alpha}{\overline{uc}}\right)^{\alpha}\left(\frac{\eta}{\overline{w}}\right)^{\eta}}{1+\lambda\kappa}\right]^{\frac{1}{1-\alpha-\eta}} \tag{A.3}$$

  where all we need to do is solve for λ.

- for bank-specific loan demand:[18]

$$\zeta = \frac{\overline{q}}{q} + \phi\left(\frac{q\ell'(q,s)}{L'}-s+S-1\right) \text{ for all } (q,s) \tag{A.4}$$

  Recognizing that equation (A.4) holds for all (q,s), we can integrate the right hand side over

[18] A useful result here is if $X = \mathbb{E}[x]$, then $\mathbb{E}[x/X] = \mathbb{E}[x]/X = X/X = 1$, and similarly

$$\mathbb{V}\left(\frac{x}{X}\right) = \mathbb{E}\left[\left(\frac{x}{X}-\mathbb{E}\left(\frac{x}{X}\right)\right)^2\right] = \mathbb{E}\left[\left(\frac{x}{X}-1\right)^2\right] = \mathbb{E}\left[\left(\frac{x}{X}\right)^2\right] - 2\mathbb{E}\left(\frac{x}{X}\right)+1 = \mathbb{E}\left[\left(\frac{x}{X}\right)^2\right]-1$$

the distribution μ to obtain:

$$\zeta = \overline{q} \int \frac{\mathrm{d}\mu(q,s)}{q} + \phi \int \frac{q\ell'(q,s)}{L'} \mathrm{d}\mu(q,s) - \phi \int s \mathrm{d}\mu(q,s) + \phi(S-1) = \overline{q}R \tag{A.5}$$

Plugging (A.5) back into (A.4) and rearranging terms gives us our bank-specific loan demand equation (5).

- for total loan demand:

$$\begin{aligned}
1+\lambda-\zeta &= \frac{\phi}{2}\int\left(\frac{q\ell'(q,s)}{L'}-s+S-1\right)^2 \mathrm{d}\mu(q,s) - \frac{\phi}{L'}\int q\ell'(q,s)\left(\frac{q\ell'(q,s)}{L'}-s+S-1\right)\mathrm{d}\mu(q,s) \\
&= \phi\int\left(\frac{q\ell'(q,s)}{L'}-s+S-1\right)\left(\frac{1}{2}\left(\frac{q\ell'(q,s)}{L'}-s+S-1\right)-\frac{q\ell'(q,s)}{L'}\right)\mathrm{d}\mu(q,s) \\
&= -\frac{\phi}{2}\left[\int\left(\frac{q\ell'(q,s)}{L'}\right)^2\mathrm{d}\mu(q,s) - \int(s-S+1)^2\mathrm{d}\mu(q,s)\right] \\
&= -\frac{\phi}{2}\left[\mathbb{V}_\mu\left(\frac{q\ell'}{L'}\right) - \mathbb{V}_\mu(s)\right]
\end{aligned} \tag{A.6}$$

We can use (5) to simplify (A.6):

$$\begin{aligned}
\mathbb{V}_\mu\left(\frac{q\ell'}{L'}\right) - \mathbb{V}_\mu(s) &= \mathbb{V}_\mu\left(1+S-s-\frac{\overline{q}}{\phi}(r-R)\right) - \mathbb{V}_\mu(s) \\
&= \mathbb{V}_\mu(s) + \mathbb{V}_\mu\left(\frac{\overline{q}}{\phi}(r-R)\right) + 2\mathbb{C}_\mu\left(s, -\frac{\overline{q}}{\phi}(r-R)\right) - \mathbb{V}_\mu(s) \\
&= \left(\frac{\overline{q}}{\phi}\right)^2 \mathbb{V}_\mu(r) - 2\frac{\overline{q}}{\phi}\mathbb{C}_\mu(s,r)
\end{aligned} \tag{A.7}$$

This delivers the aggregate demand equation (6) since we now have

$$\begin{aligned}
1+\lambda-\overline{q}R-\phi(1-S) &= -\frac{\phi}{2}\left[\left(\frac{\overline{q}}{\phi}\right)^2\mathbb{V}_\mu(r) - 2\frac{\overline{q}}{\phi}\mathbb{C}_\mu(s,r)\right] \\
\implies 1+\lambda &= \overline{q}\left[R+\mathbb{C}_\mu(s,r) - \frac{1}{2}\frac{\overline{q}}{\phi}\mathbb{V}_\mu(r)\right] \\
\implies \lambda &= \overline{q}\tilde{R} - 1
\end{aligned}$$

where $\tilde{R}$ is given by equation (7) from the main text.

A.2 Proof of Proposition 2: Bank Financing and Lending

Since $\psi'(e) > 0$, the budget constraint (9) must bind, and so we can eliminate e from the set of control variables. Mechanically, conditions (11), (12), and (13) must bind with $\ell(q,s)$ given by (5), and so we may further eliminate s', ℓ', and n'. This leaves us with a two-control-variable problem (dropping explicit dependence on μ to ease notation):

$$\begin{aligned}
V(n,s,z) &= \max_{q,d'} \psi\left(\overline{q}^d d' + z + n - q\ell(q,s)\right) + \overline{q}\pi\mathbb{E}\left[V\left(n'(q,d',s), s'(q,s); z'\right)\right] \\
\text{subject to} \quad [\lambda] \quad & \overline{q}^d d' \leq (1-\chi)q\ell(q,s)
\end{aligned}$$

where it is understood that $n'(q,d',s) = \ell(q,s) - d'$ and $s'(q,s) = \rho_q \frac{q\ell'(q,s)}{L'} + \rho_s s$ (we keep these general now for flexibility of specification). Taking first order conditions, we obtain:

$$\frac{\partial q\ell}{\partial q}\psi'(e) \quad = \quad \overline{q}\pi\mathbb{E}\left[V_n\left(n',s',z'\right)\frac{\partial n'}{\partial q} + V_s\left(n',s',z'\right)\frac{\partial s'}{\partial q}\right] + \lambda(1-\chi)\frac{\partial q\ell}{\partial q} \tag{A.8}$$

$$\overline{q}^d\psi'(e) \quad = \quad -\overline{q}\pi\mathbb{E}\left[V_n\left(n',s',z'\right)\frac{\partial n'}{\partial d'}\right] + \lambda\overline{q}^d \tag{A.9}$$

The relevant envelope conditions are:

$$V_n\left(n,s,z\right) \quad = \quad \psi'(e) \tag{A.10}$$

$$V_s\left(n,s,z\right) \quad = \quad q\frac{\partial \ell}{\partial s}\left(\lambda(1-\chi) - \psi'(e)\right) + \overline{q}\pi\mathbb{E}\left[V_n(n',s',z')\frac{\partial n'}{\partial s} + V_s(n',s',z')\frac{\partial s'}{\partial s}\right] \tag{A.11}$$

In addition, the ancillary derivatives for accumulating state variables are

$$\frac{\partial n'}{\partial q} = \frac{\partial \ell}{\partial q} \text{ and } \frac{\partial n'}{\partial s} = \frac{\partial \ell}{\partial s} \text{ and } \frac{\partial n'}{\partial d'} = -1 \tag{A.12}$$

$$\frac{\partial s'}{\partial q} = \frac{\rho_q}{L'}\frac{\partial q\ell}{\partial q} \text{ and } \frac{\partial s'}{\partial s} = \frac{q\rho_q}{L'}\frac{\partial \ell}{\partial s} + \rho_s \tag{A.13}$$

Turning first to financing results, combining equations (A.12) and (A.10) with (A.9) yields

$$\lambda = \psi'(e) - \frac{\overline{q}}{\overline{q}^d}\pi\mathbb{E}\left[\psi'(e')\right] \tag{A.14}$$

If the capital requirement is slack, then the marginal equity issuance cost today is equal to the appropriately discounted expected marginal equity issuance cost tomorrow; otherwise, these costs are relatively steep today.

Before considering the pricing policy, it is useful to simplify the envelope condition for relationship intensity (A.11). Using (A.12) and (A.10) and switching to sequential notation, we can first write

$$V_{s,t} = q_t\frac{\partial \ell_{t+1}}{\partial s_t}\left[\underbrace{\lambda_t(1-\chi) - \psi'(e_t) + \frac{\overline{q}}{q_t}\pi\mathbb{E}_t\left[\psi'(e_{t+1})\right]}_{\equiv \Pi_t}\right] + \overline{q}\pi\frac{\partial s_{t+1}}{\partial s_t}\mathbb{E}_t\left(V_{s,t+1}\right)$$

where the term in brackets represents the static flow profits associated with an additional unit of lending defined in (15). From equation (5) we know that $\frac{\partial \ell'}{\partial s} = \frac{L'}{q}$ which implies that $\frac{\partial s'}{\partial s} = \rho_q + \rho_s$, so this can be simplified further:

$$V_{s,t} = L_{t+1}\Pi_t + \overline{q}\pi(\rho_q + \rho_s)\mathbb{E}_t\left(V_{s,t+1}\right) \tag{A.15}$$

Iterating on equation (A.15) yields

$$\begin{aligned} V_{s,t} &= L_{t+1}\Pi_t + \overline{q}\pi(\rho_q+\rho_s)\mathbb{E}_t\left[L_{t+2}\Pi_{t+1} + \overline{q}\pi(\rho_q+\rho_s)\mathbb{E}_{t+1}\left(V_{s,t+2}\right)\right] \\ &= \ldots \\ &= \sum_{i=0}^{\infty}\left(\overline{q}\pi(\rho_q+\rho_s)\right)^i L_{t+i+1}\Pi_{t+i} \end{aligned} \tag{A.16}$$

Next, combine equations (A.8), (A.12), (A.13), (A.10), and the simplifications above to obtain a modified version of the pricing optimality condition

$$\frac{\partial q\ell}{\partial q}\psi'(e) = \overline{q}\pi\frac{\partial \ell}{\partial q}\mathbb{E}[\psi'(e')] + \overline{q}\pi\frac{\rho_q}{L'}\frac{\partial q\ell}{\partial q}\mathbb{E}_t\left[V_s'\right] + \lambda(1-\chi)\frac{\partial q\ell}{\partial q}$$

We can simplify by dividing through by $\frac{\partial q\ell}{\partial q}$, which involves recognizing that since $\frac{\partial q\ell}{\partial q} = q\frac{\partial \ell}{\partial q} + \ell$,

$$\frac{\frac{\partial \ell}{\partial q}}{\frac{\partial q\ell}{\partial q}} = \frac{\frac{\frac{\partial q\ell}{\partial q}-\ell}{q}}{\frac{\partial q\ell}{\partial q}} = \frac{1}{q}\left(1-\epsilon^{-1}(q\ell,q)\right)$$

where $\epsilon(q\ell, q)$ denotes the elasticity of total loan demand, $q\ell$, with respect to loan price, q, so that ϵ^{-1} is the inverse elasticity. Then, combining this expression and the simplified envelope condition (A.16), we obtain the expression from (14):

$$\Pi_t + \overline{q}\pi\rho_q\mathbb{E}_t\left[\sum_{i=0}^{\infty}\left(\overline{q}\pi(\rho_q+\rho_s)\right)^i\frac{L_{t+2+i}}{L_{t+1}}\Pi_{t+1+i}\right] = \epsilon^{-1}(q_t\ell_{t+1}, q_t)\frac{\overline{q}}{q_t}\pi\mathbb{E}_t\left[\psi'(e_{t+1})\right] \tag{A.17}$$

The last part of the proof is to give the form of the inverse elasticity term in equation (16). To derive this, simply compute the derivative of $q\ell$ with respect to q in equation (5):

$$\frac{\partial q\ell}{\partial q} = -L'\frac{\overline{q}}{\phi}\left(-\frac{1}{q}\right)^2 = \frac{\overline{q}}{\phi}\frac{L'}{q^2} \implies \epsilon(q\ell,q) \equiv \frac{\frac{\partial q\ell}{\partial q}}{\frac{q\ell}{q}} = \frac{1}{q}\frac{\overline{q}}{\phi}\frac{L'}{q\ell}$$

A.3 General adjustment cost function

Assume the quadratic adjustment cost function in (2) is replaced by:

$$L'\int\phi\left(\frac{q\ell'(q,s)}{L'},s\right)\mathrm{d}\mu(q,s)$$

where $\phi(\cdot)$ is a generic penalty function that allows for a more general relationship between relative relationship intensity and loan share. Note that this specification still embeds that total adjustment costs scale with the total size of the loan portfolio.

Extending the same analysis from Appendix A.1 shows that this specification gives rise to the modified demand system:

$$-\overline{q}\,(r-R) = \phi_1\left(\frac{q\ell'(q,s)}{L'},s\right) - \underbrace{\int\phi_1\left(\frac{q\ell'(q,s)}{L'},s\right)\mathrm{d}\mu(q,s)}_{\equiv\Phi_1} \tag{A.18}$$

$$L' = \kappa\overline{w}\left[\frac{A\alpha}{\overline{w}}\frac{1}{1+\tilde{\Lambda}(\mu)\kappa}\right]^{\frac{1}{1-\alpha}} \tag{A.19}$$

$$\text{where } \tilde{\Lambda}(\mu) = \overline{q}R + \underbrace{\int\phi\left(\frac{q\ell'(q,s)}{L'},s\right)\mathrm{d}\mu(q,s)}_{\equiv\Phi} - \int\left(\frac{q\ell'(q,s)}{L'}-1\right)\phi_1\left(\frac{q\ell'(q,s)}{L'},s\right)\mathrm{d}\mu(q,s) - 1$$

Equation (A.18) is the analog of (5) in the main text; likewise, equation (A.19) is the analog of (6) in the main text. The former equation still takes the form of specifying loan demand as a function of a pricing penalty term and the marginal cost of relationship adjustment. Likewise, the latter specifies aggregate demand as a function of average interest rates, a term describing aggregate adjustment costs (akin to the covariance term in (7)), and marginal adjustment costs. In particular, assuming that ϕ_1 is invertible, we can write the demand function as

$$\frac{q\ell'(q,s)}{L'} = (\phi_1^{-1})\left(\Phi_1 - \bar{q}(r-R), s\right)$$

This demand function satisfies the same properties as the one that arises from quadratic adjustment costs as long as ϕ_1^{-1} is increasing in both of its arguments. That is, demand rises with more relationship intensity and/or with lower interest rate spreads.

Note the change of notation from $\tilde{R}(\mu)$ in the main text. This is because Λ is actually the multiplier on the working capital constraint, which measures the excess borrowing costs. The analog of $\tilde{R}$ in this context would be such that it solves $\tilde{\Lambda} = \overline{q}\tilde{R} - 1$, or

$$\tilde{R} = R + \bar{q}^{-1}\left[\Phi - \Phi_1 + \int \phi_1\left(\frac{q\ell'(q,s)}{L'}, s\right)\frac{q\ell'(q,s)}{L'}\mathrm{d}\mu(q,s)\right]$$

A.4 CES loan demand

This subsection describes a CES specification for loan demand. The firm's problem can be written as

$$\begin{aligned}
W(\mathcal{L};\mu) \quad &= \max_{n,k,L',\{\ell'(q,s)\}} Ak^\alpha n^\eta - \overline{w}n - \overline{uc}k + L' - \int \ell(q,s)\mathrm{d}\mu(q,s) + \overline{q}\mathbb{E}\left[W(\mathcal{L}';\mu)\right] \\
\text{subject to} \quad & \kappa(\overline{w}n - \overline{uc}k) \le L' \\
& L' \le \left[\int \left(s^\theta q\ell'(q,s)\right)^{\frac{\varepsilon-1}{\varepsilon}}\mathrm{d}\mu(q,s)\right]^{\frac{\varepsilon}{\varepsilon-1}}
\end{aligned}$$

Note that we include the relationship term s directly in the CES for loan demand, in the spirit of how Gilchrist et al. (2017) interpret customer capital in product markets. θ is a parameter that affects how the relationship intensity influences the contribution of borrowing from a particular firm to total borrowing. This is interpreted as a preference shifter in the customer capital literature.

Define $\tilde{R}^{CES}$ as a habit-weighted geometric mean of interest rates:

$$\tilde{R}^{CES} \equiv \frac{1}{\left[\int (s^\theta q)^{\varepsilon-1}\mathrm{d}\mu(q,s)\right]^{\frac{1}{\varepsilon-1}}} \tag{A.20}$$

Then, we can show that the two-tier demand system becomes

$$\begin{aligned}
\frac{q\ell'(q,s)}{L'} &= s^{\theta(\varepsilon-1)}\left(\frac{1/q}{\tilde{R}^{CES}}\right)^{-\varepsilon} \\
L' &= \kappa\bar{w}\left[\frac{A\alpha}{w(1+\kappa(\bar{q}\tilde{R}^{CES}-1)}\right]^{\frac{1}{1-\alpha}}
\end{aligned}$$

As it is well know, the price-elasticity of demand with respect to $R = 1/q$ is equal to $-\varepsilon$, and therefore does not vary either with price or the intensity of relationships.

We can then take logs of the individual demand function to write an estimable version:

$$\log\left(\frac{q\ell'(q,s)}{L'}\right) = -\varepsilon r + \varepsilon \log \tilde{R}^{CES} + \theta(\varepsilon-1)\log s$$

where we use the fact that $\log(1/q) \simeq -r$. The above condition can be estimated using the techniques described in the main body of the text. In this case, while $\tilde{R}^{CES}$ is no longer an average spread, it is subsumed in firm-time FE.

A.5 Kimball loan demand

This subsection describes a specification for loan demand following Kimball (1995). The firm's problem can be written as

$$\begin{aligned} W(\mathcal{L};\mu) \quad &= \quad \max_{n,k,L',\{\ell'(q,s)\}} Ak^{\alpha}n^{\eta} - \overline{w}n - \overline{uc}k + L' - \int \ell(q,s)\mathrm{d}\mu(q,s) + \overline{q}\mathbb{E}\left[W(\mathcal{L}';\mu)\right] \\ \text{subject to} \quad & \quad \kappa(\overline{w}n - \overline{uc}k) \leq L' \\ & \quad 1 = \int G\left(s^{\theta}\frac{q\ell'(q,s)}{L'}\right)\mathrm{d}\mu(q,s) \end{aligned}$$

where G is a general aggregator. We follow Dotsey and King (2005) in assuming that this aggregator takes the form

$$G(x) = \frac{\omega}{1+\omega\nu}\left[(1+\nu)x-\nu\right]^{\frac{1+\omega\nu}{\omega(1+\nu)}} + 1 - \frac{\omega}{1+\omega\nu}$$

Note that this aggregator becomes a standard CES when $\nu = 0$, with $\omega = \frac{\varepsilon-1}{\varepsilon}$. The relevant effective price in this case is defined as

$$\tilde{R}^K = \left[\int \left(\frac{1}{s^{\theta}}\frac{1}{q}\right)^{\frac{1+\omega\nu}{1-\omega}}\mathrm{d}\mu(q,s)\right]^{\frac{1-\omega}{1+\omega\nu}} \tag{A.21}$$

This allows us to write the bank-specific demand function as

$$\frac{q\ell'(q,s)}{L'} = \frac{1}{1+\nu}\frac{1}{s^{\theta}}\left[\left(\frac{1}{s^{\theta}}\frac{R}{\tilde{R}^K}\right)^{\frac{\omega(1+\nu)}{1-\omega}} + \nu\right] \tag{A.22}$$

where $R \equiv 1/q$. The firm's aggregate credit demand is still given by an expression of the form

$$L' = \kappa(\alpha+\eta)\left[\frac{A\left(\frac{\alpha}{\overline{uc}}\right)^{\alpha}\left(\frac{\eta}{\overline{w}}\right)^{\eta}}{1+\lambda\kappa}\right]^{\frac{1}{1-\alpha-\eta}}$$

where λ is now given by a more involved expression:

$$\lambda = \tilde{R}^K q \int \left\{\left[(1+\nu)\frac{1}{s^{\theta}}\frac{q\ell'(q,s)}{L'} - \nu\right]^{\frac{1-\omega}{\omega(1+\nu)}} \frac{1}{s^{\theta}}\frac{q\ell'(q,s)}{L'}\right\}\mathrm{d}\mu(q,s) - 1$$

The Kimball demand function has advantages and disadvantages over the CES specification. The main advantage is that unlike in the CES case, the price-elasticity of demand under Kimball

(1995) is no longer constant and varies with both price and relationship intensity,

$$\epsilon(q\ell, R) = \frac{\omega(1+\nu)}{1-\omega} \frac{\left(\frac{1}{s^\theta}\frac{R}{\tilde{R}^K}\right)^{\frac{\omega(1+\nu)}{1-\omega}}}{\left(\frac{1}{s^\theta}\frac{R}{\tilde{R}^K}\right)^{\frac{\omega(1+\nu)}{1-\omega}} + \nu}$$

One disadvantage, however, is that the bank-specific demand function (A.22) no longer has a functional form that is amenable to direct estimation with linear methods. In particular, the right-hand side depends on the bank-specific interest rate, the relationship intensity term, and on the aggregate time-varying object $\tilde{R}^K$, and these terms cannot be disentangled with either linear or log-linear transformations of this expression.

A.6 Model with perfect competition

The perfectly competitive version of our model corresponds to the case in which there are no adjustment costs; that is, the case when $\phi = 0$ in the borrower's objective function (2). In this case, the state variable s is completely redundant. Furthermore, there is no reason for the borrower to diversify its loan portfolio, and in fact bank-specific demand is not well-defined and so in equilibrium all banks must charge the same loan price, $Q = R^{-1}$.

Correspondingly, the problem of the borrower is simply to choose labor, capital, and total loan demand per the following problem:

$$W(\mathcal{L}; R) \quad = \quad \max_{n,k,L'} Ak^\alpha n^\eta - \overline{w}n - \overline{uc}k + \frac{L'}{R} - L + \overline{q}\mathbb{E}\left[W(\mathcal{L}'; R)\right] \tag{A.23}$$

$$\text{subject to} \qquad \kappa(\overline{w}n + \overline{uc}k) \leq \frac{L'}{R} \tag{A.24}$$

The objective function (A.23) is modified relative to the original objective (2) to reflect that there are no loan sourcing considerations in this model and there is only a single equilibrium loan price. As a result, the loan sourcing constraint (4) is obviated in this version of the model. Finally, observe that the working capital constraint (A.24) is the same as the original constraint (3), with the modification that discount prices are accounted for directly on L' rather than on the individual ℓ'. The solution to this problem yields the same aggregate demand curve as in equation (6), with the modification that the effective interest rate $\tilde{R}(\mu)$ is replaced by the single equilibrium interest rate R:

$$L'(R) = \kappa(\alpha + \eta)\left[\frac{A\left(\frac{\alpha}{\overline{uc}}\right)^\alpha\left(\frac{\eta}{\overline{w}}\right)^\eta}{1 + \kappa(\overline{q}R - 1)}\right]^{\frac{1}{1-\alpha-\eta}} \tag{A.25}$$

The problem of the banks is similarly stripped down:

$$V(n, z; R) \quad = \quad \max_{e,\ell' \geq 0, d', n'} \psi(e) + \overline{q}\pi\mathbb{E}\left[V\left(n', z'; R\right)\right] \tag{A.26}$$

$$\text{subject to} \qquad q\ell' + e \leq n + z + \overline{q}^d d' \tag{A.27}$$

$$\chi q\ell' \leq q\ell' - \overline{q}^d d' \tag{A.28}$$

$$n' = \ell' - d' \tag{A.29}$$

The only change in the objective function in (A.26) relative to the baseline (8) is the elimination of the state variable s from the value function and the removal of the loan price q from the set of control variables. Constraints (A.27), (A.28), and (A.29) are identical to their counterparts from the baseline model, (9), (10), and (13), respectively. Since banks do not face bank-specific demand curves and the state variable s has no meaning in this version of the model, constraints (12) and (11) become irrelevant in this case.

A stationary recursive competitive equilibrium for this version of the model is defined in the standard way. The main differences relative to the equilibrium definition from the main text are that now borrower optimality specifies only aggregate demand, and the distributional consistency condition is replaced by the simple market clearing condition that aggregate demand equals aggregate supply, integrated across the entire equilibrium distribution of banks.

$$L'(R) = \int \ell'(n,z)\mathrm{d}m(n,z) \tag{A.30}$$

A.7 Switching metric

In the model, we say that a "switch" has occurred if a bank lends more in the current period than it did in the last period. We do not count reductions in lending as switches since this would imply double-counting: in a stationary equilibrium, any bank's increase in loan volume must be offset by another bank's reduction. The total volume of switches integrates the increase in lending across all banks in the distribution. It is not necessary to track individual banks; instead, we can use the state-specific policy functions and the implied endogenous transitions over bank states to compute the total volume of switches among banks in a given state today based on the evolution of their states tomorrow. We compute:

$$\begin{aligned}\text{switch volume} \quad = \quad & \underbrace{\pi \int_{\mathcal{N}\times\mathcal{S}\times\mathcal{Z}} \max\left\{g_{q\ell'}(n',s',z') - g_{q\ell'}(n,s,z), 0\right\} f(n',s',z'|n,s,z)\mathrm{d}m(n,s,z)}_{\text{net changes in volume from incumbents}} \\ & + \underbrace{(1-\pi)\int_{\mathcal{Z}} g_{q\ell'}(0,0,z)g_\ell(0,0,z)\mathrm{d}\overline{\Gamma}(z)}_{\text{volume from entrants}} \end{aligned} \tag{A.31}$$

where the transitions between bank states are summarized by the density:

$$f(n',s',z'|n,s,z) = \mathbf{1}\left[n' = g_\ell(n,s,z) - g_d(n,s,z)\right]\mathbf{1}\left[s' = \rho_q \frac{g_{q\ell'}(n,s,z)}{L'(\mu)} + \rho_s s\right]\Gamma(z'|z)$$

The first term in (A.31) accounts for the increases in lending coming from incumbent banks. The second term accounts for entrants; since entrants did not lend last period, we count their entire loan volume as switches. Note that for entrants we do not integrate across the entire distribution, but rather just the entrant-specific distribution, which reflects that $n = 0$, $s = 0$, and z is drawn from the ergodic distribution implied by the transition matrix $\Gamma(z'|z)$. Then, we simply scale the volume of switches by total loan volume, reporting the following metric in Table 4 in the main text:

$$\text{switch share} = \frac{\text{switch volume}}{L'(\mu)} \tag{A.32}$$

B Computational Appendix

B.1 Bank problem

1. Given a current guess of V, compute expected $\overline{V}$ (over z) for all (n, s, z):

$$\overline{V}(n, s, z; \mu) = \overline{q}\pi \sum_{z'} \Gamma(z, z') V\left(n, s, z'; \mu\right) \tag{B.1}$$

before entering in loop to compute policies. This saves on calculations later.

2. Fix (n, s, z). Implement nested golden section with the q choice as the outer loop and the deposit choice d' as the inner loop.

   (a) Identify the highest loan price at which the resulting loan demand can be serviced while satisfying the capital requirement ($q_{\max}(n, s, z; \mu)$) and the lowest price such that the loan demand implied by (5) is non-negative ($q_{\min}(s; \mu)$).[19]

$$\begin{aligned} q_{\min}(s; \mu) &= \left[R(\mu) + \frac{\phi}{\overline{q}}(1 - S + s)\right]^{-1} \\ q_{\max}(n, s, z; \mu) &= \min\left\{\overline{q}^d, \left[R(\mu) + \frac{\phi}{\overline{q}}\left(1 - S + s - \frac{n - \psi(\underline{e})}{\chi}\right)\right]^{-1}\right\} \end{aligned}$$

   where $\underline{e}$ is a program parameter representing the lowest reasonable dividend (i.e. the highest amount of equity issuance, which most eases the capital requirement).

   (b) For each candidate q, compute the implied loan demand $\ell(q, s; \mu)$ and next period customer capital $s'(q; s, \mu)$ via (5) and (11), respectively. Then use golden section to find the optimal level of deposits $d' \in [0, d_{\max}(q; s, \mu)]$, where

$$d_{\max}(q; s, \mu) = (1 - \chi)\frac{q\ell'(q, s; \mu)}{\overline{q}^d},$$

   that should be used to finance the implied loan quantity.

   (c) Compute the implied net worth $n'(q, d'; s, \mu)$ given the candidate policies via (13) and compute the value associated with the current (q, d') according to

$$v(q, d'; n, s, z, \mu) = \psi\left(\overline{q}^d d' + n + z - q\ell(q, s; \mu)\right) + \overline{V}\left(n'(q, d'; s, \mu), s'(q; s, \mu), z; \mu\right)$$

   This step requires interpolation on n' and s'.

   (d) Compute $w(q; n, s, z, \mu) = \max_{d'} v(q, d'; n, s, z, \mu)$ in the inner loop and then $TV(n, s, z; \mu) = \max_q w(q; n, s, z, \mu)$ in the outer loop.

B.2 Steady state

1. Begin with a guess of bank policies $g_0(x)$, the distribution of banks over states $m_0(x)$, and the bank value function $V_0(x)$.

[19] It may be desirable to solve for policies in terms of spreads, $\tau(q; \mu) = r(q) - R(\mu)$, rather than prices directly.

2. Use the consistency condition (18) to obtain the joint distribution of prices and relationships $\mu(q, s)$ implied by (g_0, m_0). Given the implied μ, use equation (7) to compute the effective interest rate $\tilde{R}(\mu)$ and the average interest rate $R(\mu)$.

3. Solve for banks' optimal policies and updated value function under the current distribution μ and current value function $V_0(x)$, $g_1(x; \mu, V_0)$ and $V_1(x; \mu, V_0)$ (using the algorithm described above). Note that it is not necessary to iterate to convergence on the value function / decision rules at this step.

4. Compute the distribution of banks over idiosyncratic states *next period* implied by the policies in the step above, $m_1(x; g_1, \mu)$; that is, iterate *just once* on equation (17).

5. Assess convergence of the distribution of banks over states and state-specific bank policies and values; that is, compute the convergence metric

$$\varepsilon = \max\left\{\max_x |g_1(x) - g_0(x)|, \max_x |m_1(x) - m_0(x)|, \max_x |V_1(x) - V_0(x)|\right\}$$

If $\varepsilon < \bar{\varepsilon}$ (the pre-specified tolerance), then the model is solved; otherwise, update

$$\begin{aligned} g_0(x) &= \psi_g g_0(x) + (1 - \psi_g) g_1(x; \mu) \\ m_0(x) &= \psi_m m_0(x) + (1 - \psi_m) m_1(x; g_1, \mu) \\ V_0(x) &= \psi_V V_0(x) + (1 - \psi_V) V_1(x; g_1, \mu) \end{aligned}$$

and return to step 2. We set $\psi_g = \psi_m = \psi_V = 0$.

B.3 Perfect foresight transitions / impulse responses

The maintained assumption throughout these steps is that both the initial and terminal steady states are known, that the initial distribution of banks over idiosyncratic states may be computed directly given the initial steady state, and that bank policies may be solved backwards given the value function implied by the terminal steady state.

1. Update the initial distribution of banks over idiosyncratic states, $m_0(x)$, to reflect the incidence of the shock being simulated.

2. Guess a sequence of aggregate prices $\{\tilde{R}_t^0, R_t^0\}_{t=1}^T$. A natural initial guess is that these prices are equal to their steady state values at all dates t.

3. Using the terminal value function $V_{T+1}(x)$ and the path of aggregate prices computed in the step above, solve backwards to obtain the sequence of bank policy functions $G = \{g_t(x)\}_{t=1}^T$.

4. Given the sequence of policy functions G, compute the implied sequence of distributions of banks over idiosyncratic states, $M = \{m_t(x)\}_{t=1}^T$.

5. Use the consistency condition to compute the implied sequence of joint distributions of loan prices and relationship intensities, $\{\mu_t(q, s)\}_{t=1}^T$. Then, compute the implied sequence of aggregate prices $\{\tilde{R}_t^1, R_t^1\}$ consistent with this sequence of distributions.

6. Assess the convergence of the aggregate prices: that is, compute the metric

$$\varepsilon = \max\left\{\max_t \left|\tilde{R}_t^1 - \tilde{R}_t^0\right|, \max_t \left|R_t^1 - R_t^0\right|\right\}$$

If $\varepsilon < \bar{\varepsilon}$, the pre-specified tolerance metric, then the transition path has been solved. Otherwise, update the guesses of the aggregate price sequence according to

$$\begin{aligned}
\tilde{R}_t^0 &= \psi_{\tilde{R}} \tilde{R}_t^0 + (1 - \psi_{\tilde{R}}) \tilde{R}_t^1 \\
R_t^0 &= \psi_R R_t^0 + (1 - \psi_R) R_t^1
\end{aligned}$$

and return to step 2.

C Additional Quantitative Results

C.1 Spread Policies across Model Economies

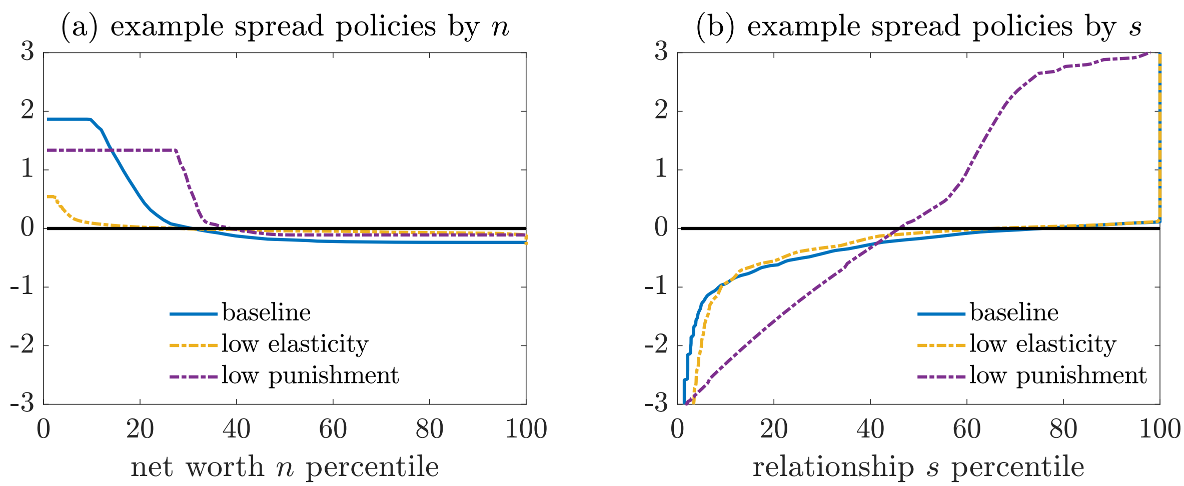


Figure C.1: **Steady state loan pricing policies**

Notes: Panels (a) and (b) plot sample pricing policy functions over the equilibrium distribution of net worth and relationships, respectively, for the indicated model variants. Percentiles for each line are from the respective equilibrium distribution for each model variant. Panels (a) and (b) each fix $z = -0.0025$. Panel (a) fixes s, and panel (b) fixes n, at the 25th percentile of the equilibrium distribution of the relevant variable in the baseline economy.

C.2 Lending policies across model economies

Figure C.2 is the analog of Figure 4 from the main text, except we show lending policies directly (expressed in units of relative loan volume, $\frac{q\ell'}{L'}$), rather than spreads. Panel (a) shows that the joint correlation between loan shares and net worth follows the pattern for relationships and net worth, a direct by-product of the accumulation process (11). Panel (b) shows that lending policies are highly elastic with respect to net worth in the bottom quarter of the distribution in the baseline; for higher levels of net worth, these policy functions are essentially flat. By contrast, for the competitive economy, relative loan volumes increase almost linearly in net worth over the entire range of the net worth distribution.

C.3 Decomposing the drivers of bank behavior

As highlighted in the optimal pricing condition (14), banks' pricing policies involve both the level of market power a bank has and how the bank allocates the rents it obtains from that market power

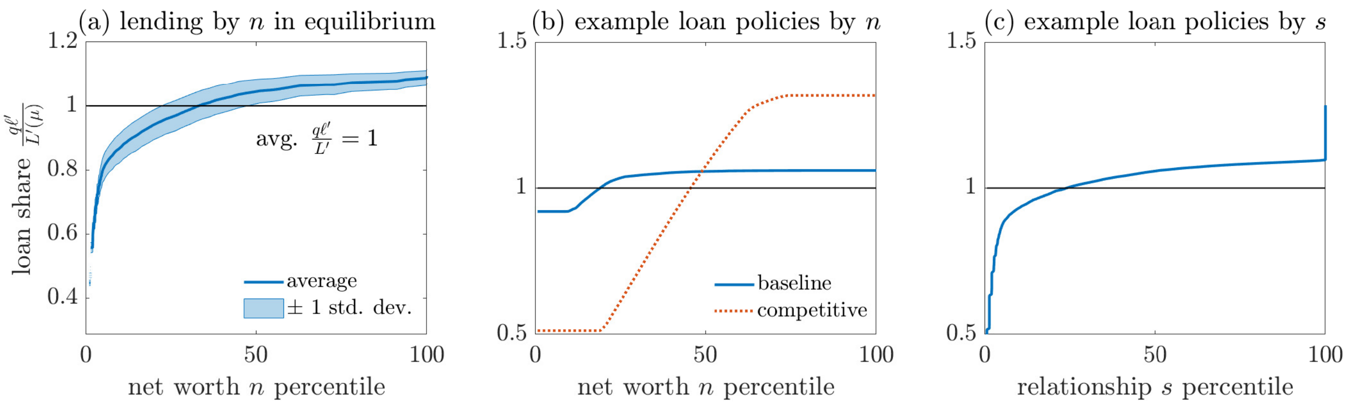


Figure C.2: **Steady state lending policies**

Notes: Panel (a) plots the first and second moments of the lending policies (expressed as loan volume relative to the average) conditional on a given level of net worth over the equilibrium distribution of banks for the baseline model. Variation at each level of n comes from dispersion in s and z. Panels (b) and (c) plot sample lending policy functions (expressed in the same units as in panel (a)) over the equilibrium distribution of net worth and relationships, respectively, for the indicated model variants. Percentiles for each line are from the respective equilibrium distribution for each model variant. Panels (b) and (c) each fix $z = -0.0025$. Panel (b) fixes s, and panel (c) fixes n, at the 25th percentile of the equilibrium distribution of the relevant variable in the baseline economy.

over time. How large are these two forces?

In order to address this question, in this subsection we construct a partial equilibrium decomposition of banks' loan pricing decisions. Specifically, given a solution to the baseline economy – and therefore given levels of aggregate loan demand, $L'(\mu)$, and an average interest rate, $R(\mu)$, from the bank-specific demand curve (5) –, we consider two alternative sets of bank policies. First, we consider policies chosen as though banks have no static market power, i.e. $\phi = 0$. Second, we compute optimal policies under the extreme scenario in which relationship intensity is fixed through time at the bank level, i.e. $\rho_q = 0$. The results of this analysis are shown in Figure C.3 below.

Since banks in the baseline economy have market power, they lend less than in the competitive economy, as highlighted in panel (a). This difference is especially strong for the least financially constrained banks, for whom the standard quantity rationing force is the strongest; in contrast, banks with low net worth generally like to lend at low volumes in both economies.

Panel (b) shows that the forces for the low ρ_q economy work in the other direction: given the increased persistence of relationships in this economy, banks generally would like to lend more than in the baseline model. Panel (c) reveals that this is achieved by charging significantly below-market interest rate spreads. Again, for the same reasons as in the competitive case, the differences between policies across models are the largest for the most financially unconstrained banks.

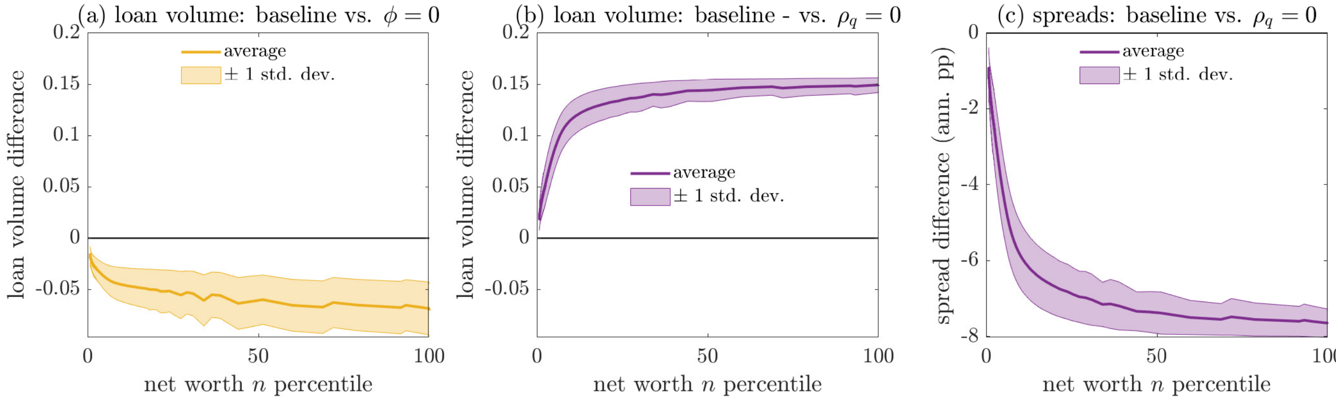


Figure C.3: **Decomposition of equilibrium lending policies**

Notes: Panels (a) and (b) plot the average and standard deviation of the difference between the loan volume in the baseline economy and the loan volume in the nearly competitive and low punishment economies, respectively. In each case, for each level of net worth in the equilibrium distribution of the baseline economy, we compute the average and standard deviation in loan volume differences across the alternative economies. Panel (c) performs a similar analysis for spreads for the low punishment model; there is no need to compare spreads for the competitive model, since these would be zero by definition, and so the levels plotted in Figure 4 would also be the differences.

C.4 Response to Aggregate Shocks across Model Economies

Figures C.4-C.6 plot the responses of selected variables to shocks to the aggregate shocks we consider in the main text, including also the responses for the low elasticity and low punishment economies.

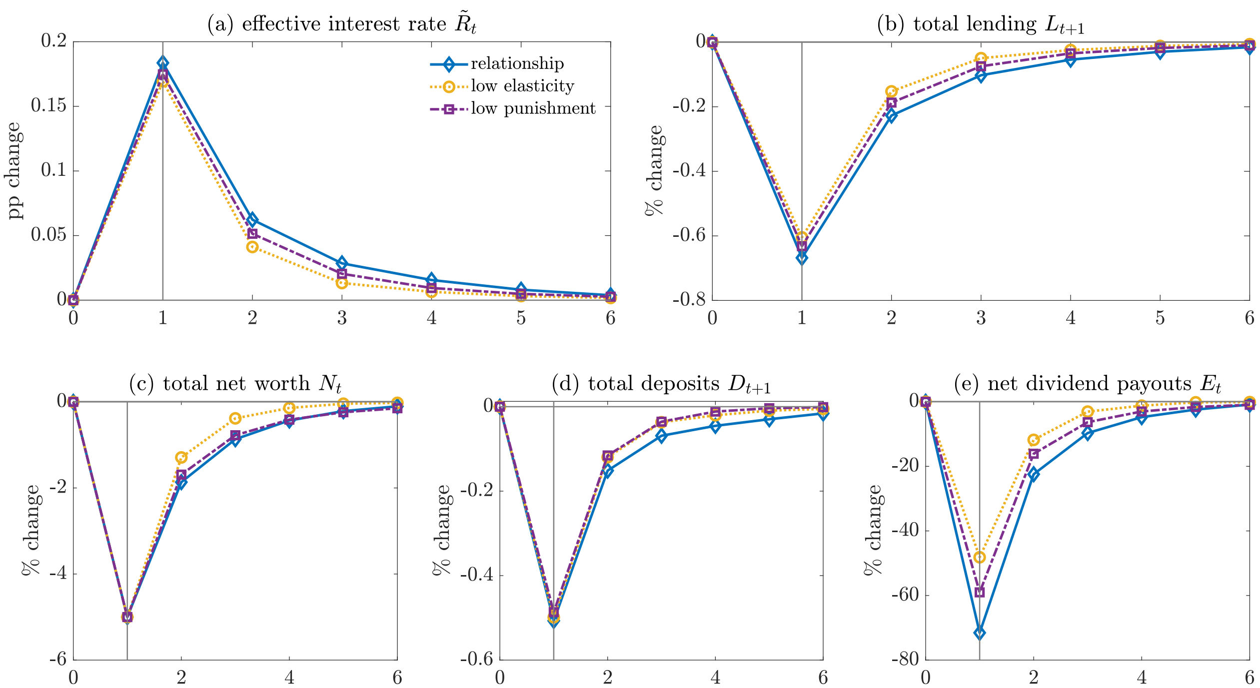


Figure C.4: **Aggregate shock to bank net worth**

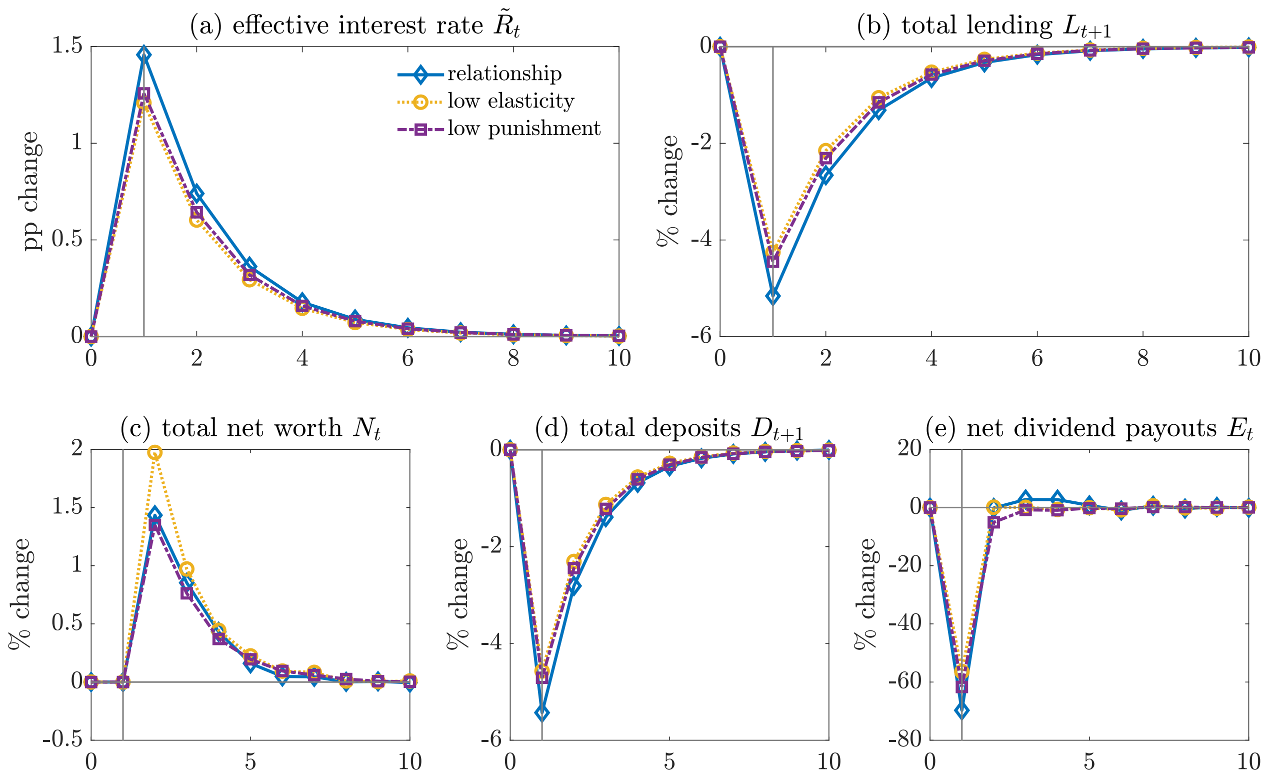


Figure C.5: **Aggregate shock to cost of funding**

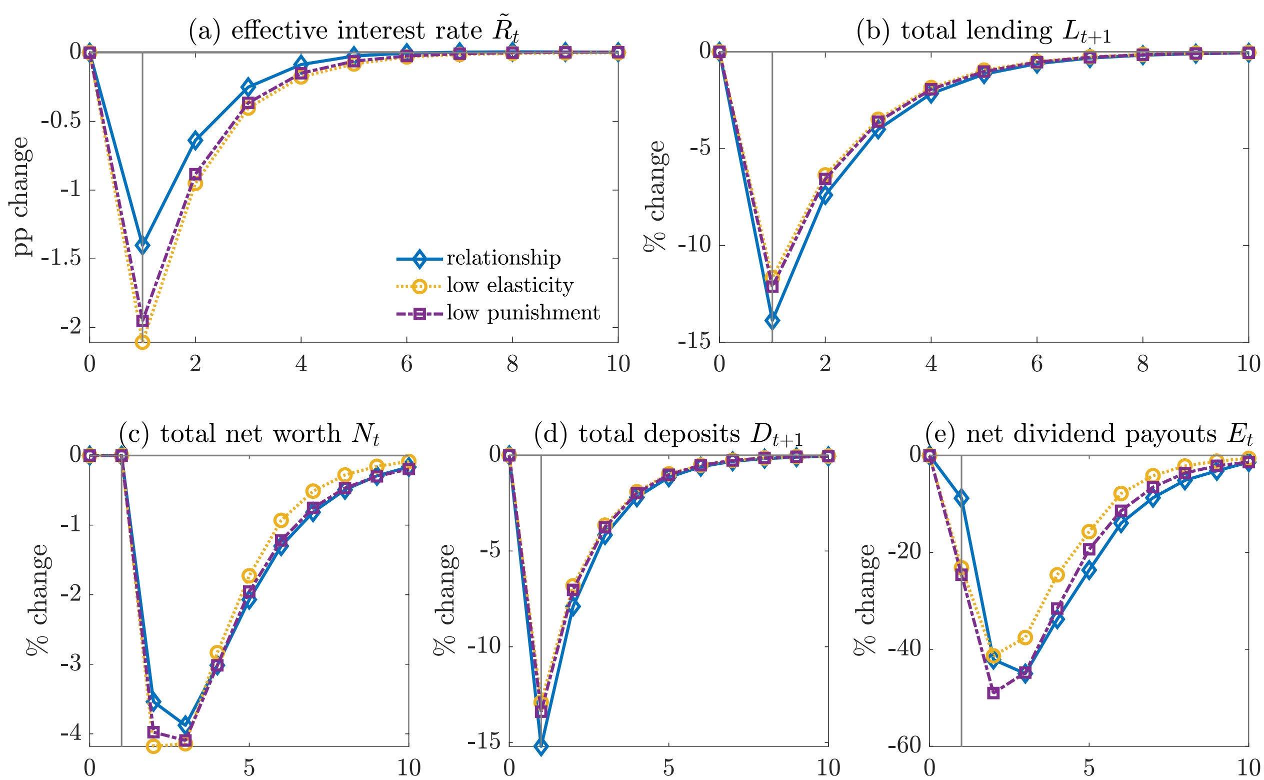


Figure C.6: **Aggregate shock to loan demand**