

# MacroPru in the Shadow of Declining Balance Sheet Model of Banking

Greg Buchak, Gregor Matvos, Tomasz Piskorski, and Amit Seru<sup>1</sup>

April 2, 2023

## Abstract

We argue that existing models and approaches for understanding the response of lending to macroeconomic or financial shocks are deficient because they do not account for the modern industrial organization of financial intermediation. We investigate two increasingly significant margins of adjustment in the credit market: banks' decision to retain loans on their balance sheets and shadow bank activity. By utilizing micro-level lending data from the largest private credit market in the United States, we demonstrate that these adjustment margins are substantial, variable, and adjust in equilibrium in response to changes in the economic environment, including regulatory shocks. Given that regional income and the composition of financial intermediation are correlated, the extent to which these margins dampen or amplify shocks varies across regions. We examine these margins in a parsimonious dynamic quantitative model featuring banks with balance sheet adjustment and shadow banks. Using the calibrated model, we illustrate that these margins act as significant dampeners on shocks to bank capital ex-post. Consequently, adverse shocks to capital result in smaller contraction in lending by banks and in the aggregate. Recovery from shocks is faster, because profitable loan sales (securitization) allow banks to build capital faster. Ex ante, because capitalization shocks are less costly, banks are less prudent with capital, and take larger advantage of balance sheet capacity when lending. Our model highlights the basic economic tension in matching bank balance sheet models to data. It has to reconcile the very weak correlation between bank capital and aggregate lending, with a model which forces total lending to depend strongly on bank balance sheet health. We show that this is only possible with a model that accounts for the bank loan sales margin. To fully comprehend the effect of policies which rely on financial intermediation pass-through one needs to utilize data from bank balance sheets, overall bank lending, and aggregate lending, in conjunction with the model of modern financial intermediation.

---

<sup>1</sup> Buchak is at Stanford Graduate School of Business (GSB) and Stanford Institute for Economic Policy Research (SIEPR), Matvos is at Northwestern University and the National Bureau of Economic Research (NBER), Piskorski is at Columbia GSB and NBER, and Seru is at Stanford GSB, the Hoover Institution, SIEPR, and NBER. We thank Declan Mirabella and Francesco Spizzuoco for excellent research assistantship.

## 1. Introduction

Macro-prudential, monetary, and fiscal policies, are frequently passed through the financial intermediation sector, or implemented directly through financial intermediaries (Kashyap and Stein 2000; Hanson, Kashyap and Stein 2011).<sup>2</sup> The modern industrial organization of financial intermediation differs from the traditional view of bank balance sheet lending in two dimensions (Buchak et al. 2018 and 2022). Specifically, non-depository institutions, that is, shadow banks, now account for a substantial share of lending in many markets, and banks now sell a significant fraction of the loans they originate through securitization. We argue that accounting for the modern industrial organization of financial intermediation is essential in two respects.

Firstly, using bank data to measure lending can lead to erroneous conclusions about the how lending responds to economic or policy shocks, and the extent to which they are amplified through financial intermediaries. We illustrate this point using micro-level lending data on the largest private credit market in the US. Secondly, failing to account for these features of financial intermediation in quantitative macro models can result in inaccurate calibrations of how shocks amplify through financial intermediaries. To study the economics of the problem we develop and calibrate a parsimonious dynamic quantitative model featuring banks with balance sheet adjustment and shadow banks.

We begin our empirical analysis with the observation that regulators and academics commonly employ three different sources for measuring the extent of lending: the data on bank lending activity from bank balance sheets (e.g., the bank call reports), data on total bank loan origination, or aggregate lending data. We focus our analysis on the US residential mortgage market, the largest private credit market in the US, with more than \$11 trillion of loans outstanding. The advantage of this market is that we observe almost all originated loans and know whether they were (i) originated by a bank or shadow bank, and (ii) whether a financial institution retained a loan on its balance sheet or sold it.<sup>3</sup>

We define two empirical multipliers, the “loan sales multiplier” and the “shadow bank lending multiplier,” to capture the mapping between bank balance sheet lending, total bank lending, and total lending. These multipliers measure the significance of two key features of modern financial intermediation described above: the extent to which banks sell a significant portion of loans and

---

<sup>2</sup> For recent papers on pass-through of macro-prudential, monetary, fiscal policies, and other shocks through financial intermediaries see, He and Krishnamurthy (2013), Brunnermeier and Sannikov (2014), Agarwal et al. (2017), Di Maggio et al. (2017), Drechsler et al. (2017), Greenwood et al. (2017), Beraja et al. (2019), Xiao (2020), Cherry et al, (2020, 2021), Corbae and D’Erasmus (2021), Elenev et al. (2021), Hachem and Song (2021), Begeau and Landvoit (2022), Bianchi and Bigio (2022), Buchak et al. (2022), Eichenbaum et al. (2022) and Wang et al. (2022).

<sup>3</sup> While we focus on mortgage market, banks now increasingly sell a significant share of their corporate loans, credit cards, auto loans, and personal loans and shadow banks now account for a substantial amount of intermediation in these markets as well (see Buchak et al. 2022 and Seru 2019).

the involvement of shadow banks in lending. We demonstrate that these multipliers are large and vary over time. For instance, banks on average sell more than half of the mortgages they originate, with this propensity fluctuating between a low of roughly 37% and a high of nearly 80% of annual bank lending. As a result, bank balance sheet lending data, which is frequently used by regulators to gauge lending conditions, accounts for less than half of the variation in aggregate lending of banks, both in terms of their level and growth rates. Even perfect bank data is insufficient to assess the evolution of aggregate lending activity: total bank lending accounts for only about 70% of variation over time in total lending. This is because shadow banks originate substantial share of loans, and their market share significantly varies over time, from a low of roughly 20% during the Great Recession and a high of 60% in 2021. In fact, the growth in bank balance sheet lending and aggregate lending can occasionally be negatively correlated, especially, as we argue below, in times when bank balance sheets are stressed.

We document a significant correlation between regional income and the composition of financial intermediation, which has a significant impact on the magnitude of both the “loan sales multiplier” and the “shadow bank lending multiplier”. Both multipliers are smallest in highest income counties, which primarily rely on bank balance sheet lending. Conversely, in lower to mid-income regions, which may be of particular interest to regulators due to their typically higher share of risky lending, the multipliers are found to be largest. These results imply that the propagation of shocks in high and low-income areas differ, as these households are situated in markets with distinct industrial organizations of financial intermediation.

Buchak et al. (2018, 2022) identify three underlying drivers that influence the two margins of substitution: bank balance sheet strength, the relative attractiveness of the loan sale market, and the regulatory burden on banks. First, banks switch away from traditional bank balance sheet lending and toward selling loans as their capitalization declines. The extent of that switching is determined in equilibrium. Second, the availability and relative attractiveness of the loan sale market affects both margins of substitution. Because shadow banks sell almost all loans they originate, the shadow bank lending margin is crucially affected by the conditions in the secondary-loan market.<sup>4</sup> Finally, changes in the regulatory environment also play a role in determining the market share of shadow banks, with Buchak et al. (2018) finding that a substantial part of the increase in shadow bank market share after the Great Recession can be attributed to increased regulatory burden on banks. Our paper provides further evidence of the importance of bank balance sheet strength in shaping the aggregate loan sale multiplier.

To understand whether modifying existing models to account for contemporary financial intermediation frictions is quantitatively important, we develop a parsimonious dynamic quantitative

---

<sup>4</sup> There are other forces that can affect shadow bank sector like deposit outflows from traditional banking sector towards non-bank sector during times of higher interest rates (see Xiao 2020 and Dreschler et al. 2022).

model of financial intermediation. We build on Buchak et al. (2022) in which the loan sales and shadow bank multipliers play a central role in a static setting.<sup>5</sup> Focusing on the dynamics allows us to understand how these margins contribute to the impact and recovery from financial shocks over time. We calibrate the model to the empirical lending multipliers measured in the data. Using the calibrated model, we examine how shocks to bank capital propagate through this augmented model of the financial intermediation sector.

We show that the financial sector is much more resilient to capital shocks *ex post* than what a bank balance sheet model would suggest. The loan sale multiplier and shadow bank multiplier significantly decrease the effect of a capital shock on lending. For example, a large negative shock to bank capital that would lead to a 40% decline in aggregate lending in a model without these multipliers, leads to only a 4% decline when these multipliers are present. The accompanying lending rate increase would be 10 basis points instead of 100 basis points. Second, the effect of the capital shock is less prolonged, with a faster recovery of banks. Third, because capitalization stocks are less costly *ex post*, banks are less prudent with capital *ex ante*. Intuitively, when the bank has a more difficult time adjusting to shocks, it keeps a larger capital buffer in excess of the statutory capital requirement. In a more complete setting with other margin present, for any given balance sheet capacity, banks originate more loans.

There are two primary features of models that determine how shocks to intermediaries are transmitted to lending, and subsequently, to real-world outcomes. The first is the responsiveness of lending to a shock to bank capital; and the second is the speed at which banks can rebuild capital after the shock. A traditional bank balance sheet model implies that the impact of reduced bank capital on lending could be substantial since it is the only means through which lending can occur. Banks inability to extend profitable loans due to inadequate capitalization further implies slow rebuilding of capital, and therefore the recovery. However, because banks can sell their loans and shadow banks can take over some of the bank lending, the impact of bank capital on lending is much less important. The recovery is faster because undercapitalized banks can instead turn to profitable loan selling, which allows them to rebuild their capital faster.

Finally, we use our setting to illustrate why models based on bank balance sheet lending struggle to quantitatively match data, and therefore have important drawbacks as guides to policy responses. Because of the banks' loan sales margin, there is a strong correlation between bank capitalization and bank balance sheet lending, but a very weak correlation between bank capital and aggregate lending. A bank balance sheet model forces total lending to depend strongly on bank balance sheet health. It is very difficult for bank balance sheet models to achieve this reconciliation. With the move towards integrating micro-data into macro models, it is natural to

---

<sup>5</sup> For other recent models featuring non-bank financial sector see, among others, Gennaioli, Shleifer, and Vishny (2013), Moriera and Savov (2017), Huang (2018), Ordóñez (2018), Jiang (2020), Jiang et al. (2020), Xiao (2020), Hachem and Song (2021), and Begenau and Landgvoit (2022). See also Adrian and Ashcraft (2016).

consider calibrating models to bank-level data instead. With this approach a researcher can exploit both the cross-section in bank capitalization, as well as time series changes within a bank. Our results highlight that using bank balance sheet lending data alone, calibrations overstate the responsiveness of lending to capital shocks. Therefore, neither aggregate lending data nor bank data alone are sufficient to empirically understand the extent of lending responses. Nor can they be used to calibrate macro models of financial intermediation in isolation. Instead, both overall lending and bank balance sheet data must be used in conjunction with the model of modern financial intermediation to fully comprehend the effect of capitalization shocks.

We conclude by discussing the broader implications of our results. There are substantial differences in the industrial organization of the financial sector between countries. The U.K., for example, does not have a large and liquid secondary market for mortgages (Benetton 2021). Our model suggests that the financial intermediation sector propagates shocks to a different extent in the two countries. This casts doubt on regulatory frameworks which propose a uniform treatment of capital requirements across countries such as the Basel framework.

Our model also implies that policies which target the intermediation sector have distributional consequences. Devoting government resources (subsidies) to recapitalize banks most benefits the highest income regions. In contrast policies that operate through secondary markets act across the income distribution. More broadly, regulators who ignore the multipliers we outline may reach inaccurate conclusions about the impact of policies on the credit market and even incorrect assessments about the current health of the lending market. For example, a regulator observing that banks are poorly capitalized would dramatically overstate how aggregate lending would react to further deterioration in bank capital. These biases will vary depending on the ease of loan sales and shadow bank lending in different markets and countries, making it difficult to apply findings from one market to another.

Our paper emphasizes the need to collect data on lending by shadow banks and loans that were not retained on the balance sheets of regulated and closely monitored financial institutions in conjunction with existing approaches, which focus on bank balance sheet data (call report data by traditional banks). In the interim, researchers and policy makers could rely on quantitative lending models, like the one proposed by us in this paper, to recognize the importance of the modern industrial organization of credit markets and allow more complete inferences from the limited data.

## **2. Data and Institutional Setting**

### **2.1 Institutional Setting**

The US residential mortgage market is the largest private debt market in the country, comprising over 50 million properties with an outstanding debt of over \$11 trillion as of 2021. The process of

securing a mortgage, called loan origination, involves a borrower submitting a loan application and documentation related to their financial and credit history to the lender. Figure A1 in the appendix shows the annual aggregate mortgage origination volume in the US residential loan market, which varies between a low of \$1.4 trillion during the Great Recession and a high of over \$4.7 trillion during the pandemic lending and refinancing boom (2020-2021).

There are three main segments of the US residential mortgage market: the conforming loan market, the jumbo loan market, and the FHA loan segment. The conforming loan market is the largest, consisting of loans usually extended to borrowers with high credit scores, conservative loan-to-value ratios, and fully documented incomes and assets. Conforming mortgages must be below the conforming loan limit, which increased from \$417,000 in 2006 to \$548,250 in 2021 for a one-unit, single-family dwelling in a low-cost area. Mortgages that exceed the conforming limit are termed "jumbo" loans.

Conforming loans are eligible for securitization with the participation of government sponsored enterprises (GSEs), while jumbo loans are not. GSEs make securitization of conforming mortgages substantially easier. For example, Fannie Mae and Freddie Mac, the two most prominent GSEs, purchase conforming mortgages and package them into mortgage-backed securities (MBS), insuring default risk. These MBS are particularly attractive to investors interested in relatively safe assets. In 2017, conforming loans packed in MBS guaranteed by Fannie Mae and Freddie Mac made up about 50% of the outstanding residential loans. (Source: *Securities Industry and Financial Markets Association* Data).

The third market segment consists of FHA loans, which are mortgages whose risk of default is directly insured by the Federal Housing Administration (FHA). They are popular among less creditworthy borrowers and first-time home buyers because they allow down payments as little as 3.5%.

The US residential mortgage market is characterized by the presence of two main groups of originators: banks and shadow banks (non-bank lenders). According to Buchak et al. (2018), traditional bank originations have seen a decline, while shadow bank market share has grown from less than 30% to more than 50% by 2015. These originators differ in several aspects. Firstly, banks (traditional banks and credit unions) partially fund their lending through insured deposits, while shadow banks do not take deposits. Secondly, they differ in their business models. Banks engage in both portfolio lending and originate-to-distribute models, with portfolio loans comprising about 47% of their originations during the sample period. On the other hand, shadow banks almost exclusively use the originate-to-distribute model (Buchak et al. 2022). Thirdly, banks face a substantially higher regulatory burden than shadow banks, including capital requirements, enhanced supervision, and compliance rules.

Furthermore, the presence of a large secondary loan market supported by government guarantees makes the residential lending market unique in comparison to other lending markets. However, the insights gained from this market may be broadly applicable to other credit market segments, such as corporate loans, credit cards, auto loans, and personal loans, where a similar shift in the industrial organization of financial intermediation has occurred. For example, in the corporate loan market, lenders can sell their loans through collateralized loan obligations (Irani et al. 2021), in the auto lending market, lenders can sell their loans in the asset-backed commercial paper market (Benmelech et al. 2017). Similarly, shadow banks have a significant presence in other markets, such as small business lending, middle-market firm loans, and personal loans.

Shadow banks have also gained a significant presence in other markets. Chen et al. (2017) indicates that large US banks significantly reduced their small business lending during the Great Recession and have yet to fully return to this market. Gopal and Schnabl (2022) document a substitution of traditional banks with non-bank lenders in the small business loan market, resulting in a significant increase in non-bank lending. Irani et al. (2021) observe that shadow banks have entered the corporate loan market as traditional banks increase their capital holdings. Chernenko et al. (2022) report that among middle-market firms over 2010-2015, one-third of all loans were directly extended by non-bank financial intermediaries. Furthermore, studies by Tang (2019) and DeRoure et al. (2022) illustrate the significant expansion of non-bank lenders in the personal loan market.

## 2.2 Data

By focusing on residential mortgage lending, we have access to comprehensive micro-level lending data on almost all loans made in this market, regardless of whether the loan was originated by a bank or shadow bank, and whether the financial institution retained or sold the loan. We collect this data from 2005 to 2021 through the Home Mortgage Disclosure Act (HMDA), which requires financial institutions to report detailed information on each loan they originate annually, unless exempt under limited exceptions. This covers a vast majority of all residential loans in the United States.

We classify a loan as a "balance sheet" loan if the financial institution does not report selling it in the year of origination. We use the HMDA data and Robert Avery's classification<sup>6</sup> to uniquely identify institutions and classify them as "shadow banks" if they are labeled as independent mortgage banks in the HMDA data. Other institutions are classified as banks.<sup>7</sup> To determine the

---

<sup>6</sup> This can be found at <https://sites.google.com/site/neilbutta/data>.

<sup>7</sup> In our analysis of aggregate and regional lending patterns we use a broad definition of banks that among others also includes other depository institutions such as credit unions. Our results are very similar if we just focus on commercial banks.

capitalization of each bank, we merge the data with the Uniform Bank Performance Report (UBPR) using a unique RSSD ID number.<sup>8</sup> Most FDIC-insured banks are required to report the composition and elements of their balance sheet each quarter to the FFIEC, and we use annualized data that averages across quarters. We use each bank's total tier one risk-based capital to risk-weighted assets (code: ubprd487) as the primary measure of bank capital ratio.

We also aggregate the data across US counties and at the national level to study regional and national lending patterns. We use a simple sum for shadow bank volume, total volume, bank volume, bank balance sheet volume, and bank sold volume. We use an annual weighted mean using bank volume as a weight for bank capitalization. We also use several county-level variables, including house price indices from the FHFA and unemployment rates from the BLS, to understand the broader economic context.

### 3. Motivating Facts

We start our analysis by presenting a set of facts that illustrate the importance of recognizing that banks are selling significant share of loans they originate (e.g., by securitizing them) and the increased role of shadow banks in the lending market. To organize our discussion of these two margins we use the following simple lending accounting framework.

#### 3.1 Lending Accounting Framework

##### 3.1.1 Balance Sheet Retention and the Loan Sales Multiplier

To analyze the loan sale multiplier, consider an amount of balance sheet lending by bank  $i$  at year  $t$ , *Bank balance sheet lending* $_{i,t}$ . This information can be inferred from the regulatory bank call reports or regulatory bank balance sheet data that measure the amount of lending by a bank in terms of loans it retains on its balance sheet. However, if a bank sells some of its loans in year  $t$ , its *Total bank lending* $_{i,t}$  in that year will be larger and given by:

$$Total\ bank\ lending_{i,t} = m_{i,t}^{Loan\ Sale} \times Bank\ balance\ sheet\ lending_{i,t} \quad (1)$$

where  $m_{i,t}^{Loan\ Sale}$  is the lending “multiplier” due to loans sales that equals

$$m_{i,t}^{Loan\ Sale} = \frac{1}{1 - Loan\ Sale\ Share_{i,t}} \quad (2)$$

---

<sup>8</sup> Obtained from <https://cdr.ffiec.gov/public/PWS/DownloadBulkData.aspx>

In the equation (2) *Loan Sale Share* $_{i,t}$  is the fraction of loans that bank  $i$  sells at time  $t$ . Aggregating across the banks we get that the total aggregate bank lending at time  $t$  is equal to

$$Total\ bank\ lending_t = m_t^{Loan\ Sale} \times Bank\ balance\ sheet\ lending_t \quad (3)$$

where *Bank balance sheet lending* $_t$  is the aggregate bank lending retained on balance sheet and  $m_t^{Loan\ Sale}$  is the aggregate lending multiplier due to loan sales.

### 3.1.2 Shadow Bank Lending Share and the Shadow Bank Lending Multiplier

To account for shadow bank lending outside of the traditional banking sector, we relate the total amount of lending in the economy at time  $t$  to total bank lending:

$$Total\ lending_t = m_t^{Shadow\ Bank} \times Total\ bank\ lending_t, \quad (4)$$

where  $m_t^{Shadow\ Bank}$  is the shadow bank lending multiplier that equals to

$$m_t^{Shadow\ Bank} = \frac{1}{1 - Shadow\ Bank\ Share_{i,t}} \quad (5)$$

and *Shadow Bank Share* $_{i,t}$  is the fraction of loans originated by shadow banks.

Combining equations (3) and (4), we get the correspondence between the aggregate amount of lending in the economy and the amount of aggregate bank balance sheet lending:

$$Total\ lending_t = m_t^{Shadow\ Bank} \times m_t^{Loan\ Sale} \times Bank\ balance\ sheet\ lending_t \quad (6)$$

Consider an example in which banks balance sheets reflect \$300 billion lending in a given year, but banks sell 50% of their loans and 50% of lending activity is done by shadow banks. In this case, both multipliers are equal to two, aggregate bank lending is equal to \$600 billion, and overall lending is equal to \$1.2 trillion.

The traditional bank balance sheet lending view corresponds to the case where both lending multipliers are equal to one:

$$m_t^{Loan\ Sale} = m_t^{Shadow\ Bank} = 1.$$

In this case, banks retain all their loans on their balance sheets, there are no shadow banks, and the total lending equals the bank balance sheet lending.

In the next section we document the magnitude of these multipliers, how they evolve over time and regions, as well as the forces shaping these multipliers. In the empirical part we emphasize the idea that these multipliers imply that measuring lending on bank balance sheets, or lending by

banks does not accurately represent aggregate lending, and that the inferences differ across regions systematically depending on county income. In the model section, we show that focusing on aggregate lending is insufficient to understand the speed of recovery from shocks or the stability of the banking system. Instead, the composition of financial intermediation and the associated multipliers are critical in determining the overall lending response.

### **3.2 Aggregate Lending: Multipliers are Large and Time Varying**

#### **3.2.1 Balance Sheet Retention and the Loan Sales Multiplier over Time**

We begin by showing that the propensity of banks to sell loans is large in magnitude and varies significantly over time. Figure 1(a) shows the fraction of banks that retain all their loans on their balance sheets, and therefore have a loan sale multiplier of one. Throughout our sample period, close to a half of banks in the US do not sell any residential mortgages they originate. However, these “traditional” banks constitute only about 4% of overall bank loan origination volume (Figure 1(b)), indicating that most bank lending activity occurs among banks that sell some of their loans.

Figure 2(a) shows that banks on average sell more than half of the loans they originate, with a mean loan sale propensity of 55%. Additionally, the loan sale propensity varied widely between 2005-2021, reaching a peak of 76% during the Great Recession and falling to a low of 38% during the 2018-2019 period before the pandemic (Figure A2 shows the banks’ loan sale propensity over a longer time period). This results in a loan sale multiplier displaying substantial variation over time that ranges from 1.6 to 4.2 depending on the year (Panel (b) of Figure 2). Figure 2(c) illustrates this further by showing the estimated loan sale multiplier in each year based on the estimation of equation (1) with individual bank data along with 95% confidence intervals.

Table 1, Column (1) shows that the aggregate bank balance sheet lending data accounts for less than half of the variation in total bank lending, at 46%. Adding bank loan sales in Column (2), by definition, it results in an r-squared coefficient of 100%. These results highlight the importance of recognizing that the propensity of banks to sell loans is large in magnitude and *varies* over time. Figure 4, panel (a) illustrates this in a simple way by showing the extent of inference errors resulting from a failure to recognize that bank loan sale propensity significantly changes over time. This figure plots an actual total bank lending volume and inferred total bank lending volume if one erroneously assumes a *constant* loan sale multiplier equal to its sample mean to infer the total bank lending from the bank balance sheet lending volume. As we observe there are significant differences between the inferred and actual aggregate banking lending ranging from the inferred bank lending underestimating the actual bank lending by close to \$600 billion in 2009 to the inferred bank lending overestimating the actual bank lending by close to \$800 billion in 2021. In

Column (1) of Table 4(a) we illustrate this more formally by regressing the true total bank lending on the inferred total bank lending from bank balance sheet lending assuming a constant loan sale multiplier. As we observe, the inferred bank lending accounts for only about 44% of variation in the actual aggregate bank lending. Overall, these results suggest that a simple mapping from bank balance sheet data to aggregate lending outcomes is highly imperfect, and as we later show, depends on economic conditions.

### **3.2.2 Shadow Bank Lending Share and the Shadow Bank Multiplier**

Even perfect bank data is insufficient to evaluate the evaluation of lending activity because the shadow bank multiplier is also large and varies significantly over time. Panel (a) of Figure 3 illustrates the proportion of residential mortgages originated by shadow banks along with the shadow bank loan multiplier. As can be seen, shadow banks originate a considerable share of loans, and their inclination to sell loans significantly increases over time, from about 30% in 2005 to about 60% by 2021 (Figure A2 shows the shadow bank market share over a longer time period). The associated shadow bank multiplier increases from a low of 1.3 in 2007 to 2.4 in 2021, indicating that by 2021 for every dollar of bank-originated loan, there is about 2.4 dollars of total lending accounting for lending done by both banks and shadow banks (Figure 3, panel b). Because shadow bank market share is time varying, bank lending data—including loans sales—accounts for only about 68% variation in total lending (Column (3) of Table 1).<sup>9</sup> The variability in the shadow bank multiplier makes it difficult to infer aggregate lending from just traditional bank information. Figure 4, panel (b) illustrates this in a simple way by showing the extent of inference errors on total lending volume if one uses a *constant* shadow bank lending multiplier (equal to its sample mean) to infer the total lending volume from aggregate bank lending volume (broadly defined, including loan sales). As we observe there are significant differences between the inferred and the true aggregate total lending level ranging from the inferred lending overestimating the actual lending by close to \$500 billion in 2009 to the inferred lending underestimating the actual total lending by more than \$1.4 trillion in 2021. In Column (2) of Table 4(a) we illustrate this more formally by regressing the actual total lending on the inferred lending volume from the bank balance sheet lending assuming a constant shadow bank lending multiplier. As we observe, the inferred total lending accounts for only about 67% of variation in the actual total lending. Overall, these results illustrate that a simple mapping from bank data to aggregate lending is highly imperfect, and as we later show, depends on economic conditions.

### **3.2.3 Aggregate Lending Growth Rates**

---

<sup>9</sup> Column (4) shows that the addition of shadow bank lending volume mechanically explains 100% of the variation.

The variation in the multipliers is also visible in lending growth rates in Figure 5, which would be uniform under constant multipliers. For instance, in 2007, bank balance sheet lending increased by 6%, respectively, while total bank lending and overall lending, including shadow banks, decreased by, respectively 5% and 20%. In contrast, in 2009, bank balance sheet lending declined by 13%, while total bank lending and total lending increased by 22% and 28%, respectively. More systematic analysis in Table 3 illustrates that the aggregate bank balance sheet lending growth rate accounts only for about 49% variation in the aggregate lending growth rate of banks (Column 1). Including growth in bank loans sold results in an r-squared equal to 97% (Column 2).

Column (3)-(4) of Table 3 show the similar analysis for the total lending growth rate. Bank lending growth rate including loan sales accounts for about 93% of variation in total lending growth rate. Recognizing in addition the shadow bank lending growth allows us to explain almost all variation in the total bank lending growth rate with an r-squared equal to 99% (Column 4).

### **3.3 The Impact of Bank Loan Sales and Shadow Banks on Regional Lending Patterns**

Here we show that there are substantial differences in the bank loan sale propensity and shadow bank market share across U.S. regions. We illustrate that this regional heterogeneity further complicates usage of bank balance sheet data, or even data on bank lending, to learn about overall regional lending patterns. We then show that regions (counties) with different income levels are substantially different in the loan sale and shadow bank multipliers. As we argue later these differences also result in differential shock propagation across the regions.

We start by visually illustrating the significant heterogeneity in the bank loan sale propensity and shadow bank market share across US counties in Figure 6. This heterogeneity together with a time-varying nature of these factors implies a substantial regional variation in the loan sale and shadow bank multipliers across and within regions. Only about 53% of variation in the county-level lending of banks is explained by bank balance sheet lending (Table 1, Column 5). There are also regional differences in how multipliers change over time. Bank balance sheet lending, together with county-level and year fixed effects, only explains roughly 74% of the variation in county-level bank lending volume (Table 1, Column 6). This implies that even after considering level differences in the loan sale multiplier across counties (county fixed effects) and aggregate yearly variation in the loan sale multiplier (year fixed effects), the within-region time variation of the loan sale multiplier is large enough to leave more than a quarter of the variation in total bank lending unexplained. Similarly, shadow bank multipliers also differ substantially across regions: bank lending (balance sheet and loan sales) accounts for about 54% of variation in the county-level total lending volume, which increases to about 86% once we add county and time fixed effects. The latter are a result of an aggregate shift to shadow banks, but even with this aggregate variation

accounted for, significant variation in the shadow bank multiplier exists. Variation in the multipliers is also reflected in regional growth rates in lending, which we confirm in Table 3 and Figure 7. This analysis shows that grossing up bank balance sheet lending by a fixed multiplier is insufficient for making precise quantitative statements about the total quantity of lending.

To illustrate this further in Panel (b) of Table 1 we regress the true total bank lending at the county-level on the inferred total bank lending from bank balance sheet lending assuming a constant loan sale multiplier based on its sample mean in the aggregate data. As we observe, the inferred bank lending accounts for only about 53% of variation in the actual aggregate bank lending at the county level (Column 1), and with county fixed effects for just 63% of that variation (Column 2). Similarly, as shown in Column (5) and (6) the inferred total lending from bank lending assuming a constant shadow bank multiplier accounts for only about 49% of variation in the total lending at the county level (59% with county fixed effects). Even allowing the multipliers to vary as in the aggregate data and including county-level fixed effects does not significantly improve the inference on county-level lending patterns. For example, using this method the inferred total lending just accounts for about 74% of variation of actual lending at the county-level. This evidence confirms that not only multipliers vary over time and across regions but also that these region-level differences evolve over time in a nontrivial way complicating further the inference of regional lending patterns from just bank balance sheet or bank data.

In Figure 8, we show that banks in counties with the highest income are most likely to keep their loans on their balance sheets (panel a). As Buchak et al. (2018) show, this was in part a reflection of tightened regulation on banks post 2007 crisis that led to banks retreating from lending to low-income households. In addition, because high income households tend to have jumbo loans (larger loans), lack of securitization market for such loans in the post 2007 period implies that in counties with such households, the loan sale and shadow bank multipliers are smallest. Because shadow bank entered in sectors where banks retreated and where market for securitization existed, the share of shadow bank lending monotonically decreases with income (Figure 8, panel b), resulting in the largest shadow bank multiplier for the poorest counties. As we later discuss, the heterogeneity in multipliers across space implies that recovery from shocks will differ across regions of different incomes.

#### **4. Bank Capitalization and Secondary Markets Drive the Margins of Adjustment**

Buchak et al. (2022) have identified two key drivers behind the loan sales multiplier and shadow bank multiplier. Firstly, banks switch between traditional bank balance sheet lending and selling loans based on their balance sheet strength and the extent of that switching is determined in equilibrium. Panel (a) of Figure 9 shows that the bank loan retention propensity is lower (loan sale

propensity is higher) when the aggregate bank capitalization is lower. We show that the relationship between bank capital and share of loans retained on balance sheets holds across banks (Figure 9, Panel (b)) as well as within a bank across time and accounting for loan characteristics (Figure 9, Panel c). These results extend the results from Buchak et al (2022). As banks' balance sheet capacity declines, banks shift towards the originate-to-distribute model and then move back towards balance sheet lending as their balance sheet capacity improves.

Second, both multipliers—including the shadow bank multiplier, given that their business model relies on selling loans—also depend on the availability and relative attractiveness of the loan sale market. For example, policies that acquire mortgage-backed securities, such as quantitative easing, can lower the cost of capital for the originate-to-distribute model, resulting in increased loan sale propensity by banks. Because shadow banks do not originate on the balance sheet, the shadow bank lending multiplier is also crucially affected by the conditions in the secondary-loan market (Buchak et al. 2022).

Overall, our empirical evidence shows that the shadow bank multiplier and the loan sale multiplier are large and evolve over time as a function of the composition of the financial intermediation system: balance sheet capacity of banks and the presence of shadow banks. These empirical findings have two direct consequences. The first is on measurement. Bank data are frequently used to measure how lending responds to policy, financial or real shocks or to calibrate models in which lending plays a central role. The time and regional varying multipliers suggest that bank balance sheet lending does not accurately represent bank lending, which in turn does not accurately represent aggregate lending. Importantly, simply grossing up bank balance sheet lending by a fixed, common multiplier across time or regions is insufficient to both measurement and policymaking.

Additionally, bank capitalization and the presence of shadow banks are major determinants of these multipliers. Because policy frequently impacts bank capitalization, these multipliers are therefore not policy invariant. For example, imposing a stricter capital requirement on banks is likely to have a large impact on *bank balance sheet* lending, but also simultaneously alter banks' incentives to sell their originated loans and shadow banks' competitiveness with bank lenders. Therefore, understanding only the bank balance sheet lending response to capital requirement changes is insufficient for evaluating its impact on total lending.

The second implication of our empirical exercise is that the severity and recovery from shocks behaves differently than predicted by models in which all lending is through bank balance sheets. These heterogeneous responses across the U.S. are difficult to gauge by looking at just the lending on bank balance sheets.

## 5. Model

Our empirical results document that the loan sale and the shadow bank lending multipliers are not only large, but also highly variable and endogenous to many relevant policies. Thus, a realistic quantitative policy analysis requires modeling how these multipliers evolve in equilibrium. To illustrate how these multipliers affect policy analysis, we develop a simplified parsimonious dynamic quantitative model of financial intermediation. Our model builds on Buchak et al. (2022), which estimates a rich heterogeneous agent demand system for bank and shadow bank loans in which the loan sales and shadow bank multipliers play a central role, but in a static setting. Here, we instead focus on dynamics to understand how these multipliers contribute to the recovery from financial shocks over time, and how they affect banks' ex-ante incentives to retain capital buffers. This allows us to compare the recovery dynamics relative to standard bank balance sheet models of financial intermediation, which omit these margins.

We calibrate the model to the empirical loan sales and shadow bank lending multipliers. Using the calibrated model, we examine how shocks to bank capital propagate through this augmented model of the financial intermediation sector. Importantly, we counterfactually examine economies where these margins of adjustment are absent. Our model highlights the impact of these margins on how shocks to bank capital propagate, as well as the speed of recovery from shocks to intermediaries. Ex-post, these margins ameliorate shocks to bank capital both from the perspective of the initial impact of the shock, and the speed of recovery from the shock. We then use our model to illustrate where calibrating a more standard bank balance sheet model leads to incorrect inferences.

### 5.1 Model Specification

We first provide a high-level overview of the model. Banks compete with non-banks in imperfect competition to provide loans that mature in the following period. Banks have capital and are long-lived. Bank capital gives banks the ability to make on-balance sheet loans, but regulatory capital requirements impose a severe penalty when bank capital falls below a statutory minimum. Additionally, banks can choose to finance their loans through securitization in a secondary market. Banks make loan pricing, financing, dividend, and equity raising decisions to maximize their discounted present value.

Non-banks have no capital and cannot make balance sheet loans. Therefore, they must finance all originations through securitization. Non-banks set interest rates to maximize profits. By assumption, lenders are symmetric within type.

Lenders' loans are imperfect substitutes for one another. This captures unmodeled horizontal product differentiation such as, differences in lenders' branch networks or other amenities offered to borrowers. This product differentiation gives individual lenders market power, and consequently there are positive markups and variable profits in equilibrium. Lenders therefore earn (variable)

rents not because, e.g., an incentive compatibility constraint, but because of well-documented imperfect competition. Loan origination has a fixed per-loan “labor” cost, and additionally must be financed through balance sheet retention or securitization for banks, or through securitization for non-banks. Securitization has a higher direct marginal cost, but balance sheet financing negatively impacts banks’ capital ratios.

### 5.1.1 Loan Demand

Given a vector of economy-wide interest rates  $\mathbf{r}_t \equiv \{r_{1t}, r_{2t}, \dots, r_{Nt}\}$  of  $N$  lenders at time  $t$ , total demand for lender  $i$ ’s loans is given by

$$q_{it} \equiv q_i(r_{it}, \mathbf{r}_{-it}; \theta) \quad (\text{M1})$$

Where  $\theta$  are non-price characteristics, such as the lender type. Quantities are expressed in units of aggregate bank risk-weighted assets, e.g., a quantity of 0.10 means that the bank’s flow lending is equal to 10% of its risk-weighted assets.

### 5.1.2 Non-bank Loan Supply

A fixed<sup>10</sup> number of non-banks,  $N_{sb}$ , compete with a fixed number of banks,  $N_b$ . Non-bank loan provision has a labor cost  $mc_{sb}^l$  and a cost of securitization  $mc^s$ , so that non-bank marginal cost is equal to  $mc_{sb} = mc_{sb}^l + mc^s$ . Taking other interest rates  $\mathbf{r}_{-j}$  as given, non-banks maximize profits as follows:

$$\pi_{sb} = \max_r q_i(r, \mathbf{r}_{-it}; \theta)(r - mc_{sb}) \quad (\text{M2})$$

In equilibrium because loans are differentiated, the variable markup  $(r - mc_{sb})$  will be positive.

### 5.1.3 The Bank’s Problem

A representative long-lived bank’s capital ratio is given by  $c_t$ . In each period, it decides whether to raise equity  $I_t$  (pay dividends) thereby directly increasing (decreasing) its capital ratio. Additionally, it sets an interest rate on loans  $r_{it}$  and a financing policy  $\phi_{it} \in \{0,1\}$ , where  $\phi_{it} = 1$  means the bank retains the loan on balance sheet.

**Bank investment:** The bank can raise equity ( $I_t > 0$ ) or issue dividends ( $I_t < 0$ ), i.e.,  $I_t$  represents net investment. In order to contribute  $I_t$  of equity the banker must pay an effective cost of  $\psi(I_t)$ , with  $\psi(I_t) > I_t$ . That is, the banker pays a cost both when raising equity and when receiving dividends. We assume a convex cost of issuing dividends and a fixed cost of raising equity. Bank

---

<sup>10</sup> Our model can easily accommodate non-bank entry, but for simplicity we omit it in the baseline model. The effect of allowing entry is to increase the aggregate elasticity of loan supply, meaning that quantities adjust more, and prices adjust less to shocks.

investment takes time, so that investment at time  $t$  only impacts the bank's capital ratio at time  $t + 1$ .

**Bank loan supply:** Like non-banks, each bank offers a differentiated loan to borrowers. Bank loan provision has a marginal labor cost,  $mc_b^l$ . If the bank securitizes the loan, it additionally pays a marginal securitization cost, equal to the non-bank's securitization cost,  $mc^s$ .

Alternatively, the bank can finance the loan by retaining it on its balance sheet. We capture the fact that the loan is a long-lived asset through the assumption that on-balance-sheet loan origination negatively affects the bank's capital ratio at the time it is originated, and pays out—thus increasing the bank's capital ratio—only in the following period. Loans receive a regulatory risk weight  $\xi$ , so that when originating quantity  $q$  of on-balance-sheet loans reduces regulatory capital by  $\xi q$ .

Bank profit is as follows:

$$\pi_b(r, \phi) = q_i(r, \mathbf{r}_{-it}, \phi; \theta)(r - mc_b^l - (1 - \phi)mc^s) \quad (\text{M3})$$

As with non-banks, because loans are differentiated, in equilibrium, bank variable markups will be positive.

**Bank capital:** We model directly the bank's capital ratio, which follows the following law of motion:

$$c' = \exp(z) c + I + \pi_b(r, \phi) \quad (\text{M4})$$

$c$  is the current period capital ratio,  $c'$  is next period's capital ratio,  $z$  is a shock to bank capital,  $I$  is net investment, and  $\pi_b$  is profits from lending, defined above. We assume that  $z$  follows an exogenous Markov process, for example, an AR(1) with some persistence.

Note that the quantity of on-balance-sheet lending does not appear directly in the law of motion for bank capital, because the one-period loans mature in the following period and become cash, which has a risk-weight of zero.

The bank is subject to a regulatory capital requirement, which we model as a severe penalty to flow utility. The penalty is assessed at the end of the period, after the shock  $z$  has realized, after the firm has made its investment decision, after the firm has made its lending rate and financing choice, but importantly, *before* the firm has realized any profits from on-balance sheet lending. Thus, its effective capital at the time of the regulatory assessment is:

$$c^{eff}(r, \phi) = \exp(z) c - \xi q_b^b(r) \phi \quad (\text{M5})$$

We denote with  $\rho(r, \phi) = f(c^{eff}(r, \phi))$  the regulatory penalty. Finally, we assume that the bank, when making decision to retain or securitize its loans, receives an independent utility shock  $\epsilon_{ret}$  and  $\epsilon_{ota}$ , which captures in reduced form heterogeneity in bank capital across banks. Thus, its period utility is as follows:

$$u(r, I, \phi) = -\psi(I) + \phi(\rho(r, 1) + \epsilon_{ret}) + (1 - \phi)(\rho(r, 0) + \epsilon_{ota}) \quad (\text{M6})$$

Recall, for positive net investment,  $I > 0$ ,  $\psi(I) > 0$ , i.e., the banker gets disutility from putting more equity into the firm. Conversely, for negative net investment (paying dividends),  $I < 0$  and  $\psi(I) < 0$ , i.e., the banker receives utility from dividends.

**The banks' problem:** With these ingredients defined, we express the bank's problem as follows:

$$\begin{aligned} v_0 &= \max_{\{r, \phi, I\}_t} \sum_{t=0} \beta^t E [u(r, I, \phi)] \\ \text{s.t.} \quad c_{t+1} &= \exp(z_t) c_t + I_t + \pi_b(r_t, \phi_t) \end{aligned}$$

#### 5.1.4 Equilibrium and model solution

Equilibrium in our model is a set of policy functions  $\mathbf{r}(c, z)$ ,  $\phi(c, z)$ , and  $I(c, z)$ , which depend on the state variables  $(c, z)$  and satisfy:

1. Banks choose rates, loan retention, and net investment to maximize lifetime utility
2. Non-banks choose rates to maximize per-period profits
3. Loan demand equals loan supply

**Functional forms and parameters:** For quantification, we impose functional form assumptions on the general components of the model.

We assume that each lender  $i$  faces logistic loan demand:

$$q_i(r_i, \mathbf{r}_t; \theta) = m \times \frac{\exp(-\alpha r_i + \delta_i)}{\sum_j \exp(-\alpha r_j + \delta_j)} \quad (\text{M7})$$

This form arises naturally out of the standard IO discrete choice framework where a mass  $m$  individual borrowers make a discrete choice among  $N$  lenders and an outside option of not borrowing.  $\alpha$  determines how sensitive borrowers are to interest rates. We assume that non-price

attributes (e.g., amenities or convenience of the lender),  $\delta_i$ , is common across lender types, so that banks have non-price attribute  $\delta_b$  and non-bank lenders have a non-price attribute  $\delta_s$ .<sup>11</sup>

We assume banks have a convex cost adjustment cost for raising equity and issuing dividends, plus a fixed cost when issuing equity, which captures in reduced form both underwriting costs as well as financing frictions around issuing equity. In particular, we specialize  $\psi(I)$  as:

$$\psi(I) = I + \frac{\gamma}{2}I^2 + \begin{cases} 0 & I \leq 0 \\ C & I > 0 \end{cases} \quad (\text{M8})$$

$\gamma$  determines the convexity of equity or dividend issuance costs.  $C$  determines the fixed cost of raising equity.

We define the regulatory cost function,  $\rho(x)$  as:

$$\rho(x) = \begin{cases} 0 & x \geq \bar{c} \\ -\exp(-\lambda \times (x - \bar{c})) - \lambda \times (x - \bar{c}) + 1 & x < \bar{c} \end{cases} \quad \text{when} \quad (\text{M9})$$

This formulation imposes no regulatory cost above a threshold,  $\bar{c}$ , an exponentially increasing cost below the threshold, that is continuous in level and derivative first at the threshold.  $\lambda > 0$  controls how quickly regulatory costs rise for an out-of-compliance bank; a lower  $\lambda$  corresponds to a more forbearing regulator.

For the exogenous shocks, we assume that shocks to capital  $z_t$  follow an AR(1) process,  $z_{t+1} = \theta z_t + \epsilon_t^z$ ,  $\epsilon_t^z \sim N(0, \sigma_z^2)$ . Finally, the securitization and retention shocks  $\epsilon_{ota}$  and  $\epsilon_{ret}$  follow a type-1 extreme value distribution with scale parameter  $\sigma_{fin}$ .

**Characterizing the solution to the banks' problem:** To provide intuition, the bank's problem can be separated into a static problem and a dynamic problem.

*The static problem:* In the static problem, the bank takes as given its current state,  $(c, z)$  and a candidate next-period capital  $c'$ . Conditional on its current state and desired next-period capital, the bank chooses interest rates, a financing policy, and an investment policy such that its state transition is feasible, and its current period flow utility is maximized. In particular, the bank's intra-temporal problem is to maximize:

$$u^*(c, z, c') \equiv \max_{r, \phi, I} E[u(r, I, \phi)] \quad (\text{M10})$$

---

<sup>11</sup> In particular, this functional form arises when borrower  $b$  chooses among  $i = \{0, 1, \dots, N\}$  alternatives each with indirect utility  $u_i = -\alpha r_i + \delta_i + \epsilon_i$  where  $\epsilon_i$  is a type-1 extreme value shock. We normalize the indirect utility of the outside option to  $u_0 = 0$ .

Subject to the law of motion,

$$c' = \exp(z) c + I + \pi_b(r, \phi) \quad (\text{M11})$$

Letting  $\mu$  be the Lagrange multiplier on the capital law of motion, one has, if the bank securitizes the loan:

$$\psi'(I) = \mu \quad (\text{M12})$$

$$\pi'_b(r, 0) = 0 \quad (\text{M13})$$

Observe that the bank's price-setting decision when originating to distribute does not depend on its capital; it simply aims to maximize profits. In contrast, if the bank retains the loan:

$$\psi'(I) = \mu \quad (\text{M14})$$

$$\xi(q_b^b)'(r) \rho'(c^{eff}) + \mu \pi'_b(r, 1) = 0 \quad (\text{M15})$$

That is, the bank's balance sheet pricing decision depends additionally on how the newly originated loans will impact its regulatory cost. When  $c^{eff}$  is far above the regulatory constraint,  $\rho'(c^{eff}) = 0$ , and hence this term drops out. In this case, the bank is again setting interest rates to maximize profits. In contrast, when  $c^{eff}$  is relatively low,  $\rho'(c^{eff}) > 0$ , and the bank will choose higher rates to offset the increased regulatory cost. Observe that in order for the static plan to be feasible,  $I|_{\phi=0} \neq I|_{\phi=1}$ , because balance sheet and securitized lending generate different levels of retained profit for the bank, which changes how much investment must be made (dividends paid out).

Finally, given the functional form assumption on the retention shocks, the bank's retention policy takes the following logistic form in expectation, which we interpret as the share of heterogeneous banks following the retention policy.

$$\phi = \frac{\exp\left(u(r|_{\phi=1}, I|_{\phi=1}, 1) - u(r|_{\phi=0}, I|_{\phi=0}, 0)\right)}{1 + \exp\left(u(r|_{\phi=1}, I|_{\phi=1}, 1) - u(r|_{\phi=0}, I|_{\phi=0}, 0)\right)} \quad (\text{M16})$$

The retention policy depends on the difference in flow utilities for retaining versus securitizing the loans. The intuitive tradeoff is as follows: Balance sheet lending has lower marginal costs, because the bank does not have to pay the cost of securitization. Thus, profits from balance sheet lending are higher. This means that in order to maintain a given target level of capital for the next period,

the bank can pay out a larger dividend. When the bank is well capitalized, it will prefer to do this. In contrast, when the bank is poorly capitalized, balance sheet lending leads to a greater regulatory cost, and the value of avoiding this regulatory cost outweighs the benefits of the larger dividend (smaller equity issuance) that balance sheet lending would enable.<sup>12</sup>

*The dynamic problem:* The static problem produces, for each state  $(c, z)$  and candidate next-period capital  $c'$ , an optimized flow utility  $u^*(c, z, c')$ . Given this function, the bank's problem can be written as a straightforward dynamic optimization problem where the bank chooses the optimal next-period capital in each state. In recursive form, this is:

$$v(c, z) = \max_{c'} u^*(c, z, c') + \beta E[v(c', z')] \quad (\text{M17})$$

Once the optimal dynamic policy function,  $c' = g(c, z)$  is solved, the optimal interest rate, retention policy, and investment policy can be recovered directly from the solution to the static problem described above.

## 5.2 Calibration

The model with specialized functional forms has parameters that we calibrate in order to produce quantitatively reasonable simulations and counterfactuals. Several of the parameters correspond to values available directly from the literature, which others do not. Where possible, we take values directly from the literature or from regulations. We calibrate the remaining parameters via simulated method of moments.

### 5.2.1 Parameters calibrated from existing literature

We set the subjective discount rate,  $\beta$ , to 0.95. Following Buchak et al. (2022), we set the regulatory risk weight on loans to be  $\xi = 0.25$ , which corresponds to the statutory requirement for conforming loans. Also following Buchak et al., we set marginal costs as  $mc_b^l = mc_s^l = 3\%$ , which incorporates both labor costs and a “baseline” cost of capital averaged over the post-crisis period, and  $mc^s = 0.67$ . We set  $\alpha = 1.65$ , which is the average price sensitivity estimated over the US mortgage market. We set  $N_b = 25$  and  $N_{sb} = 50$ , which corresponds roughly to the average number of bank and non-bank lenders in a given MSA.

---

<sup>12</sup> For computational tractability, to solve the model, we currently calculate optimal interest rates in an “all balance sheet retention” regime, and an “all securitization” regime, and taking these as given and constant across each  $(c, z, z')$  tuple, calculate the necessary investment policy and the optimal choice probability  $\phi$ . Aggregate quantities, rates, and investment are then a weighted average of the all balance sheet retention regime and the all securitization regime, weighted by  $\phi$ . The effect of this simplification is to essentially turn off second-order competitive responses to banks offering higher interest rates when poorly capitalized. That is, in equilibrium, seeing that other banks are offering higher interest rates, competing banks and shadow banks would slightly raise their interest rates. This channel is therefore absent in the currently-calibrated version of the model.

### 5.2.2 Parameters calibrated through the simulated method of moments

There are several remaining parameters to calibrate. While the calibration is done jointly through the simulated method of moments, it is useful to describe intuitively which moments are most informative about which parameters.

The parameters to be calibrated first concern the capital adjustment cost function:  $\gamma$ , the investment convexity parameter,  $C$ , the fixed cost of raising equity. Next, parameters of the regulatory cost function,  $\bar{c}$  and  $\lambda$ , and parameters of the exogenous shock process,  $\theta$  and  $\sigma_z^2$ . These parameters most directly influence time-series properties of the capital ratio process, and intuitively we exploit various historical moments to inform them. In particular, we use the historical mean, standard deviation, and autocorrelation of levels and changes in the aggregate capital ratio (6 moments) to calibrate these six parameters.

Next, parameters of the demand function,  $m$ , the market size, and  $\delta_b$  and  $\delta_{sb}$ , the non-price demand characteristics of banks and non-banks, respectively. These parameters broadly concern aggregate lending quantities as well as means and higher-order moments of bank market shares, or, equivalently, the shadow bank multiplier. In particular, we use four moments to calibrate these three parameters: The historical average flow of lending relative to bank assets (a measure of lending quantities), the mean and standard deviation of the shadow bank multiplier, and the correlation of total lending with bank capitalization.

Finally, the scale parameter of the retention utility shock,  $\sigma_{fin}$ , broadly governs the level and volatility of the loan sales multiplier. We calibrate this parameter using the mean and variance of the loan sales multiplier, together with the correlation of bank balance sheet shocks to lending quantity shocks.

Observe that our calibration is overdetermined, making use of 13 moments to calibrate 10 parameters. Despite this overdetermination, our calibration is able to reasonably match the targeted moments. Table 4, Panel (a) shows the targeted moments used in calibration and the moments produced by the model. Panel (b) shows the calibrated parameters and summarizes the key moments used in identification. Panel (c) shows the parameters taken from the literature.

### 5.3 Model Discussion

Our model captures the two key multipliers, the loan sales multiplier and the shadow bank multipliers, which we emphasize in our reduced-form results. Both margins relate directly to the key state variable, the level of bank capital. To help illustrate these forces, Figure 9 plots the bank's policy functions against the level of bank capital. We focus first on the baseline case (solid yellow line) in which both margins are present.

When the bank is well-capitalized, the bank's balance sheet provides the lowest-cost source of financing in the economy. Shadow banks, which must finance through more expensive securitization, are at a funding disadvantage relative to banks. Intuitively, banks can replicate the funding of shadow banks, but shadow banks cannot replicate the funding of banks. While product differentiation and imperfect competition lead shadow banks to have a non-trivial market share, the fraction of loans that shadow banks originate is relatively low because they must pass their higher marginal costs on to borrowers. Banks, facing the choice between low-cost balance sheet financing and higher-cost securitization, tend to choose to retain loans on balance sheet (Figure 10 Panel (a)). Implicit bank capital cost heterogeneity implies that this fraction is not 100%. Additionally, well-capitalized banks can afford to issue dividends (negative net investment, shown in Figure 10 Panel (b)). As low balance sheet funding costs are partially passed on to borrowers, lending rates are relatively low, and lending quantities are relatively high (Figure 10 Panels (c) and (d)).

As bank capital deteriorates, the shadow cost of on-balance sheet financing increases. Additional on-balance sheet loan originations reduce bank capital and push the bank closer to the region, in which capital regulation imposes significant costs on the bank. As bank balance sheet financing costs rise, the bank endogenously begins to substitute towards an originate-to-distribute model, as shown in Panel (a) of Figure 10. Because the bank is forced to substitute towards higher-cost sources of financing and pass these costs on to borrowers, bank lending quantities fall, as shown in Panels (c) and (d). Finally, more poorly capitalized banks lower dividends or even raise equity, as shown in Panel (b).

To quantify the role of the loan sales and shadow bank multipliers, we re-solve our model sequentially turning off these margins, while still using the same calibrated parameters. The *No Balance Sheet Margin* counterfactual removes the loan sales margin so banks only originate loans on balance sheet, but maintains the presence of shadow banks. The *No Shadow Banks Margin* counterfactual removes the shadow bank margin and but allows banks to substitute between on balance sheet and securitized lending.. The *Neither Margin* counterfactual removes both, which corresponds to the pure bank balance sheet model of lending with no loan sales by banks and no shadow bank lending. Figure 9 shows these counterfactual policy functions in narrow-dashed blue, wide-dashed green, and dotted red lines, respectively.

First, Panel (a) of Figure 10 shows that mechanically, under the *No Balance Sheet Margin* and *Neither Margin* scenarios, banks retain 100% of their loan originations, because these counterfactuals assume that banks must retain all their loans. Removing the shadow bank margin in the *No Shadow Banks Margin* scenario, does not meaningfully alter the bank's optimal financing decision because that decision is essentially made conditional on having made the loan and does not depend much on the broader competitive environment.

The ability to sell loans lowers bank profits and thus dividends when capital is plentiful, but also allows them to maintain higher dividends during recapitalization when capital is low. Panel (b) shows the bank's optimal investment (dividend) policy under the counterfactuals. Holding the bank balance sheet margin fixed, shadow banks presence uniformly decreases dividends that banks pay. Because the lending sector is more competitive and bank profits are lower, banks have less to pay out in steady state. Next, holding shadow bank presence fixed, in the well-capitalized states, dividend payouts are lower in the presence of the loan sales margin. This occurs because of product differentiation: Due to bank heterogeneity, even when the sector as a whole is well-capitalized, some banks still choose to engage in an originate-to-distribute business model. This model is less profitable than balance-sheet lending from a well-capitalized bank, and thus total banking profits, and consequently dividends, are lower. In contrast, in the poorly capitalized states, in the absence of the loan sales margin, banks cut dividends much more sharply and switch to net investment sooner. This occurs for two reasons: First, because banks have no option to switch to loan sales (securitized lending), higher marginal costs reduce their profits, so they are less able to issue dividends. Second, dynamically, because these banks rely more on capital to engage in profitable lending, retained capital is more valuable so they are quicker to recapitalize.

Giving banks the option to sell loans does not always lead to lower average interest rates, despite the fact that it allows each bank to lower its funding cost. Panels (c) and (d) of Figure 10, consider first the case of well-capitalized banks. Some banks, even though the sector is well-capitalized, choose to engage in securitized lending, which has higher marginal costs. Because of product differentiation, these higher-cost lenders maintain non-trivial market share and raise prices in equilibrium, increasing the average interest rate in the economy. When the banking sector is poorly capitalized, on the other hand, allowing banks to securitize loans decreases average rates. In this scenario, the shadow cost of balance sheet lending increases dramatically. While the costs of securitized finance are greater than those of balance sheet lending for a well-capitalized bank, they are lower than the costs for a poorly capitalized bank. Some of the cost savings are passed on to consumers, resulting in lower average rates when banks are poorly capitalized.

The largest effects are on quantities of lending, with dramatically different sensitivities of total lending to bank balance sheet capital when these margins are active versus when they are not. When banks cannot sell originated loans, around the point in the state space where bank capital becomes impaired, aggregate lending declines dramatically. In contrast, when banks can sell these loans, the aggregate lending response is much more muted. Thus, a regulator observing a relatively poorly capitalized banking sector thinking only about balance sheet lending would incorrectly and dramatically overstate the impact of further deterioration of bank capital on total lending.

To summarize, the presence of the loan sales and shadow bank margins have the effect of moderating outcomes in the loan market when bank capital deteriorates. When these margins are not present, bank costs rise as capital falls, and these costs are passed through to borrowers, leading

to higher prices and lower quantities. When these margins are present, banks shift their financing business model towards securitization. Costs rise, but much less dramatically. Additionally, shadow banks provide an important source of lending particularly when bank capital is low and costs are relatively high.

## 5.4 Capital Shocks under Counterfactual Financing Models

We illustrate the quantitative importance of these margins in the recovery from a negative shock to the intermediation sector. We simulate a negative shock to bank capital in four economies: (1) the *baseline* economy with both margins active, (2) the *No Shadow Banks* economy, with no shadow banks but the loan sales margin still active, (3) the *No Balance Sheet* economy, with no loan sales but shadow banks still present, and (4) the *neither margin* economy, where both margins are shut down. We draw exogenous capital shocks  $z$  from the calibrated distribution except at time  $t = 0$ , for which we impose a large deterministic negative capital shock. We run the simulations 500 times and examine average outcomes across each of the four economies and plot the implied impulse response functions from these simulations. The results are shown in Figure 11 Panels (a)–(e).

### 5.4.1 Bank Capital Response

Here we show that the loan sales margin substantially increases the resilience of the financial intermediation sector to bank capital shocks *ex post*, but also results in less prudent banks *ex ante*. We describe the path of bank capital across the four counterfactuals, which is shown in Figure 11, Panel (a). The shock at time  $t = 0$ , dramatically (and mechanically) decreases capital for banks in each counterfactual.

Ignoring the margins of adjustment substantially underestimates the resilience of the financial intermediary sector to shock to intermediary capital. When capital falls below the statutory minimum, banks begin to recapitalize through both a combination of retained earnings (i.e., decreased dividend payouts) as well as direct investment. Following the shock, bank capital slowly rebuilds. Our counterfactuals do not incorporate any regulatory interventions which would increase banks' capital—all capital increases are voluntary given the required capital ratio. When banks can adjust on balance sheet retention margin, they rebuild capital substantially faster because the bank can still engage in profitable lending off balance sheet. When banks are restricted on balance sheet lending the recovery is slow: while their incentives to rebuild capital are high, their limited capital prevents them from generating substantial retained earnings.<sup>13</sup>

*Ex ante*, because capitalization stocks are less costly, banks are less prudent with capital, and take larger advantage of balance sheet capacity when lending. In other words, when banks have access to securitization, they are willing to lend *more* on balance sheet for any given level of capital and

---

<sup>13</sup> See Kashyap and Stein (2004) for a discussion of the cyclical implications of bank capital standards.

capital requirements. More formally, without the ability to substitute, banks keep a 3pp larger capital buffer away from the regulatory constraint. One can see this by the higher steady-state level of capital in economies without the loan sales margin (the *No Balance Sheet* economy and the *Neither Margin* economy). Intuitively, when the bank has a more difficult time adjusting to shocks, it keeps a larger capital buffer in excess of the statutory capital requirement. This implies that for a given level of capital requirements and capital banks' access to securitization allows them to take larger advantage of their balance sheet capacity.

#### 5.4.2 Lending Price and Quantity Response

Panels (b) and (c) show interest rates and lending quantities relative to the average pre-shock level in each economy. A common feature across all counterfactuals is that well-capitalized bank balance sheets provide the lowest cost loan financing. As bank balance sheets are impaired, balance sheet financing becomes costlier, and this is passed through to borrowers in the form of higher rates and consequently lower quantities. Whether these costs are passed through to borrowers through higher prices or lower quantities depends primarily on borrowers' demand elasticity.

While the direction of the effect is common across each counterfactual, the quantitative magnitude differs dramatically. In the standard bank balance sheet model, there is no margin for adjustment. In this *Neither Margin* counterfactual, the price and quantity response is the largest. because bank balance sheets are the *only* source of financing in the economy, the increases in the effective cost of balance sheet financing are unavoidably passed through to borrowers. In our simulations, interest rates increase by over 100 basis points and quantities decrease by 60%.

The presence of banks offsets some of banks' inability to lend off balance sheet. Bank lending necessarily becomes more expensive, but some of these borrowers are able to substitute towards less expensive shadow banks financed through securitization. Price increases are lower—roughly 35 basis points—and quantity decreases less dramatic—roughly 40%. This counterfactual shows that while shadow banks are a partial but imperfect substitute for bank balance sheet lending. Two forces are responsible for this result. First, on the supply side, securitized financing through shadow banks is more expensive than balance sheet financing from a well-capitalized bank. When lending switches to shadow banks, prices must increase to reflect these higher costs. Second, on the demand side, the products that shadow banks offer are not perfect substitutes for the products that banks offer, and thus, even adjusting for the higher price, some would-be bank borrowers would prefer to exit on the extensive margin rather than switching to shadow bank financing on the intensive margin.

When there is no shadow bank margin (but banks can sell loans), as shown in the wide-dashed green line, the effects of the capital shock on lending are even more muted. Rates increase by only 25 basis points and quantities decrease by only 20%. The relatively more important role for the bank securitization margin as compared to the shadow bank margin is consistent with the findings

in Buchak et al. (2022). It is explained by the fact that securitized *bank* lending is a better substitute for *bank* balance sheet lending than securitized *shadow bank* lending is. While costs for securitized financing through banks is higher than costs for balance sheet lending from a well-capitalized bank (a supply side imperfection), from the perspective of borrowers, both products are still bank loans, and thus are closer substitutes on the demand side. For example, some borrowers may prefer the convenience of accessing all bank services in one place. In other words, from the borrower's perspective, a securitized bank loan is *almost* the same as a bank balance sheet loan, while a securitized bank loan is *not* the same as a securitized shadow bank loan.

Finally, in the baseline scenario where both margins are active, the price and quantity effects are the most muted. Prices increase by roughly 10 basis points and quantities decrease by roughly 10 percentage points. In other words, the two margins of adjustment significantly dampen the effect of capital shocks to the intermediation sector at impact.

### **5.4.3 Shadow Bank and Loan Sale Multipliers**

We next show that the multipliers which we record in the data respond to capital shocks in the intermediary sector. In other words, our model generates large and variable Shadow Bank and Loan Sale Multipliers (Figure 11, Panels (d) and (e)). Following the capital shock, the loan sale lending multiplier increases from below 2 to nearly 4, meaning that the share of *bank* lending that financed through bank balance sheets declines from one half to one quarter. The shadow bank multiplier is smaller, increasing from about 2 to 2.5, implying an increase in shadow bank lending share of about 10pp.

To understand how the two margins of adjustment interact, we study the importance of shadow bank lending when we prevent bank loan sales. The shadow bank multiplier becomes volatile, and following the shock, increases dramatically to well above 5. Because banks can only originate on balance sheet, nearly all lending migrates to shadow banks following the shock. In contrast, when banks are able to securitize loans, banks continue to originate a large fraction of loans in the economy off balance sheet, and thus the response of the shadow bank multiplier is much more muted on average.

### **5.5 The Consequences of Model Misspecification**

Finally, we use our model to illustrate why models based on bank balance sheet lending struggle to quantitatively match data, and therefore have important drawbacks as guides to policy responses. To undertake this exercise, we first re-calibrate our baseline model but restrict the model to bank balance sheet lending. We use aggregate lending data in the calibration to give the bank balance sheet model the best chance to capture the relationship between bank capital and aggregate lending. We then study the same shock as we did above, and point to the basic economic tension

when bank balance sheet models are confronted with data: that aggregate lending is not very responsive to bank capital.

### 5.5.1 A Bank Balance Sheet Model Calibration

To highlight the basic economic tension in the bank balance sheet model, we seek to approximate ways in which a researcher would calibrate a model that links bank balance sheet strength to aggregate lending activity without considering the margins we emphasize. This calibration exercise differs from our prior calibration in two concrete ways. First, on the model side all lending in the economy must be financed on depository institutions' balance sheets. There is no securitization nor are there shadow banks. This is the sense in which the model is mis-specified. Second, on the data side, we overlook the empirical distinction between bank balance sheet lending and total lending, because in the eyes of the model, they are the same. We work from the assumption that aggregate lending is the regulators main concern. Therefore, to give the model the best chance to perform well on the dimension of aggregate lending, we only match moments concerning total lending. Table 4, Panel (a) shows the moments we seek to match in the baseline and bank balance sheet calibrations.

The main economic tension can be seen when the bank balance sheet model targets the correlation between bank capital and aggregate lending, which is very small, at 0.01. On the other hand, the correlation between the change in bank capitalization and *balance sheet* lending, in the data is roughly 0.23. This difference does not arise from the micro-macro data wedge, but is instead driven by the two multipliers we describe in the empirical section. The mean shadow bank multiplier and loan sales multiplier are approximately 2.3 and 2.1, respectively, with non-trivial variation over time, which provide discipline for the two margins we study. Because the balance sheet calibration does not recognize these channels, it has to reconcile the very weak correlation between bank capital and aggregate lending with a model which forces total lending to depend strongly on bank balance sheet health. This means that when bank balance sheets are impaired in the model, total lending *must* decrease by a quantitatively significant amount.

The bank balance sheet model achieves this difficult reconciliation by both by making equity issuance less costly, and by directly shrinking the size of the region in which capital impairment is costly. Table 4, Panel (b) shows the results of the calibration. The key parameter differences between the baseline specification and the mis-specified calibration. Broadly, the key difference for the mis-specified calibration as compared to the baseline specification are lower equity issuance costs (both in terms of the fixed cost of issuing equity and the convex adjustment cost) as well as what appears to be a more permissive regulatory regime.

This exercise thus highlights the basic economic tension in calibrating bank balance sheet models. If they are calibrated to aggregate data, they have to reconcile the low correlation between bank

capital and aggregate lending. If, on the other hand, they are calibrated to bank balance sheet data, then they severely overestimate the effect of bank balance sheet shocks.

### **5.5.2 Policy Consequences of using Bank Balance Sheet Models**

We next show that using bank balance sheet models for policy analysis come with quantitative drawbacks. We compare the lending response to a negative capital shock in our model with the bank balance sheet model. The nature of the shock is exactly the same as that imposed earlier: a large negative shock to bank capital from the pre-shock steady state. Figure 12 shows the impulse response.

Even calibrated to actual data, the bank balance sheet model overstates the increase in interest rates and decrease in lending quantities. In the baseline model with both margins active, the interest rate increase is modest of 10 bp; in the balance sheet model, rates increase by 50 basis points. Similarly, Panel (c) shows that lending quantities decrease substantially in the bank balance sheet model.<sup>14</sup>

The bank balance sheet model makes two mistakes and tries to thread the needle between them. On one hand, this model dramatically *overstates* the impact of bank capital shocks on total lending, both in prices and quantities. This is because the model does not allow for the relevant margins of substitution. On the other hand, the model dramatically *understates* the impact of bank capital shocks on bank balance sheet lending because a realistic calibration seeks to match the relationship between bank capital shocks and total lending needs to match an empirically small correlation. When bank balance sheet lending is the only type of lending in the economy, this necessarily forces the model to generate a counterfactually small correlation between bank balance sheet capital and bank balance sheet lending.

In sum, the bank balance sheet model implies that bank balance sheets are *more* important for total lending than they really are, but are *less* important for bank balance sheet lending than they really are. This is a particularly important distinction, because some sectors of lending have flexible margins of adjustment, e.g., residential mortgage lending, and other sectors lack these margins of adjustment, e.g., business lending.

## **6. Discussion and Implications**

Several key insights emerge from our analysis. First, we show that the evaluation of any policy that targets credit must incorporate the lenders' ability to sell their loans and the equilibrium interaction of banks and shadow banks. A policy analysis that does not recognize these margins of adjustment will misdirect resources based on faulty perceptions of their effect on aggregate lending as well as where the risk resides in the economy. Shocks to bank capital are neither as severe, nor

---

<sup>14</sup> The responses in the calibrated bank balance sheet model are substantially smaller than the declines in the “Neither Margin” counterfactual. This difference arises because the separate calibration of the bank balance sheet model allows it to better match the muted lending responses to bank capital shocks in the data.

as long lasting as suggested by bank balance sheet models. Moreover, the design of policies need to be mindful that ignoring these margins could have distributional consequences, because the margins of adjustments differ across intermediaries which serve households across the income distribution.

We also show that bank balance sheet models face a basic economic tension when used in a quantitative setting: either they can reconcile aggregate lending data or bank balance sheet lending data, but not both. Macro financial models which hope to match data in a quantitative sense therefore must account for the industrial organization of the modern financial sector. More broadly, our paper underscores the limitations of bank balance sheets as a source of data on lending (i.e., focusing solely on call report data by traditional banks) and highlights the critical importance of collecting and making available data on overall lending that includes lending done by shadow banks and loans that were not retained on the balance sheets of financial institutions.

### References

- Adrian, T., and A. B. Ashcraft, 2016, Shadow Banking: A Review of the Literature. In *Banking Crises*, pp. 282-315. *Palgrave Macmillan UK*.
- Agarwal, S., Amromin, G., Ben-David, I., Chomsisengphet, S., Piskorski, T., Seru, A., 2017. Policy Intervention in Debt Renegotiation: Evidence from Home Affordable Modification Program, *Journal of Political Economy* 125, 654-712.
- Begenau, J, and T. Landgvoit, 2022, Financial Regulation in a Quantitative Model of the Modern Banking System, *Review of Economic Studies* 89, 1748–1784
- Benetton, Matteo, 2021, Leverage Regulation and Market Structure: An Empirical Model of the UK Mortgage Market, *Journal of Finance*, 2997-3053.
- Benmelech, E., R. R. Meisenzahl, R. Ramcharan, 2017, The Real Effects of Liquidity During the Financial Crisis: Evidence from Automobiles, *Quarterly Journal of Economics* 132, 317-365.
- Beraja, Martin, Andreas Fuster, Erik Hurst, and Joseph Vavra, 2019, Regional Heterogeneity and the Refinancing Channel of Monetary Policy, *Quarterly Journal of Economics* 134, 109–83.
- Bianchi, Javier and Saki Bigio, 2022, Banks, Liquidity Management and Monetary Policy, *Econometrica* 90, 391-454.
- Buchak, G., G. Matvos. T. Piskorski, and A. Seru, 2018, Fintech, Regulatory Arbitrage and the Rise of Shadow Banks, *Journal of Financial Economics* 130, 453-483.
- Buchak, G., G. Matvos. T. Piskorski, and A. Seru, 2022, Beyond the Balance Sheet Model of Banking: Implications for Bank Regulation and Monetary Policy, forthcoming in the *Journal of Political Economy*.
- Brunnermeier, and Y. Sannikov, 2014, A Macroeconomic Model with a Financial Sector, *American Economic Review* 104, 379-421.

- Chen, B.S., S.G. Hanson, and J.C. Stein, 2017, The Decline of Big-Bank Lending to Small Business: Dynamic Impacts on Local Credit and Labor Markets, NBER Working Paper 23843.
- Chernenko, Sergey, I. Erel and R. Prilmeier, 2022, Nonbank Lending, *Review of Financial Studies* 35, 4902–4947.
- Cherry, S., E. Jiang, G. Matvos, T. Piskorski, and A. Seru, 2021, Government and Household Debt Relief during COVID-19, *Brookings Papers on Economic Activity* 52, 1–86
- Cherry, S., E. Jiang, G. Matvos, T. Piskorski, and A. Seru, 2022, Shadow Bank Distress and Household Debt Relief: Evidence from the CARES Act, *AEA Papers and Proceedings* 112: 509–515.
- Corbae D., and P. D’Erasmus, 2021, Capital Buffers in a Quantitative Model of Banking Industry Dynamics, *Econometrica* 89, 2975–3023.
- De Roure, C., L. Pelizzon and A.V. Thakor, 2022, P2P Lenders versus Banks: Cream Skimming or Bottom Fishing?, *Review of Corporate Finance Studies* 11, 213-262.
- Di Maggio, Marco, Amir Kermani, Benjamin J. Keys, Tomasz Piskorski, Rodney Ramcharan, Amit Seru, and Vincent Yao, 2017, Interest Pass-Through: Mortgage Rates, Household Consumption and Voluntary Deleveraging, *American Economic Review* 107, 3550–3588.
- Drechsler, I., A. Savov, and P. Schnabl, 2017, The Deposits Channel of Monetary Policy, *Quarterly Journal of Economics* 132, 1819-1876.
- Drechsler, I., A. Savov, and P. Schnabl, 2022, How Monetary Policy Shaped the Housing Boom, *Journal of Financial Economics* 144, 992-1021.
- Eichenbaum, Martin, Sergio Rebelo, and Arlene Wong, 2022, State Dependent Effects of Monetary Policy: The Refinancing Channel, *American Economic Review* 112, 721-761.
- Elenev, V., T. Landvoigt, and S. Van Nieuwerburgh, 2021, A Macroeconomic Model with Financially Constrained Producers and Intermediaries. *Econometrica* 89, 1361-1418.
- Gennaioli, N., A. Shleifer, and R. W. Vishny, 2013, A Model of Shadow Banking, *Journal of Finance* 68, 1331–1363.
- Gopal, M., and P Schnabl, 2022, The Rise of Finance Companies and FinTech Lenders in Small Business Lending, *Review of Financial Studies* 35, 4859-4901.
- Greenwood, R., J. C. Stein, S. G. Hanson, and A. Sunderam, 2017, Strengthening and Streamlining Bank Capital Regulation, *Brookings Papers on Economic Activity*, 479–565.
- Hanson S.G, A. Kashyap, and Jeremy C. Stein, 2011, A Macroprudential Approach to Financial Regulation, *Journal of Economic Perspectives* 25, 3-28.
- He, Zhiguo, and Arvind Krishnamurthy, 2013, Intermediary Asset Pricing, *American Economic Review*, 103, 732-70.
- Hachem, K., and Z. Song, 2021, Liquidity Rules and Credit Booms, *Journal of Political Economy* 129, 2721–2765.

- Huang, J., 2018, Banking and Shadow Banking, *Journal of Economic Theory* 178, 124-152.
- Irani, R.M., R. Iyer, R.R. Meisenzahl and J.L. Peydro, 2021, The Rise of Shadow Banking: Evidence from Capital Regulation, *Review of Financial Studies* 34.5, 2181-2235.
- Jiang, E., 2020, Financing Competitors: Shadow Banks' Funding and Mortgage Market Competition, working paper.
- Jiang, E., G. Matvos, T. Piskorski, and A. Seru, Banking without Deposits: Evidence from Shadow Bank Call Reports, working paper.
- Kashyap, A. K., and J. C. Stein, 2000, What Do a Million Observations on Banks Say about the Transmission of Monetary Policy?, *American Economic Review*, 90, 407-428.
- Kashyap, A., and J. Stein, 2004, Cyclical Implications of Basel II Capital Standards, *Federal Reserve Bank of Chicago, Economic Perspectives*, 1st Quarter, 18-31.
- Moriera, Alan and Savov, Alexi, 2017, The Macroeconomics of Shadow Banking, *The Journal of Finance* 72, 2381-2432.
- Ordonez, G., 2018, Sustainable Shadow Banking, *American Economic Journal: Macroeconomics* 10, 1-25.
- Plantin, Guillaume, 2015, Shadow Banking and Bank Capital Regulation, *Review of Financial Studies* 28, 146–75.
- Seru, Amit, 2019, “Regulating banks in the era of fintech shadow banks”, Andrew Crockett Memorial Lecture, Bank for International Settlements.
- Tang, Huan, 2019, Peer-to-Peer Lenders versus Banks: Substitutes or Complements?, *Review of Financial Studies*, 32, 1900-1938.
- Wang Yifei, Toni Whited, Yufeng Wu and Kairong Xiao, 2022. "Bank Market Power and Monetary Policy Transmission: Evidence from a Structural Estimation," *The Journal of Finance*, vol 77(4), pages 2093-2141.
- Xiao, Kairong, 2020. Monetary Transmission through Shadow Banks, *The Review of Financial Studies* 33, 2379-2420.

**Table 1: Variation in the Total Bank Lending Volume Accounted by the Bank Balance Sheet Lending Volume and the Variation in the Total Lending Volume Accounted by the Total Bank Lending Volume**

This table presents the OLS estimates from the regression of *Total Bank Lending* on *Bank Balance Sheet Lending* (Column 1), *Total Bank Lending* on *Bank Balance Sheet Lending* and *Bank Balance Sheet Sold* (Column 2), *Total Lending* on *Total Bank Lending* (Column 3), and *Total Lending* on *Total Bank Lending* and *Shadow Bank Lending* (Column 4) on the national data. *Total Bank Lending* is defined as annual aggregate residential mortgage origination volume originated by banks. *Bank Balance Sheet Lending* is defined as the annual aggregate residential mortgage origination volume originated by banks that the banks retain on their balance sheet in the year of its origination. *Bank Sold Lending* is the annual aggregate residential mortgage origination volume originated by banks that the banks sell in the year of its origination. *Total Bank Lending* is defined as the annual aggregate residential mortgage origination volume originated by banks and shadow banks. Finally, *Shadow Bank Lending* is defined as the annual aggregate residential mortgage origination volume originated by shadow banks. Columns (5)-(10) present the corresponding results for the county-level data where we scale the variables by the county-level mean of the dependent variable. Columns (6)-(7) and (9)-(10) also include the county and year fixed effects. The estimation sample is 2005 to 2021. Standard errors are reported in parentheses. Data Source: HMDA Data.

	National Data				County Data					
	Total Bank Lending		Total Lending		Total Bank Lending			Total Lending		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Bank Balance Sheet Lending	1.05 (0.31)	1.00 (0.00)			1.00 (0.00)	1.11 (0.00)	1.00 (0.00)			
Bank Sold Lending		1.00 (0.00)					1.00 (0.00)			
Total Bank Lending			2.28 (0.40)	1.00 (0.00)				1.08 (0.00)	1.05 (0.00)	1.00 (0.00)
Shadow Bank Lending				1.00 (0.00)						1.00 (0.00)
County FEs	No	No	No	No	No	Yes		No	Yes	Yes
Year FEs	No	No	No	No	No	Yes	1.00	No	Yes	Yes
Observations	17	17	17	17	54,829	54,829	54,829	54,840	54,840	54,840
R-squared	0.44	1.00	0.68	1.00	0.53	0.74	1.00	0.54	0.86	1.00

**Table 2: Errors in Inference on Total Bank and Overall Lending Volume**

Panel (a) of this table presents the OLS estimates from the regression of *Total Bank Lending* volume on the *Inferred Total Bank Lending* volume from *Total Bank Balance Sheet Lending* volume (in \$ billions) assuming a constant loan sale multiplier equal to the sample mean (Column 1). Panel (b) shows the OLS estimates from the regression of *Total Lending* volume on the *Inferred Total Lending* volume from *Total Bank Lending* volume (in \$ billions) assuming a constant shadow bank multiplier equal to the sample mean (Column 2). Panel (b) shows the corresponding results for the county-level data where county-level inference uses the corresponding multipliers at the national level. Columns (1)-(4) of panel (b) show the OLS estimates of *Total Bank Lending* volume in a county on the *Inferred Total Bank Lending* volume in a county from *Total Bank Balance Sheet Lending* volume in a county. Columns (5)-(8) of panel (b) show the OLS estimates from the regression of *Total Lending* volume in a county on the *Inferred Total Lending* volume in a county from *Total Bank Lending* volume in a county. In panel (b) in Columns (1)-(2) and (5)-(6) we infer the county-level lending levels using aggregate loan sale and shadow bank multipliers, respectively, that are based on national data and constant over-time and equal to their sample mean. In panel (b) in Columns (3)-(4) and (7)-(8) we infer county level-lending lending patterns using aggregate time-varying loan sale and shadow bank multipliers, respectively, that are based on national and are equal to their sample mean each year. In panel (b), Columns (2), (4), (6), (8) we also add county state fixed effects. The variables in the county-level regressions are scaled by the sample mean of the dependent variable. Data Source: HMDA Data.

**Panel A: National Lending**

	Total Bank Lending	Total Lending
	(1)	(2)
Inferred Total Bank Lending (constant loan sale multiplier)	0.43 (0.12)	
Inferred Total Lending (constant shadow bank multiplier)		1.36 (0.24)
Observations	17	17
R-squared	0.44	0.67

**Table 2: Errors in Inference on Total Bank and Total Lending Volume (continued)**

**Panel B: County-Level Lending**

	Total Bank Lending				Total Lending			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Inferred Total Bank Lending (constant loan sale multiplier)	0.41 (0.002)	.48 (0.002)						
Inferred Total Bank Lending (time-varying loan sale multiplier)			0.39 (0.002)	0.46 (0.002)				
Inferred Total Lending (constant shadow bank multiplier)					0.62 (0.003)	0.75 (0.003)		
Inferred Total Lending (time-varying shadow bank multiplier)							0.58 (0.002)	.6594 (0.002)
County FEs	No	Yes	No	Yes	No	Yes	No	Yes
Observations	55,515	55,515	55,515	55,515	55,527	55,527	55,527	55,527
R-squared	0.53	0.63	0.51	0.60	0.49	0.59	0.66	0.74

**Table 3: Variation in the Total Bank Lending Growth Rate Accounted by the Bank Balance Sheet Lending Growth Rate and the Variation in the Total Lending Growth Rate Accounted by the Total Bank Lending Growth Rate**

This table presents the OLS estimates from the same specification as Table 1 but estimated on the annual growth rates of these variables. In Column (5)-(8) the regressions are volume weighted, where the volume is the county-level bank lending volume in dollars in Column (5)-(7) and the county-level total lending volume in dollars Column (8)-(10). The estimation sample is 2005 to 2021. Standard errors are reported in parentheses. Data Source: HMDA Data.

	National Data				County Data					
	Total Bank Lending		Total Lending		Total Bank Lending			Total Lending		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Bank Balance Sheet Lending	0.82 (0.22)	0.41 (0.06)			0.34 (0.05)	0.36 (0.03)	0.34 (0.03)			
Bank Sold Lending		0.55 (0.03)					0.13 (0.02)			
Total Bank Lending			1.20 (0.09)	0.64 (0.06)				0.99 (0.10)	0.78 (0.04)	0.64 (0.02)
Shadow Bank Lending				0.37 (0.04)						0.24 (0.02)
County FEs	No	No	No	No	No	Yes	Yes	No	Yes	Yes
Year FEs	No	No	No	No	No	Yes	Yes	No	Yes	Yes
Observations	17	17	17	17	54,446	54,446	54,294	54,596	54,596	54,057
R-squared	0.49	0.97	0.93	0.99	0.25	0.76	0.81	0.79	0.93	0.97

**Table 4: Model Calibration**

This table shows key inputs and results from the model calibration. Panel A shows the targeted moments and model simulated moments for the “baseline” model, which has the shadow bank and loan sale multipliers active and aims to target aggregate moments that reflect these, and the “mis-specified” model, which runs the calibration with these multipliers shut down and aims to target aggregate moments that do not distinguish bank balance sheet lending from total lending. Panel B shows the calibrated parameters from each calibration and the key set of identifying moments. Panel C shows the externally calibrated parameters: the parameter from the model, its description, the calibrated value, and the source. These parameters are constant across the calibrations.

**Panel A: Targeted Moments**

	Target		Model	
	<i>Baseline</i>	<i>Mis-specified</i>	<i>Baseline</i>	<i>Mis-specified</i>
E[Shadow Bank Multiplier]	2.294	1.000	2.005	1.000
SD[Shadow Bank Multiplier]	0.726	0.000	0.590	0.000
E[Loan Sale Multiplier]	2.145	1.000	2.112	1.000
SD[Loan Sale Multiplier]	0.459	0.000	0.101	0.000
SD[d Total Lending]	0.348	0.348	0.362	0.367
Corr[d Total Lending, d CR]	-	0.014	0.002	0.001
Corr[d Balance Sheet Lending, d CR]	0.236	-	0.226	0.046
Lending Quantity / Assets	0.250	0.250	0.253	0.193
E[Capital Ratio]	0.109	0.109	0.162	0.099
SD[Capital Ratio]	0.013	0.013	0.098	0.037
AR[Capital Ratio]	0.985	0.985	0.939	0.895
E[d Capital Ratio]	0.000	0.000	0.000	0.000
SD[d Capital Ratio]	0.002	0.002	0.034	0.017
AR[d Capital Ratio]	0.173	0.173	-0.045	-0.013

**Panel B: Calibrated Parameters**

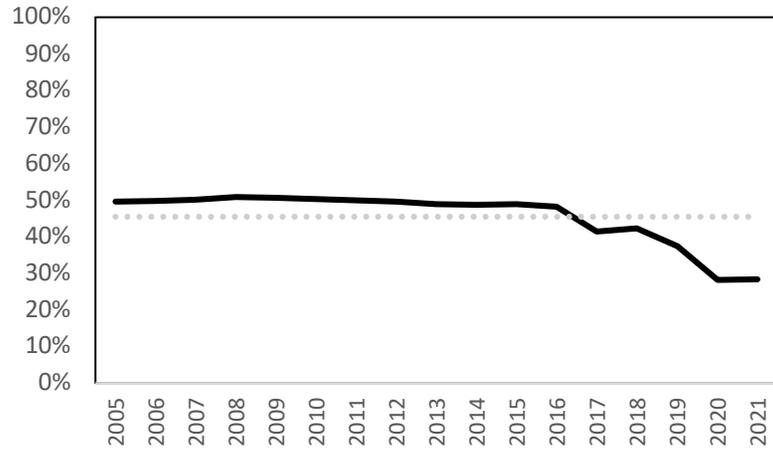
Parameter	Description	Baseline	Mis-specified	Key Moments
$\gamma$	Investment cost convexity	43.10	32.46	Capitalization time-series
$C$	Equity issuance cost	2.08	0.59	Capitalization time-series
$\bar{c}$	Regulatory capital requirement	0.07	0.04	Capitalization time-series
$\lambda$	Regulatory intensity	18.39	19.14	Capitalization time-series
$\theta$	Shock AR coefficient	0.17	0.23	Capitalization time-series
$\sigma_z^2$	Shock variance	0.31	0.36	Capitalization time-series
$m$	Market size	1.00	1.00	Lending volumes and shares
$\delta_b$	Bank non-price demand	1.33	1.32	Lending volumes and shares
$\delta_{sb}$	Non-bank NP demand	1.42	N/A	Lending volumes and shares
$\sigma_{fin}$	Financing shock scale	0.01	0.04	Balance sheet shares
$\sigma_d^2$	Loan demand variance	0.13	0.13	Lending volumes and shares

**Panel C: Externally Calibrated Parameters**

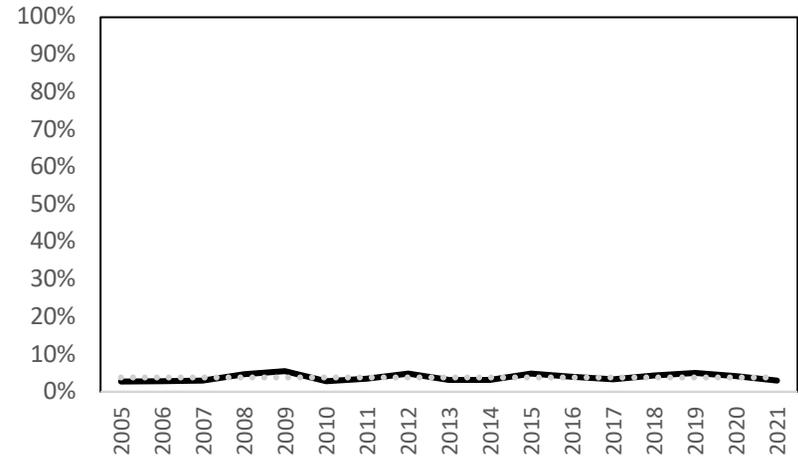
Parameter	Description	Value	Source
$\beta$	Discount factor	0.95	Standard
$\xi$	Asset risk weight	0.25	Buchak et al. (2022)
$mc_b^l$	Bank origination MC	3.00%	Id.
$mc_s^l$	Non-bank origination MC	3.00%	Id.
$mc^s$	Securitization cost	0.67%	Id.
$\alpha$	Price sensitivity	1.65	Id.

**Figure 1: Traditional Balance Sheet Lending Banks in the Residential Mortgage Market**

Panel (a) of this figure shows the percentage of banks among the US mortgage loan originators that do not sell any residential loans in each year, and hence follow the traditional bank balance sheet lending model. Panel (b) shows the percentage of annual bank loan origination volume these traditional balance-sheet only lending banks account for in each year. While up to half of all banks do not sell any loans during our sample period depending on the year, they make up on average only around 7% of the loan origination volume. A loan is retained if it is still on the originating institution's balance sheet at the end of the year. Data Source: HMDA Data.



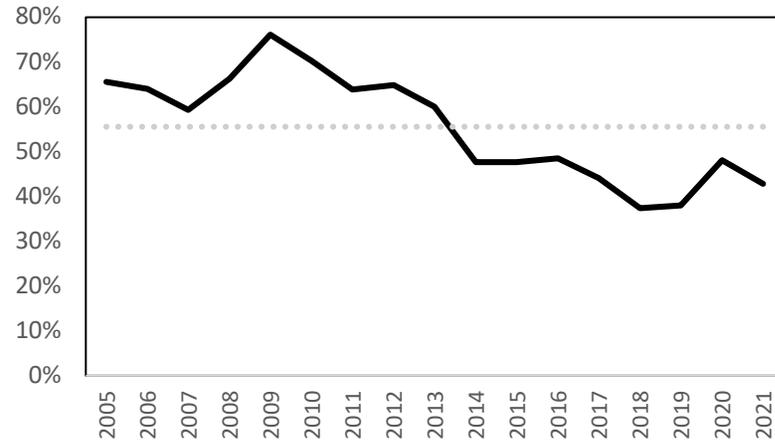
(a) Share of banks following the balance sheet lending business model



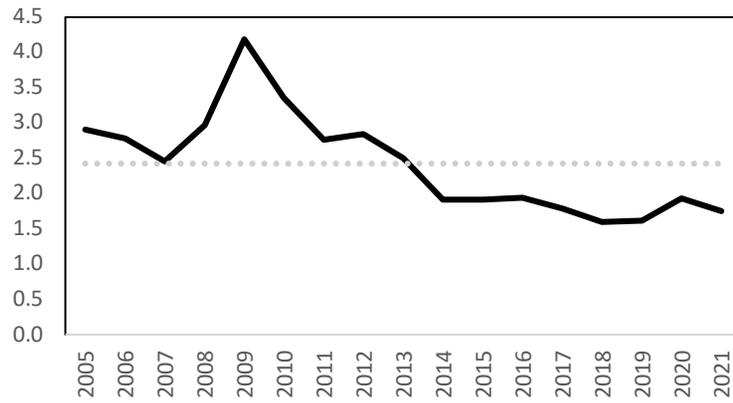
(b) Share of annual lending volume due to banks following the balance sheet lending business model

**Figure 2: Bank Loan Sales Propensity and the Loan Sale Lending Multiplier**

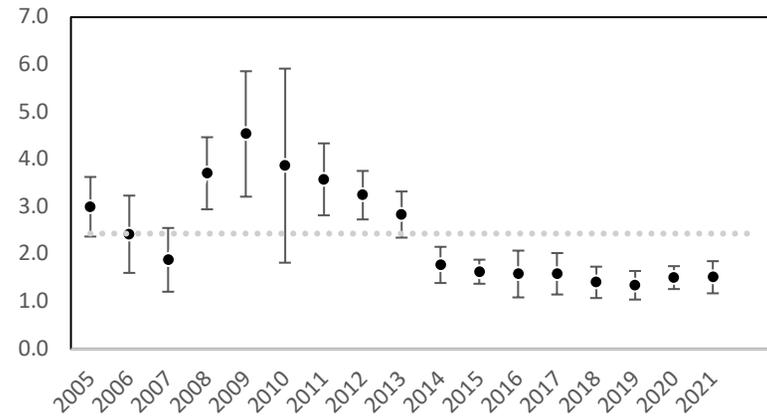
Panel (a) of this figure shows the percentage of residential loan sold among the loans originated by banks per year. The dashed line shows the sample mean. Panel (b) shows the implied loan sale multiplier that indicates the ratio of aggregate bank lending to retained on balance sheet bank lending. Panel (c) shows the estimated loan sale multiplier in each year based on the estimation of equation (1) with individual bank data with 95% confidence intervals. The coefficient of interest indicates the estimated dollar volume of loans originated by a bank in that year for each dollar of lending retained on the bank balance sheet. Data Source: HMDA Data.



(a) Percentage of loans sold by banks



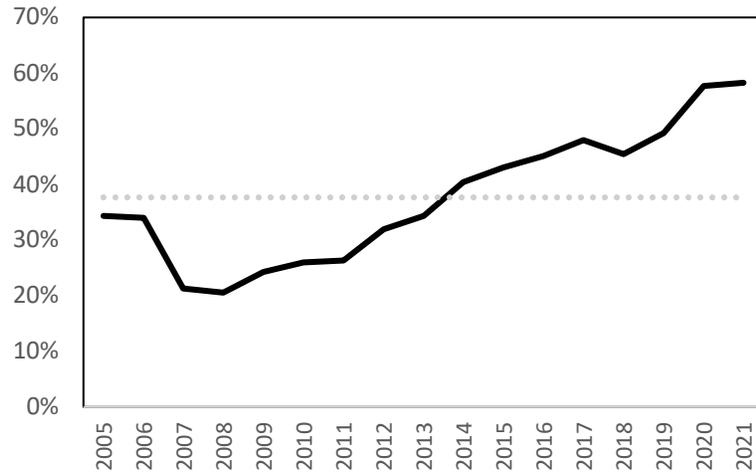
(b) Loan sale multiplier ( $m_t^{Loan Sale}$ )



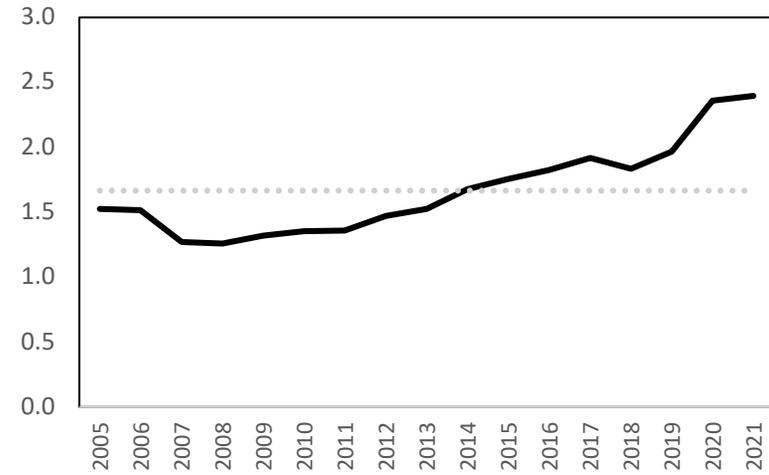
(c) Estimated loan sale multiplier across banks

**Figure 3: Shadow Bank Market Share and the Shadow Bank Loan Multiplier**

Panel (a) of this figure shows the percentage of residential mortgage loans originated by the shadow banks in the US. The dashed line shows the sample mean. Panel (b) shows the implied shadow bank multiplier, with the dashed line being the sample mean again. Data Source: HMDA Data



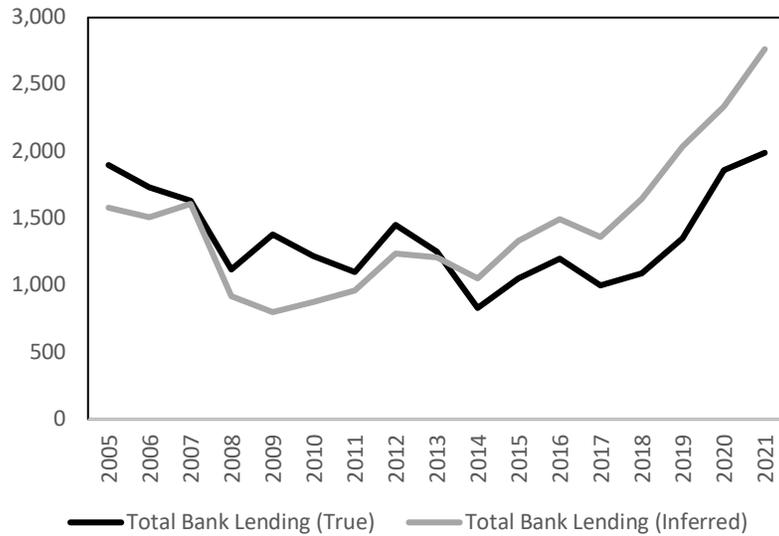
(a) Shadow bank market share



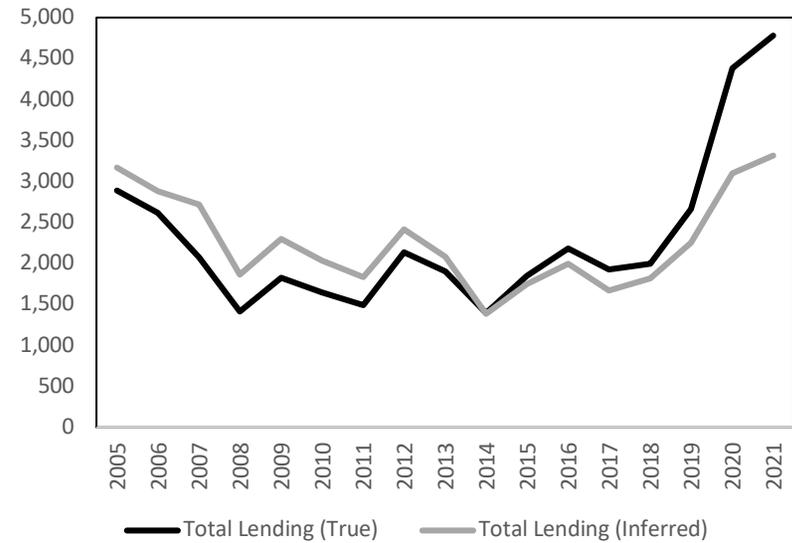
(b) Shadow bank lending multiplier  $m_t^{Shadow Bank}$

**Figure 4: Errors in Inference of Total Bank Lending and Total Lending Volume due to Time-Varying Loan Sale and Shadow Bank Multipliers**

Panel (a) of this figure shows the total annual bank volume (black) and the inferred total bank volume (grey) from the aggregate bank balance sheet lending volume (in \$ billions) assuming a constant loan sale propensity, resulting in the constant loan sale multiplier equal to 2.44 (the sample mean). Panel (b) shows the total annual lending volume and the inferred total lending volume from total bank volume assuming a constant shadow bank market share equal, resulting in the constant shadow bank lending multiplier equal to 1.67 (the sample mean). Source: HMDA Data.



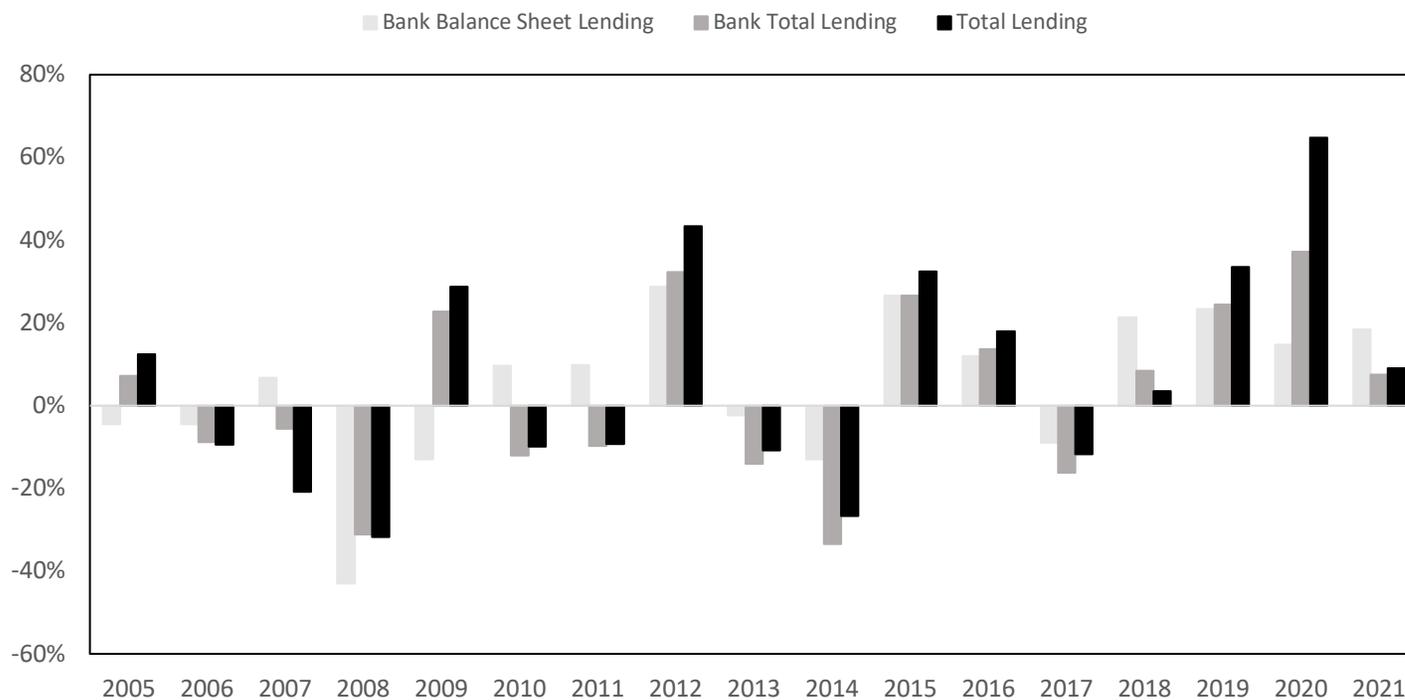
(a) Total bank lending volume (true vs inferred)



(b) Total lending volume (true vs inferred)

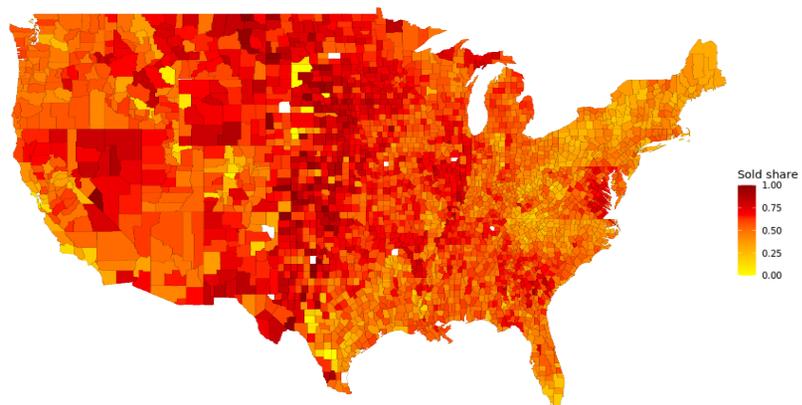
**Figure 5: Growth Rates of Aggerate Bank Balance Sheet Lending, Bank Lending, and Overall Lending**

This figure shows the annual growth rates of aggerate bank balance sheet lending (light grey), aggregate bank lending (in dark grey), and aggregate total lending (black). The aggregate total lending includes lending by both banks and shadow banks. As we observe, not only these growth rates are different in magnitudes but in several years have opposing signs. Source: HMDA Data.

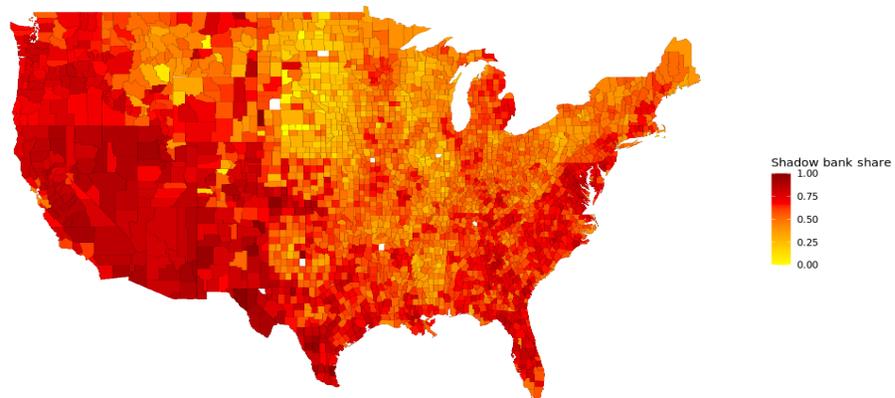


### Figure 6: Regional Heterogeneity --- Fractions of Loans Sold by Banks and Shadow Bank Share

Panel (a) of this figure shows the fraction of loans in a county that banks originate and sell in 2021, the latest year in our data. Panel (b) shows the fraction of loans in a county originated by shadow banks in 2021. Source: HMDA Data.



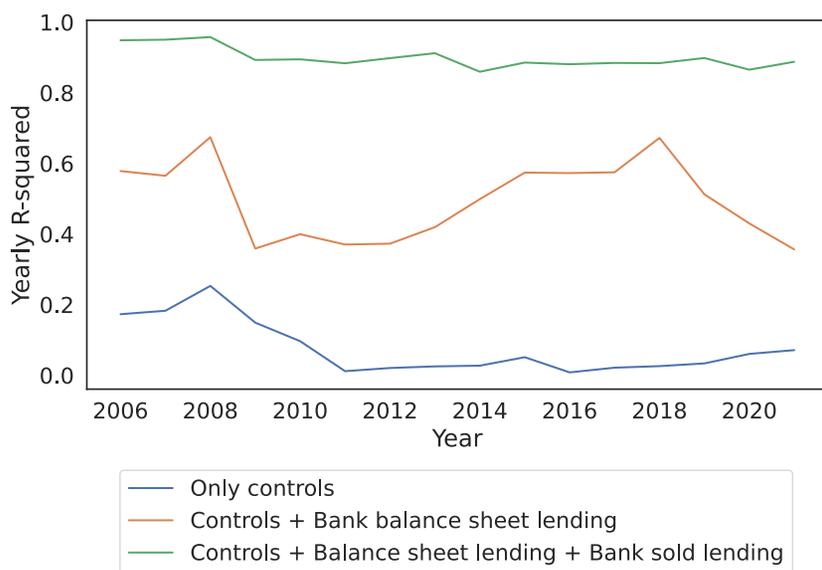
(a) Fraction of loans sold by banks



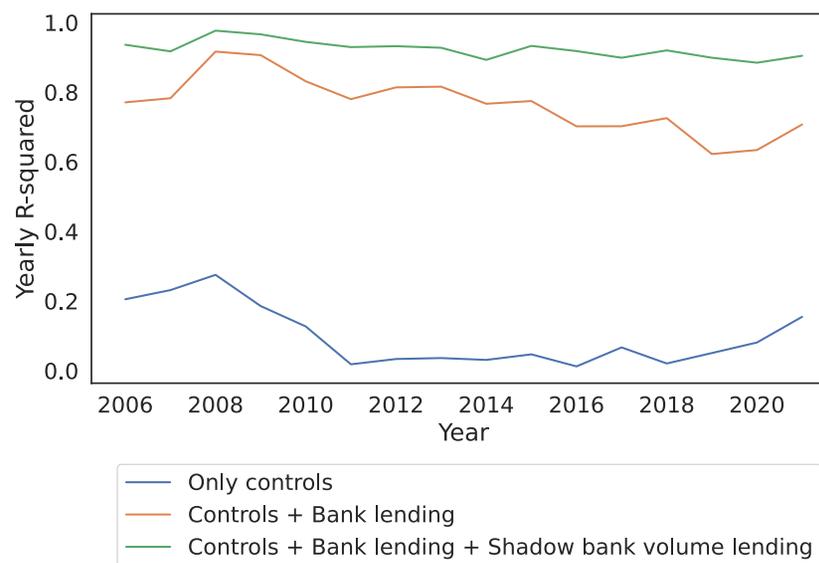
(b) Shadow bank market share

**Figure 7: Accounting for the Regional Variation in the Bank and Total Lending Growth Rates over Time**

This figure shows the R-squared from the county-level regression estimated in each of the plotted years in the cross-section of counties. Panel (a) shows the results for the county-level growth in the bank lending as the dependent variable while panel (b) shows the corresponding results for the county-level growth in total lending. In panel (a) we show these results from three specifications: (i) one with a set of regional controls including the county-level growth in house prices and unemployment, (ii) one that in addition to controls in the specification (i) adds the growth in the county-level bank balance sheet lending, and (iii) the one that in addition adds the county-level growth in bank sold lending as a control variable. In panel (b) we show corresponding results for three specifications: one with a set of regional controls, the second one that in addition adds the growth in the county-level total bank lending, and the one that in addition adds the county-level growth in shadow bank lending as a control variable. Source: HMDA Data.



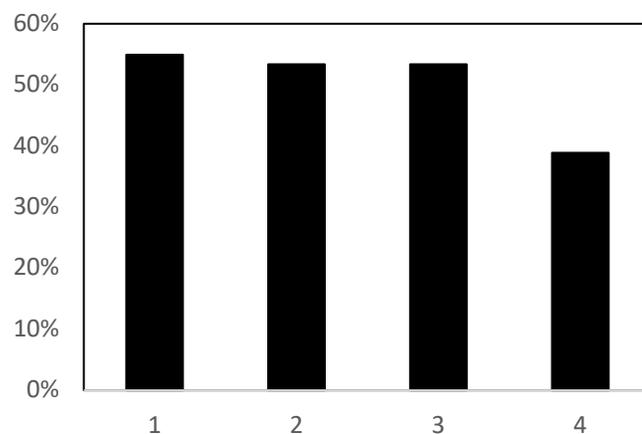
(a) Variation in the bank lending growth accounted by the regional controls, and bank balance sheet and bank sold lending growth



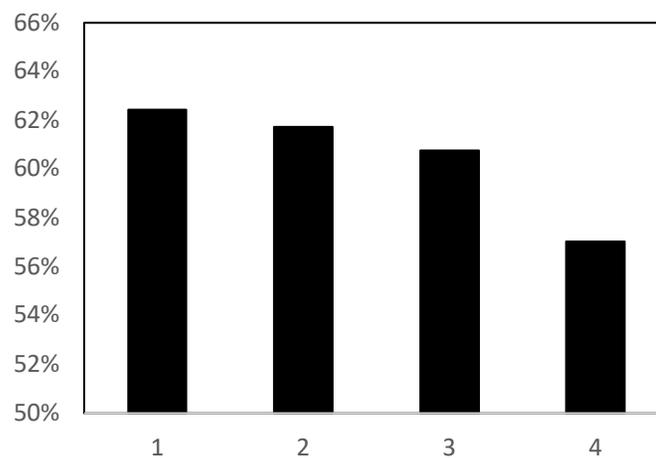
(b) Variation in the total lending growth accounted by the regional controls, and bank lending and shadow bank lending growth

### Figure 8: Regional Heterogeneity --- Fractions of “Loans Sold by Banks” and “Shadow Bank Share” by County Income

Panel (a) of this figure shows the fraction of loans in a county that banks originate and sell in 2021 sorted by the county income quartile with (one being the lowest and fourth being the highest income). Panel (b) shows the fraction of loans in a county originated by shadow banks in 2021 sorted by the county income decile. The means are weighted by the county-level lending volume. Source: HMDA Data.



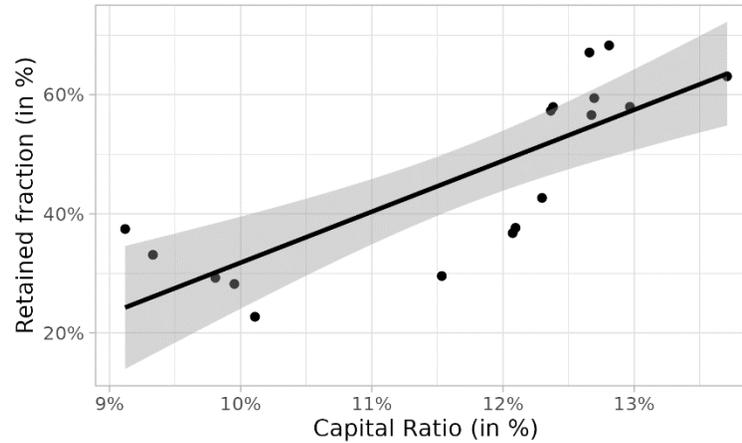
(a) Bank sold share by county income



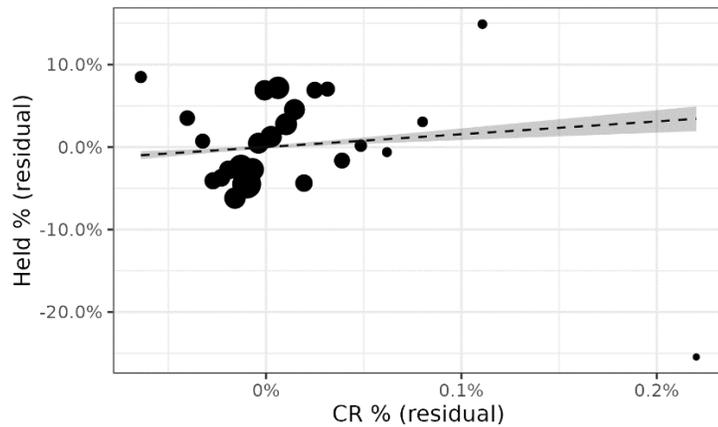
(b) Shadow bank share by county income

### Figure 9: Bank Capitalization and Loan Retention

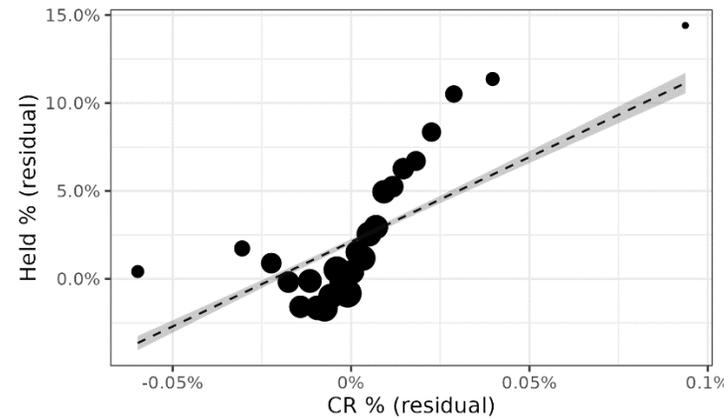
Panel (a) of this figure plots the relation between average aggregate bank capitalization in a year and the aggregate fraction of loans that banks retain on their balance sheet as a share of all loans they originate in each year during 2005-2021 period. The average bank capitalization is weighted by the total volume of the loans they originate. Panel (b) and (c) use bank-level data and show binned scatterplots (25 equal-sized bins) of a bank percent of loans retained on balance sheet in a given year versus a bank capital ratio in a given year. Both loan retention and capital ratios are residualized using a set of bank-level controls and year time dummies. “Within” analysis in panel (c) also removes the bank fixed effects. Estimation sample is 2005-2021. Source: HMDA Data and FFIEC Bank Call Reports.



(a) Aggregate bank capitalization and banks' loan retention



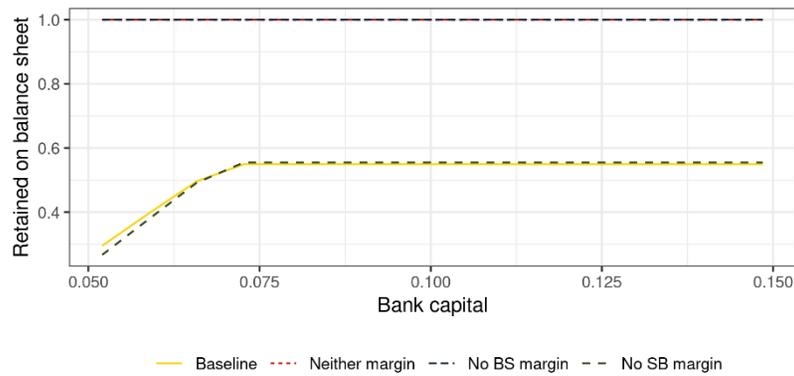
(b) Loan retention and bank capitalization, across banks



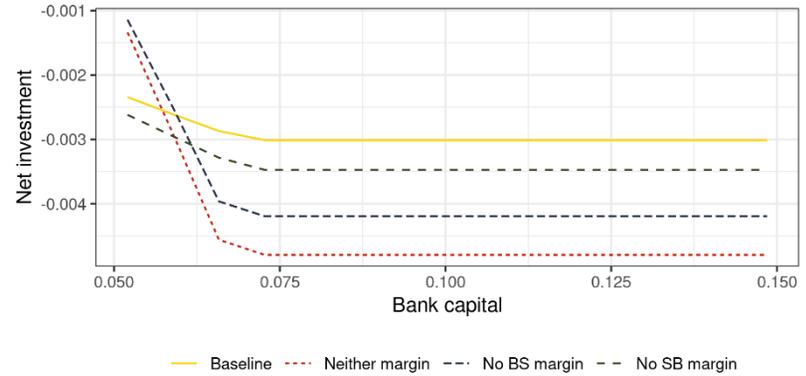
(c) Loan retention and bank capitalization, within banks

### Figure 10: Policy Functions

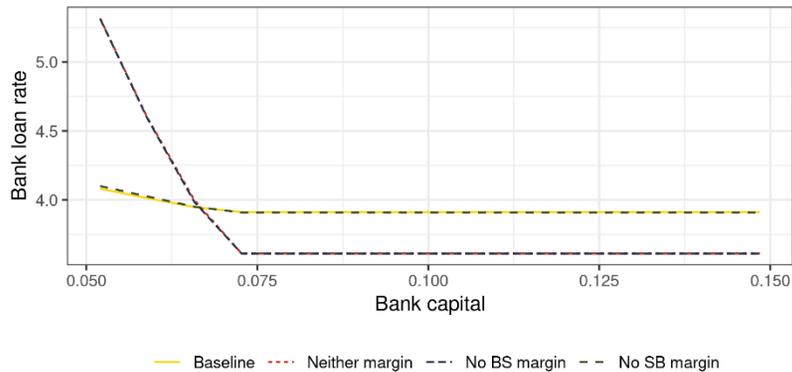
This figure shows the optimal bank policy functions from the calibrated model for the baseline specification (solid yellow line), the model with no loan sales margin (narrow-dashed blue line), the model with no shadow bank margin (wide-dashed green line), and the model with neither margin (dotted red line). In each figure, the x-axis is the key state variable, the bank capital ratio. The y-axis shows the bank's optimal policy under each scenario for each value of the capital ratio. Panel (a) shows the fraction of loans retained on balance sheet. Panel (b) shows net investment, with a negative number denoting a dividend. Panel (c) shows the bank's optimal loan rate, and panel (d) shows the bank's lending quantities.



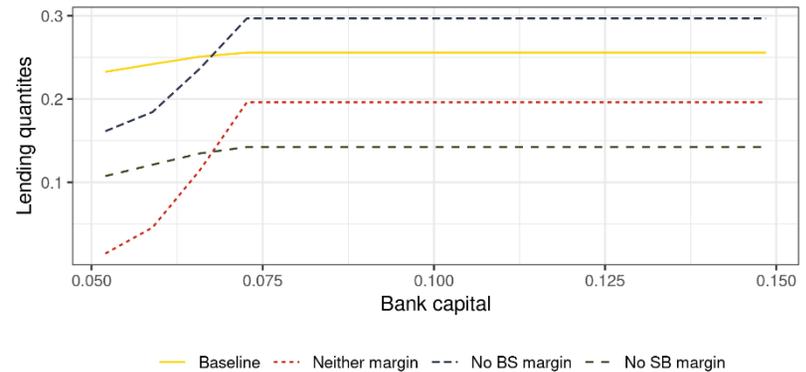
(a) Loans retained on balance sheet



(b) Net investment



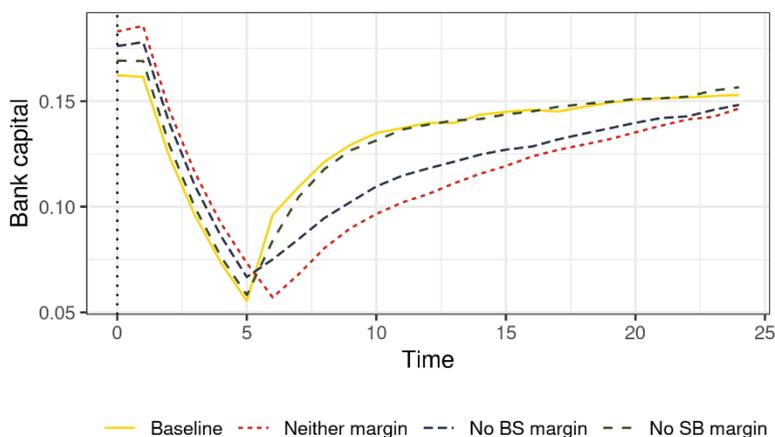
(c) Lending rate



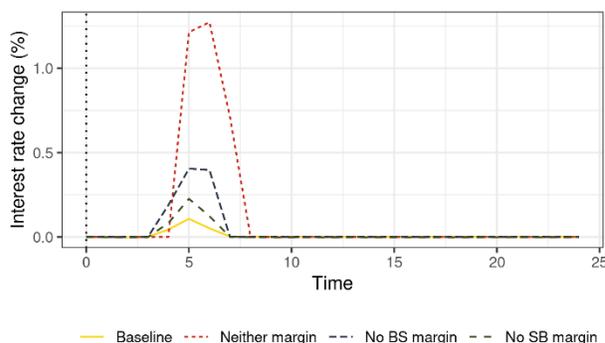
(d) Lending quantity

### Figure 11: Counterfactual Impulse Response

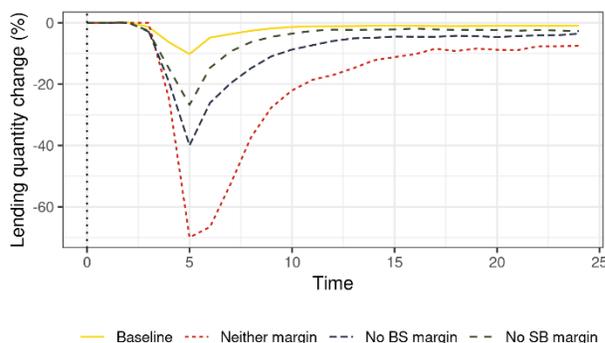
This figure shows the simulated impulse response of a large negative shock to bank capital at time  $t = 0$  across the baseline (yellow) and counterfactual scenarios (no loan sales, narrow-blue dash; no shadow banks, wide green dash; neither margin, red dots). The x-axis is the time relative to the shock in quarters. Panel (a) shows bank capital; Panel (b) shows the average lending rate in pp deviations from the steady-state value; Panel (c) shows the percentage change in lending from the steady-state value. Panel (d) shows the shadow bank multiplier; Panel (e) shows the loan sale multiplier.



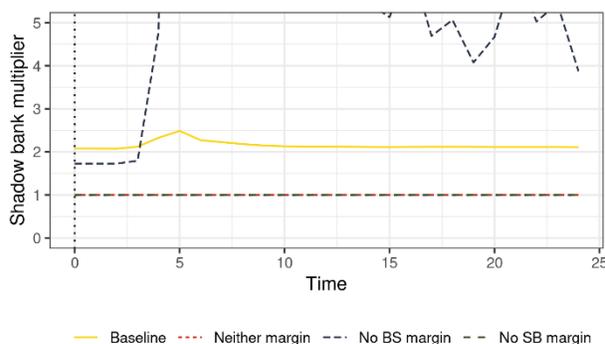
(a) Bank capital



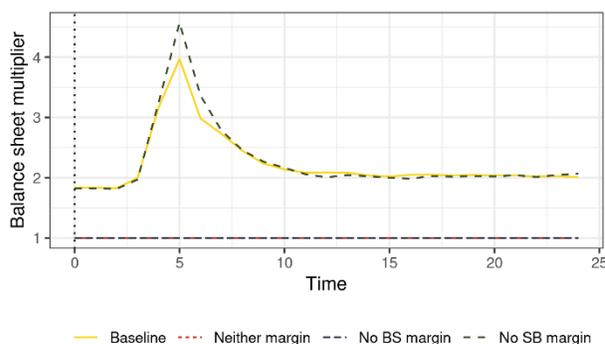
(b) Interest rate deviation from SS



(c) Lending quantity deviation from SS



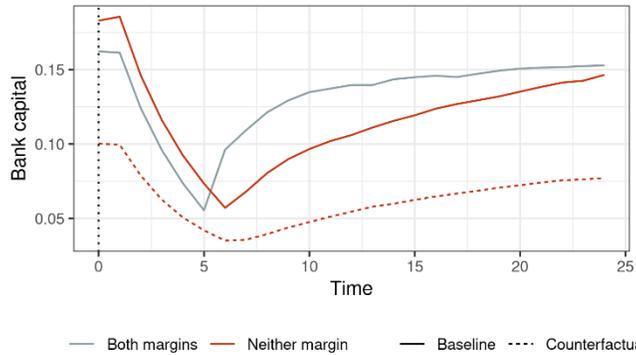
(d) Shadow bank multiplier



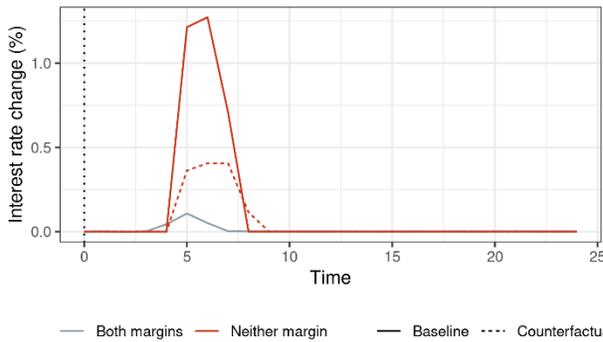
(e) Loan sale multiplier

## Figure 12: Mis-specified Calibrations

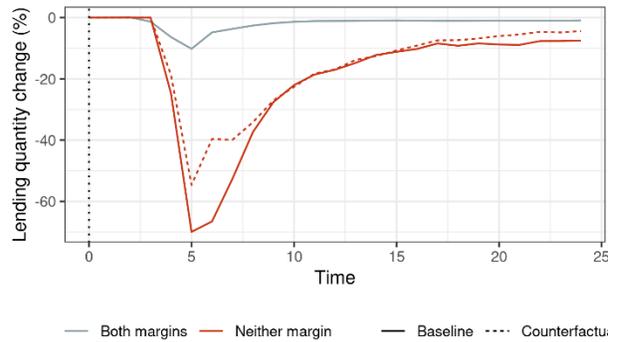
This figure shows the simulated impulse response of a large negative shock to bank capital at time  $t = 0$  across three scenarios: The solid lines show counterfactuals where both margins are active (gray) and where neither margin is active (red), under the baseline calibration, which matches overall lending, balance sheet lending, and their correlations to bank balance sheet capital under a model where both margins are active. The dotted red line shows counterfactuals where neither margin is active under the “mis-specified” calibration, which matches overall lending and its correlation to bank balance sheet capital under a model where neither margin is active. The x-axis is the time relative to the shock in quarters. Panel (a) shows bank capital; Panel (b) shows the average lending rate in pp deviations from the steady-state value; Panel (c) shows the percentage change in lending from the steady-state value.



(a) Bank capital



(b) Interest rate deviation from SS

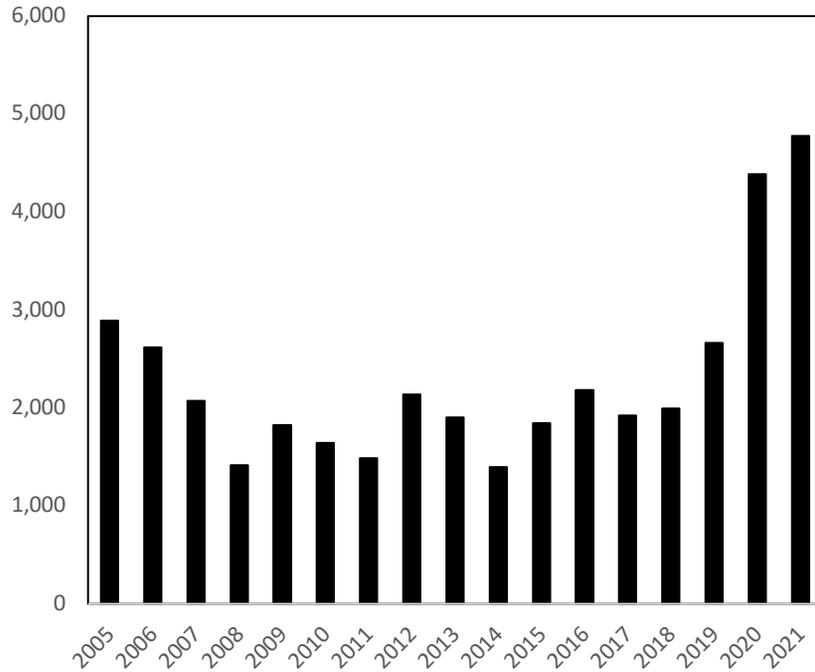


(c) Lending quantity deviation from SS

## **Appendix**

**Figure A1: Aggregate Residential Mortgage Origination Volume and the Refinancing Share**

Panel (a) shows total annual mortgage origination volume in billions of US dollars in 2005-2021 period. Panel (b) shows the percent of mortgage originations that were refinances in each year. *Source: HMDA data.*



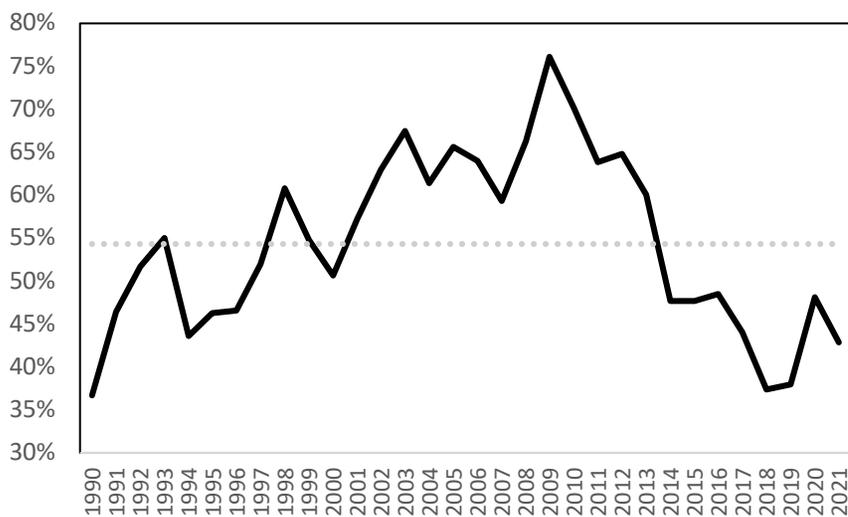
(a) Total mortgage origination volume, in billions.



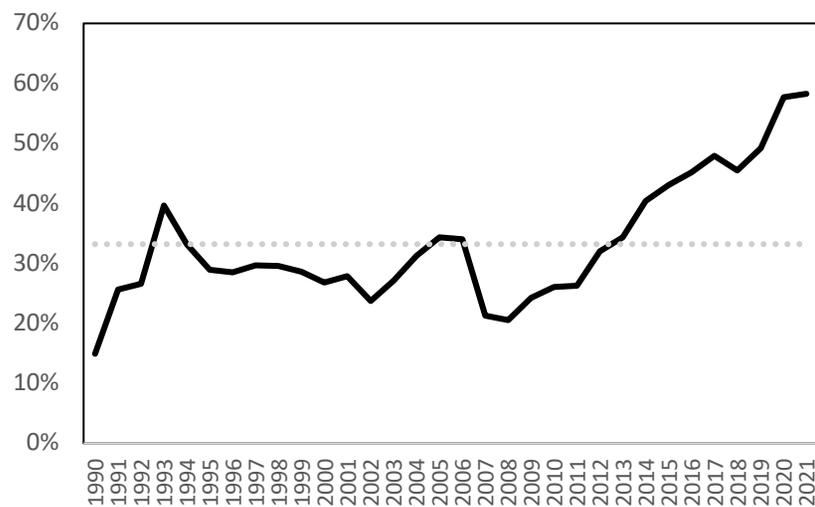
(b) Percentage of mortgage originations that are refinances.

**Figure A2: Bank Loan Sale Propensity and Shadow Bank Market Share over Longer Period**

Panel (a) of this figure shows the percentage of residential loan sold among the loans originated by banks per year during 1990-2021 period. The dashed line shows the sample mean. Panel (b) of this figure shows the percentage of residential mortgage loans originated by the shadow banks in the US. The dashed line shows the sample mean. Source: HMDA Data.



(a) Percentage of loans sold by banks



(b) Shadow bank market share (in %)