# THE DECLINING ROLE OF DEPOSITS IN CREDIT CREATION*

Naz Koont
Stanford University

Stefan Walz
Boston College

August 26, 2026

**Abstract**

Does deposit funding still drive credit creation? Deposit inflows affect credit directly, through banks' balance sheet response, and indirectly, through a feedback loop in which credit generates new deposits. Using exogenous deposit shocks and a structural model, we show that the total bank-credit impact of a $1 deposit inflow has fallen from $2.13 to $0.88 over two decades. Both channels have weakened: banks extend less credit out of each marginal deposit, reflecting greater costs of balance sheet expansion, and fewer credit proceeds recycle into stable deposits, consistent with declining deposit specialness and a greater use of non-bank alternatives. Central bank policy also matters: Quantitative Easing lowers the credit impact of a marginal deposit through higher capital costs, while Quantitative Tightening can raise it through scarcer liquidity. We show that non-bank credit is unlikely to offset the decline. These findings point to an implicit narrowing of banking that dampens credit creation and reshapes monetary and fiscal policy transmission.

*We thank Michele Andreolli, Luigi Bocola, Greg Buchak, Darrell Duffie, Zhiguo He, Arvind Krishnamurthy, Jane Li (Discussant), Wenhao Li, Rodney Ramcharan, Martin Schneider, Amit Seru, Phil Strahan, Adi Sunderam, Christopher Tonetti, Vikrant Vig and seminar participants at the USC Macro-Finance Workshop, Banking and the Deposit Franchise SITE Session, Boston College, UCSC, and Stanford GSB for helpful comments.

# 1 Introduction

Does deposit funding still drive credit creation? Stable bank deposits are widely thought to support bank lending and maturity transformation, as they relax banks' liquidity management (Diamond and Dybvig, 1983; Kashyap et al., 2002a; Drechsler et al., 2017). Beyond this direct effect, lending can generate new stable deposits, creating a feedback loop that amplifies the effect of deposit inflows on credit creation (Tobin, 1963; Friedman and Schwartz, 1963). Despite its potential importance, this amplification channel has received relatively little attention in the modern banking literature. As a result, we know little about the total effect of deposit inflows on bank credit provision, or about how that effect changes as the financial system evolves and intermediation shifts outside banks.

In this paper, we decompose the total effect of deposit inflows on credit into a direct credit response and an indirect amplification through the credit–deposit feedback loop. We document a pronounced decline in both channels over the past two decades, and use a structural model to trace these declines to shifts in primitives driven by higher balance sheet costs and non-bank disintermediation. Our results point to an implicit narrowing of banking: deposits have become substantially less effective at generating credit, even as formal deposit-taking remains integrated with lending, with implications for the cyclicality of credit, the transmission of monetary and fiscal policy, and the supply of safe assets.

We begin by providing empirical evidence on the magnitude and evolution of banks' credit–deposit multiplier using weekly data from 2005 to 2019. The multiplier embeds both banks' direct credit response to a deposit inflow and the recycling of credit into new deposits, and thus captures how an initial inflow propagates through successive rounds of credit creation. To estimate it, we use high-frequency variation in the timing of U.S. Treasury cash settlements as an instrument for deposit inflows and trace the cumulative deposit response over the following weeks.

We find that the credit-deposit multiplier is sizable: a \$1 deposit inflow ultimately raises cumulative deposits by \$2.06 through balance sheet propagation. However, the strength of this amplification has weakened markedly over time, with the multiplier declining from \$2.38 in the first half of our sample to \$1.36 in the second half. This decline indicates a weakening of one or both of the underlying margins that governs the loop: banks may lend less against a marginal deposit dollar, or fewer loan proceeds may return as stable deposits that support another round

of lending. Post-crisis regulation and the growth of non-bank savings and credit markets can affect both margins. Because the empirical multiplier embeds the two margins jointly, we next develop a structural model that disentangles them and recovers the effect of deposit inflows on bank and total credit.

Our structural model is built around the forces we seek to quantify: balance sheet costs, non-bank competition, and the recycling of credit into deposits. Depositors and borrowers choose between banks and non-bank outside options; a banking system sets deposit and loan rates, and a non-bank credit intermediary competes with banks for borrowers. The size and composition of the banking system's balance sheet are disciplined by the costs of maintaining capital and liquidity buffers, while non-bank competition shapes the elasticity of bank credit demand. We characterize the equilibrium of this economy in two layers. The first is a static portfolio equilibrium: taking household financial wealth as given, markets clear as a consistent allocation of that wealth across claims, with banks and non-banks optimizing subject to their constraints. This layer pins down banks' direct credit response to deposit inflows as an equilibrium object shaped by balance sheet costs and competition. The second layer is a monetary-creation channel, summarized by a single recycling rate: the fraction of credit that returns as stable deposits.

We recover the direct credit response structurally by differentiating the equilibrium conditions, and discipline the recycling rate with data. We estimate the model's parameters via deterministic minimum distance, following the logic of structural banking models that use externally disciplined demand elasticities and equilibrium restrictions to recover supply-side frictions (Wang et al., 2022). The banking sector's deposit and loan first-order conditions discipline the shadow costs of illiquidity and leverage, while the non-bank first-order condition disciplines the cost of expanding non-bank credit supply. We jointly fit these pricing residuals and the reduced-form credit-deposit multipliers in the early and late periods.

The estimated model delivers the two sufficient statistics that govern the credit–deposit multiplier: banks' credit response to deposit inflows and the recycling rate of credit proceeds into stable deposits. Both have declined sharply. The credit response falls from 0.88 to 0.58: over 30% less of each marginal deposit dollar is channeled into new credit, with the remainder absorbed by reductions in wholesale funding. The recycling rate falls from 0.65 to 0.47, representing a similar decline in the share of credit-generated payment flows that return to banks as the stable, low-cost funding that supports further credit expansion. Two complementary forces are consistent with the decline in the recycling rate. First, bank deposits may have become less

special as a savings instrument, with households and firms increasingly allocating funds toward non-bank alternatives such as money market funds. Second, the composition of bank credit has shifted from loans toward securities, which may generate different payment flows: loan proceeds may resettle more with households and firms, whereas securities route funds to institutional sellers less likely to hold stable deposits.

Turning to the central question of the paper, we find that the total bank-credit effect of a deposit inflow has declined from 2.13 to 0.88: each dollar of deposits today generates less than half of the bank credit it once did. The total bank credit effect combines banks' direct credit response with the amplification of deposits through the multiplier. The decline reflects both a weaker direct credit response and diminished deposit recycling, which weakens the amplification from one round to the next. As a result, the financial system has lost an important amplification channel through this implicit narrowing of the banking sector. Consequently, concerns about deposits migrating from banks to non-banks such as stablecoins may have more muted bank-credit effects today than they would have had in earlier periods, when deposit funding translated more powerfully into credit creation.

We next use the model to study the drivers of this decline in the credit impact of deposit flows. We decompose the decline into two channels. The first exercise moves the structural primitives governing bank behavior one at a time from their early- to their late-period estimates. Two forces emerge as first-order: the higher capital wedge and reduced recycling each lower the total bank-credit response by about half a dollar per deposit dollar. The liquidity wedge is small and works in the opposite direction, slightly offsetting the decline by raising the liquidity benefit of the marginal deposit dollar. The second exercise varies balance-sheet quantities, asking how Quantitative Easing and Tightening affect the banking sector's capacity to intermediate deposits into credit. The object here is the marginal bank-credit response to a new deposit dollar—not the direct credit-supply effect of QE itself. In the model, the two operate on the same balance-sheet frictions in opposite directions. QE injects reserves and deposits together: liquidity improves, but the balance sheet expands, tightening the capital position and lowering the marginal value of a deposit. QT instead can tighten the liquidity constraint, as emphasized by Acharya and Rajan (2024), removing reserves without a commensurate decline in deposit claims and raising the marginal value of stable funding. Together, the two exercises show that the decline reflects higher balance sheet costs, reduced deposit specialness, and the transformation of the balance sheet itself.

A natural question is whether the non-bank sector is able to compensate for the dampened credit response in banks. We approach this from two angles. First, we examine the direct competitive response of non-banks to a deposit shock and find that non-bank credit contracts rather than expands: the inflow reaches non-banks only through lower bank loan rates, which pull borrowers away from non-bank credit, so non-bank credit contracts. A more elastic non-bank sector deepens the contraction rather than closing the gap. Second, we ask how strong a spillover from non-recycled credit proceeds to non-bank funding conditions would need to be to restore the aggregate credit response to its early-period level. Even in the most favorable case, in which all non-recycled proceeds flow into money markets, restoring the early-period response requires a sizable pass-through from money-market depth to non-bank intermediation costs. More broadly, because non-bank intermediation chains are longer, neither recycled non-bank proceeds nor leaked bank flows are likely to compensate for the loss in bank recycling. On the direct margin, non-banks contract rather than substitute; on the funding margin, they do not replace the multiplier.

Overall, we find that deposits injected into the banking system generate far less cumulative amplification of credit than they did in the past, and non-bank intermediaries are unlikely to fully offset this decline. As a result, both money and credit have become less responsive to deposit inflows, dampening the amplification of the credit cycle. In this sense, our findings document an implicit narrowing of banking, driven not by institutional separation, but by regulation and non-bank competition in the payments market, which has weakened the traditional feedback between deposits and credit. Deposits have become less effective at supporting credit creation, even as formal deposit-taking remains integrated with lending.

**Literature Contribution**

Whether deposit funding matters for bank credit is a classic question in banking. It connects to seminal theories of why banks jointly provide deposit-taking and credit provision services (Diamond and Dybvig, 1983; Kashyap et al., 2002a), as well as to a more recent debate about the consequences of non-bank intermediaries eroding banks' deposit base (Whited et al., 2022). Our contribution is to quantify how deposit inflows translate into bank credit creation and to link this relationship to underlying economic primitives. By doing so, we provide a framework for evaluating how shifts in the composition of money and savings, including the migration of funds toward non-bank instruments such as money market funds or stablecoins, affect the supply of credit in the modern financial system.

Moreover, the idea that deposits and loans are connected through a feedback loop back at least to Tobin (1963), yet remains underexplored. The literature has tended to emphasize two separate views: one in which deposits fund loans, and another in which loans create deposits. In the traditional "intermediary" view, banks collect deposits from savers, hold a fraction as reserves, and lend out the remainder, so deposit market size constrains lending (Kashyap et al., 2002b; Bernanke and Blinder, 1988; Xiao, 2019). In the "money creation" view, lending itself creates deposits (Tobin, 1963; Donaldson et al., 2018), and in the absence of a binding reserve requirement is shaped by regulations, loan demand, and banks' risk appetite. Empirically, micro evidence shows that banks create deposits when they lend (Werner, 2014), aggregate deposits far exceed cash deposits (Thakor and Edison, 2024), and reduced-form estimates document a positive elasticity from deposit shocks to credit growth (Kundu et al., 2021). Our contribution is to bridge these perspectives by quantifying the strength of the endogenous credit-deposit multiplier and the amount of deposits and credit that are ultimately created from an initial deposit inflow. This multiplier highlights that deposits both constrain and are created by lending: the availability and stability of deposits shape banks' willingness to expand credit, while banks' lending decisions feed back into the deposit base. By decomposing the evolution and drivers of the multiplier, we show how the balance between these forces shifts over time.[1]

The credit–deposit multiplier relates to amplification mechanisms emphasized in the macro-finance literature. A large body of work studies how intermediary balance sheet strength shapes credit supply through net-worth-based amplification. Seminal contributions show how collateral values and internal liquidity constrain borrowing and investment (Kiyotaki and Moore, 1997; Holmstrom and Tirole, 1997), while Bernanke et al. (1999) highlights how stronger balance sheets reduce financing frictions and expand borrowing. Building on these foundations, intermediary-based macro models emphasize how intermediary net worth governs leverage and credit creation (Gertler and Kiyotaki, 2010; Gertler and Karadi, 2011), with implications for risk premia and asset prices (Brunnermeier and Sannikov, 2014; He and Krishnamurthy, 2013). More recent heterogeneous-bank frameworks study how these elasticities vary across institutions (Corbae and D'Erasmo, 2021; Jamilov and Monacelli, 2025). Our framework studies a closely related amplification mechanism operating through funding rather than intermediary net worth separately. The marginal lending response captures how strongly banks expand credit in

[1]While we estimate a long-term liquidity multiplier that supports credit provision, Denbee et al. (2021) and Piazzesi and Schneider (2018) study a related short-term payments liquidity multiplier, showing how transactional liquidity shapes banks' reserve management and monetary transmission.

response to deposit inflows and therefore plays a role analogous to the credit supply elasticities emphasized in this literature. While we focus on short-run balance sheet propagation through funding flows, the two mechanisms would interact at longer horizons. Lending generates profits that raise depositor wealth and bank equity, which in turn strengthen the feedback loop between deposit creation and credit supply. In this sense, the credit–deposit multiplier complements the net-worth-based amplification mechanisms emphasized in the macro-finance literature.

We highlight the importance of accounting for the recycling of credit proceeds into stable deposit funding. This connects to recent work on deposit franchise value and depositor heterogeneity in both the cross-section and the time series (Drechsler et al., 2017, 2021; Li et al., 2023; DeMarzo et al., 2024; Blickle et al., 2025; Egan et al., 2025; Lu et al., 2024; Koont et al., 2024), as well as to the growing literature on non-bank intermediaries and the implications of a shrinking traditional banking sector (Begenau and Landvoigt, 2022; Buchak et al., 2024; Schneider et al., 2025). The recycling rate measures how much of banks' own credit creation returns as stable funding: a high rate means that credit provision replenishes the deposit franchise, while its decline implies that credit expansion generates less franchise value per dollar. Similarly, as non-banks disintermediate banks on both sides of their balance sheets, a smaller share of funding flows through this amplification mechanism, raising the question of whether non-bank intermediation can compensate. We show that it is unlikely to: in equilibrium, non-bank credit contracts rather than expands in response to a deposit inflow, and because non-bank intermediation chains are longer, neither recycled non-bank proceeds nor leaked bank flows are likely to make up the loss in bank recycling.

# 2 Defining the Total Effect of Deposit Inflows on Bank Credit

## 2.1 Banks' Lending Response to Deposit Inflows

The central question in this paper is how deposit inflows translate into bank credit creation. When deposits enter the banking system, banks may expand their balance sheets by increasing lending or other assets. A natural starting point for thinking about this relationship is banks' marginal credit response to deposits, which we denote by $\lambda = dL/dD$: how much additional credit banks supply when their deposit funding increases.

However, the total effect of deposit inflows on bank credit may differ from this marginal response because deposits and credit are jointly determined within the banking system. When banks extend credit, the associated payment flows circulate, and a fraction, which we denote by $\theta$, returns as the stable deposits that fund further credit. As a result, the ultimate increase in credit following a deposit inflow depends both on banks' marginal credit response $\lambda$ and on the recycling rate $\theta$.

Our object of interest is the total effect of a deposit inflow $D_0$ on bank credit $L$, given by $dL/dD_0$, which embeds both the marginal response $\lambda$ and the amplification generated through $\theta$. The remainder of this section develops this object: how deposits and credit are jointly determined in a fractional-reserve system, a simple example of the propagation mechanism, and a formal decomposition into these two margins.

## 2.2 The Link Between Deposits and Bank Credit

To understand how deposit inflows translate into bank credit creation, we begin by discussing how deposits and credit are jointly determined in the modern banking system, where banks operate with fractional reserves and are not constrained by binding reserve requirements. In such an environment, the effect of deposit inflows on credit depends on how bank balance sheets adjust and whether credit provision generates additional deposits within the financial system.

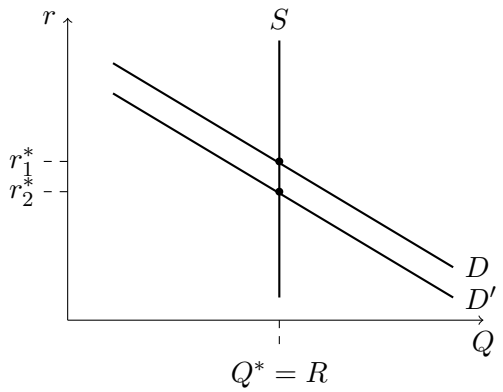
As a benchmark, consider a fully narrow banking system in which deposits are backed one-for-one by central bank reserves, as illustrated in Figure 1 panel (a). Under narrow banking, the aggregate quantity of deposits is fixed by reserves, so that $Q^* = R$, where $R$ denotes aggregate reserves. In this case, the aggregate supply of deposits is perfectly inelastic. Changes in the demand for deposits shift equilibrium deposit rates $r^*$ but do not affect the quantity of deposits.

In contrast, in a fractional reserve banking system as in Figure 1 panel (b), deposits are jointly determined by (i) households' willingness to hold deposits and (ii) banks' willingness to expand their balance sheets and issue deposit liabilities (e.g., to fund loans and securities). Unlike in a narrow system, banks are not required to hold reserves one-for-one against deposits. Instead, they choose balance sheet size to maximize their profits subject to liquidity and capital management or regulatory constraints that limit how much asset expansion can be supported by a given stock of liquid resources. In this environment, shifts on either side of the market affect both equilibrium rates and quantities. Bank balance sheet expansion can therefore be

**Figure 1**: Deposit determination under narrow and fractional reserve banking

**Narrow Banking**

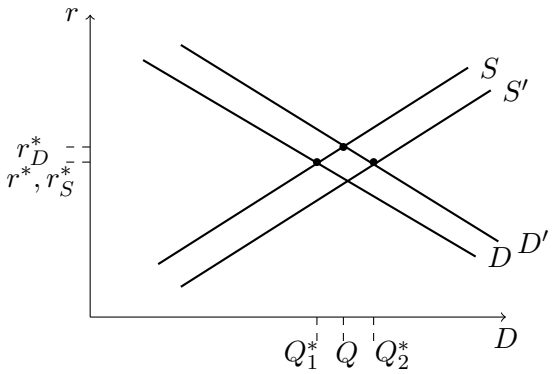
| **Assets** | | **Liabilities** | |
|---|---|---|---|
| Reserves | $R$ | Deposits | $R$ |
| | $R$ | | $R$ |




(a) Narrow banking: deposits are backed one-for-one by reserves, so deposit quantity is pinned down by aggregate reserves. Demand shifts change equilibrium rates but not quantities.

**Fractional Reserve Banking**

| **Assets** | | **Liabilities** | |
|---|---|---|---|
| Reserves | $R$ | Deposits | $D$ |
| Loans / Securities | $A$ | | |
| | $R + A$ | | $D$ |




(b) Non-narrow banking: deposits are jointly determined by an upward-sloping deposit supply curve (banks' willingness to expand balance sheets) and a downward-sloping deposit demand curve (households' willingness to hold deposits). Shifts in either demand or supply move both equilibrium rates and quantities.

initiated either by an increase in deposit demand, arising from changes in household preferences or portfolio allocations toward deposits, or by a shift in credit supply, reflecting improvements in lending opportunities or changes in banks' regulatory environment that make asset expansion more attractive.[2]

We see evidence of this deposit creation in the U.S. banking sector when we examine the evolution of key monetary aggregates over time. We focus on two standard measures of money that are most directly linked to the banking system's role in money creation. M0, or the monetary base, includes currency in circulation and reserves held by banks at the central bank

[2]This example abstracts from banks' non-deposit wholesale funding adjustment. Allowing for this margin would imply that balance sheet expansion occurs net of funding adjustments, as formalized in our model in Section 4.

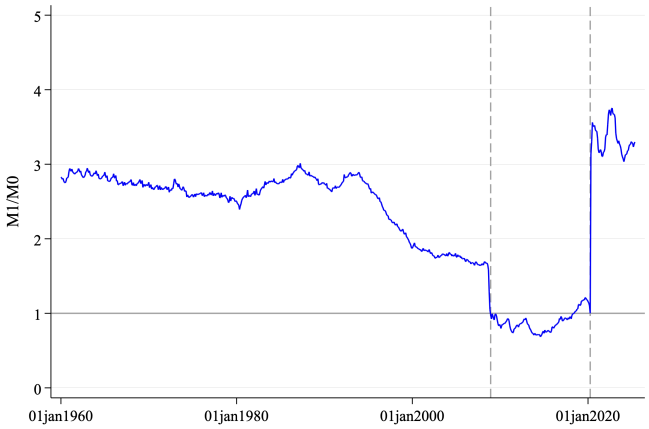



**Figure 2**: Evolution of M1/M0. Monthly data is from FRED (1960–2025). Vertical lines denote the onset of QE after the 2008 financial crisis and pandemic-related stimulus. In May 2020, savings deposits were reclassified as transaction accounts and added to M1.

and represents the supply of base money injected by the central bank. M1, or narrow money, includes currency in circulation as well as demand deposits and other highly liquid checkable deposits. Comparing these two aggregates illustrates how the stock of bank deposits compares to reserves. Figure 2 plots the evolution of M1 relative to M0. The fact that the ratio fluctuates over time, and generally exceeds 1, is consistent with banks creating deposits through credit creation.[3]

The fact that banks create deposits to expand their balance sheets does not, by itself, imply the existence of a feedback loop between deposits and credit. Such amplification arises only if banks respond to additional deposits by expanding credit provision, so that the marginal credit response to deposits, $dL/dD$, is positive. This condition is likely to hold in practice because deposits, when they constitute stable funding, relax banks' liquidity constraints at the margin and thereby support additional balance sheet expansion. As a result, deposits created through credit can return to the banking system as funding and support further balance sheet expansion. The ultimate increase in credit can therefore exceed banks' initial marginal credit response.

**Table 1**: Simple Example: Credit-Deposit Multiplier

In this numerical example, a $100 deposit inflow replaces wholesale funding (b). The banking sector expands credit by $\lambda = 0.7$ per dollar of stable deposits, and loan proceeds appear as payment balances (PB) that circulate as borrowers spend (c). A fraction $\theta = 0.9$ of these flows settles as stable deposits and the remainder as wholesale funding (d), initiating the next round of credit creation (e) and recycling (f).

(a) Initial

| **Assets** | | **Liabilities** | |
|---|---|---|---|
| R | $100 | W | $100 |
| | $100 | | $100 |

(b) Deposit Inflow

| **Assets** | | **Liabilities** | |
|---|---|---|---|
| R | $100 | D | $100 |
| | | W | $0 |
| | $100 | | $100 |

(c) Credit Creation

| **Assets** | | **Liabilities** | |
|---|---|---|---|
| R | $100 | D | $100 |
| L | $70 | PB | $70 |
| | $170 | | $170 |

(d) Deposit Recycling

| **Assets** | | **Liabilities** | |
|---|---|---|---|
| R | $100 | D | $163 |
| L | $70 | W | $7 |
| | $170 | | $170 |

(e) Credit Creation

| **Assets** | | **Liabilities** | |
|---|---|---|---|
| R | $100 | D | $163 |
| L | $114 | PB | $44 |
| | | W | $7 |
| | $214 | | $214 |

(f) Deposit Recycling

| **Assets** | | **Liabilities** | |
|---|---|---|---|
| R | $100 | D | $203 |
| L | $114 | W | $11 |
| | $214 | | $214 |

## 2.3 A Simple Example of the Credit-Deposit Feedback Loop

The example in Table 1 illustrates how an initial deposit inflow can propagate through the banking system via successive rounds of balance sheet expansion, using a consolidated balance sheet for the aggregate banking system. A deposit-demand shock first lets banks replace wholesale funding with stable deposits, leaving total assets unchanged but improving the funding mix. Because stable deposits relax banks' liquidity management, suppose that banks expand credit. The credit proceeds initially appear as flighty payment balances at the issuing bank, and the intermediary view of banks stops here, whereby an increase in deposit funding leads to an effect on banks' willingness to extend credit.

However, the quantity of deposits is endogenously determined, and this is what sets the feedback loop in motion. As borrowers spend the credit proceeds, payments circulate through the economy. Over time, some fraction of these proceeds settles as deposit funding in the banking system, and the rest as flightier wholesale funding. The recycled deposits again relax banks' liquidity management and support another round of credit expansion. Some of the resulting credit proceeds again return as deposit funding, continuing the feedback loop.

[3]The ratio drops below 1 during the QE period primarily because central bank reserves expanded disproportionately relative to the bank deposit categories in M1.

**Table 2**: Aggregate Banking System: Incremental Expansion by Round

| **Round** $h$ | **Deposits** | **Credit** |
|---|---|---|
| 1 | \$100.0 | \$70.0 |
| 2 | \$63.0 | \$44.1 |
| 3 | \$39.7 | \$27.8 |
| 4 | \$25.0 | \$17.5 |
| **Limit** $h \to \infty$ | **\$270.3** | **\$189.2** |

This table reports the round-by-round expansion of deposits and loans under the credit-deposit multiplier, using the numerical example from Table 1. Each round's new credit creation equals $\lambda = 0.7$ times newly retained deposits, which in turn equal $\theta = 0.9$ times the previous round's credit. Successive rounds therefore shrink by the factor $\lambda\theta = 0.63$, and the final row reports the limiting value of the resulting geometric series.

This process continues until it converges, as shown in Table 2. Amplification arises because deposits are superior to wholesale funding: they relax the banking sector's liquidity management, and proceeds from credit creation partially recycle into new deposits. The interaction between banks' credit response and the retention of payment flows governs the strength of the credit–deposit multiplier.

## 2.4 The Total Effect of Deposit Inflows on Bank Credit

Formally, we define the Credit-Deposit Multiplier $m_D$ as the total change in the deposit base, $\Delta D_{total}$, generated by an initial deposit inflow $\Delta D_0$. Two sufficient statistics govern this process. The first is $\lambda$, banks' marginal credit response to deposit inflows.[4] The second is $\theta$, the marginal propensity for credit-financed expenditures to be redeposited within the banking system as stable bank deposits.

Given a deposit inflow $\Delta D_0$, banks extend new credit in the amount $\lambda\Delta D_0$, of which a fraction $\theta$ returns as deposits, raising the deposit base to $\Delta D_0 + \lambda\theta\Delta D_0$. These new deposits $\lambda\theta\Delta D_0$ then support additional credit provision of $\lambda^2\theta\Delta D_0$, of which a fraction $\theta$ recycles as deposits, and the process continues. Iterating forward, we can express the total deposit expansion

[4] In general, the first-round response depends on the source of the inflow—a preference shift, a reserve injection, or recycled payment flows—while every subsequent round involves the same response to recycled deposits. Section 4.8.1 formalizes this distinction; the multiplier logic here requires only the recycled-deposit response, which governs all rounds beyond the first.

from an initial inflow $\Delta D_0$ as a geometric series:

$$\Delta D_{t-1,t+h} = \Delta D_0 \left(1 + \lambda\theta + (\lambda\theta)^2 + \cdots + (\lambda\theta)^h\right). \tag{1}$$

The corresponding credit-deposit multiplier at horizon $h$ is:

$$m_D^h = 1 + \lambda\theta + (\lambda\theta)^2 + \cdots + (\lambda\theta)^h. \tag{2}$$

As $h \to \infty$, the process converges (provided $|\lambda\theta| < 1$), yielding the long-run credit-deposit multiplier:

$$m_D = \frac{1}{1 - \lambda\theta}. \tag{3}$$

The credit-deposit multiplier $m_D$ describes how an initial deposit inflow expands the deposit base through successive rounds of credit creation and redepositing. To recover the total effect of deposit inflows on bank credit $\Delta L_{total}^{bank}$, we combine this amplification mechanism with banks' marginal lending response to deposits,

$$\Delta L_{total}^{bank} = \lambda m_D \Delta D_0. \tag{4}$$

In Appendix A.1 we consider whether non-bank institutions can generate a similar funding multiplier. They are unlikely to: because non-bank liabilities do not serve as the economy-wide settlement asset, credit proceeds return to non-banks only through longer intermediation chains, attenuating at each link, and the limited recycling that occurs within these subsystems typically does not support long-term credit provision to the real economy. We therefore focus on banks in the main text and abstract from non-bank multipliers.

# 3 Empirical Analysis

## 3.1 Empirical Strategy and Data

To estimate banks' credit-deposit multiplier, we measure how an initial deposit inflow translates into a total change in deposits. A natural way to implement this is to estimate local projections

of deposits at horizon $h$ on a shock to deposits at time $t$:

$$\Delta D_{t+h} = \alpha_h + m_D^h \cdot \Delta D_t + \Gamma_h X_t + \varepsilon_{t+h}. \tag{5}$$

In Equation (5), $\Delta D_t$ is deposit growth from period $t-1$ to $t$ and $\Delta D_{t+h}$ is cumulative deposit growth from $t-1$ to $t+h$. $X_t$ denotes a vector of controls dated at time $t$, and $\varepsilon_{t+h}$ is the error term. In this setting, the coefficient $m_D^h$ represents the credit-deposit multiplier at horizon $h$. As $h$ grows large, $m_D^h$ converges to the long-run multiplier $m_D$.

If all other determinants of deposit growth were fully accounted for in $X_t$, then local projections would deliver unbiased estimates of $m_D^h$ at each horizon, converging to the true multiplier $m_D$ for large $h$. However, in practice, many relevant factors that influence deposit dynamics, such as local demand conditions, interest rate expectations, or shifts in portfolio preferences, are difficult or impossible to observe. As a result, $\Delta D_t$ may be correlated with the error term $\varepsilon_{t+h}$, giving rise to endogeneity and biasing our estimate of $m_D$. For instance, we may overestimate $m_D$ if $\Delta D_t$ and $\Delta D_{t+h}$ co-move positively because of an unobservable factor that drives deposit flows at both horizons, such as a surge in aggregate risk appetite or liquidity demand. To address this endogeneity concern, we seek a shifter of contemporaneous deposit growth, $\Delta D_t$, that is uncorrelated with the unobservable factors that jointly drive both current and future deposit growth. We construct such a shifter using high-frequency innovations to Treasury cash flows from the Daily Treasury Statement (DTS).

Treasury cash flows reflect the government's operating receipts and outlays, including corporate and income taxes, tax refunds, federal salaries, grants to states, vendor payments, and interest on Treasury debt. These flows move funds between the public sector and private bank deposits. We aggregate daily Treasury withdrawals and deposits of operating cash into weekly net cash flows, $CF_t$, treating deposits of operating cash as negative values so that positive realizations correspond to net Treasury payments to the private sector. Our goal is to isolate variation in routine Treasury operations, so we exclude categories that may reflect financial system transfers or endogenous policy responses.[5] The cash-flow innovation is constructed using the full weekly DTS series. After constructing the innovation, the estimation sample excludes January weeks because the turn of the year combines large scheduled federal payments, tax-

[5]We include all operating cash categories except for internal transfers to and from the Federal Reserve, the TGA, and depository institutions, and counter-cyclical means-based transfers including unemployment insurance, food assistance, and TANF, which may correlate with broader economic conditions. We exclude October 2008, which coincides with extraordinary fiscal and financial interventions during the Great Financial Crisis.

withholding timing, and year-end balance-sheet adjustments that add noise to the weekly link between Treasury settlements and core deposit growth.

We next purge this weekly series of predictable variation by residualizing on quarter-year fixed effects $\alpha_q$, calendar week-of-year fixed effects $\alpha_w$, and the lag set in Equation (6). The short lags absorb settlement dynamics, while the longer lags capture quarterly recurrence in weekly Treasury flows. We winsorize the residual at the 1% level and normalize by lagged aggregate deposits, $Z_t \equiv \widetilde{u}_t / D_{t-1}$, so that a given dollar shock is scaled relative to the deposit base over which it is absorbed.[6]

$$CF_t = \alpha_q + \alpha_w + \sum_{k \in \mathcal{K}} \rho_k \, CF_{t-k} + u_t, \qquad \mathcal{K} = \{1, 2, 3, 4, 12, 13\}. \tag{6}$$

This innovation captures unexpected timing variation in Treasury settlement flows. The identifying variation is the administrative timing of Treasury settlements after removing predictable seasonal, short-run, and quarterly recurrence. We test the scaled instrument for serial correlation, and Ljung–Box tests through six lags do not reject the null at conventional levels in any sample.

Instrument validity requires that the relevance and exclusion assumptions be satisfied. The relevance condition is supported by the first-stage estimates reported in Appendix Table A.1. Intuitively, Treasury payments and receipts credit and debit private bank accounts, generating changes in aggregate core deposits at weekly horizons. The exclusion restriction requires that innovations in Treasury cash-flow timing affect future deposit growth only through its impact on contemporaneous deposit growth $\Delta D_t$, and not through any other channel. This is plausible because, after removing predictable seasonality and persistence, the remaining high-frequency variation in Treasury cash-flows largely reflects administrative settlement timing across a large number of payment streams.

We implement this strategy in a 2SLS framework. We measure deposits using aggregate core deposits from FRED, defined as total deposits held by all U.S. commercial banks excluding large time deposits. Deposit growth $\Delta D_t$ is the log difference of core deposits, winsorized at the 1% level. Our sample runs weekly from 2005 to 2019; we estimate the multiplier over the full sample and separately for an early period (2005–2010) and a late period (2011–2019). In the vector of controls $X_t$ we include the lag of deposit growth $\Delta D_{t-1}$. Because some Treasury

[6]We focus on the surprise component of government spending, while Greenwood et al. (2015) instrument for Treasury bill issuance using the predictable seasonal component of bill supply, the fitted values from regressing four-week changes in bills outstanding on week-of-year dummies.

payments settle with short delays, we instrument contemporaneous deposit growth with both the contemporaneous and one-period-lagged innovation, so that the first stage accommodates settlement effects that reach deposit balances in the following week. Our first-stage specification is:

$$\Delta D_t = \alpha + \beta_1 Z_t + \beta_2 Z_{t-1} + \Gamma_h X_t + \mu_t. \tag{7}$$

The first-stage estimates reveal a shift in the timing with which Treasury cash flows reach bank deposits. In the full sample and in the early period, both the contemporaneous and lagged innovations enter positively, consistent with fiscal payments flowing into core deposit accounts over the contemporaneous and following week. In the late period, however, the lagged innovation no longer predicts deposit growth once the contemporaneous innovation is included. Accordingly, our preferred subperiod specification uses both $Z_t$ and $Z_{t-1}$ in the early period, and uses only $Z_t$ in the late period. This preserves the early-period information in both timing margins while avoiding a weak late-period lagged instrument.

In the second stage, we regress cumulative deposit growth at horizon $h$ on instrumented deposit growth at time $t$ within the local projection framework. For horizon $h$, the second stage regression equation is,

$$\Delta D_{t-1,t+h} = \alpha_h + m_D^h \widehat{\Delta D_t} + \Gamma_h X_t + \varepsilon_{t+h}. \tag{8}$$

We consider three potential threats to identification. First, in the post-crisis period, Treasury cash flows may affect reserves through movements in the Treasury's account at the Federal Reserve. This could generate co-movement between deposits and reserves that is not driven by deposit inflows. In Appendix A.1, we show that our results are not sensitive to controlling for contemporaneous changes in aggregate reserves, indicating that the estimates primarily reflect deposit inflows rather than reserve co-movement at weekly frequency. Second, serial dependence in Treasury settlement timing could cause current cash-flow innovations to predict future deposit growth within the local-projection horizon. We address this by purging both short-run and quarterly recurrence from Treasury cash flows and by reporting serial-correlation tests for the resulting instrument. Third, deposit inflows could affect credit through real-side channels rather than balance sheet propagation. For example, deposit-funded lending may stimulate economic activity, raising incomes and wealth and thereby shifting out deposit demand. However,

such real-side adjustments operate over longer horizons, and the six-week window we study is likely too short for these mechanisms to play a meaningful role. As a result, real-side feedback effects are unlikely to drive our estimated deposit multiplier.

Thus, our empirical strategy allows us to estimate how cumulative deposit growth responds to an initial exogenous inflow through balance sheet propagation, thereby recovering a causal estimate of the deposit multiplier.

## 3.2 Estimates of the Credit-Deposit Multiplier

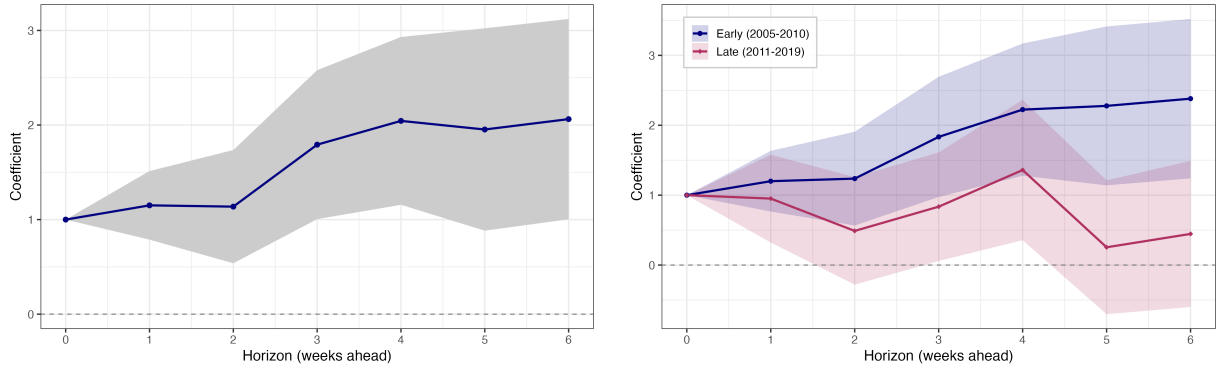
Figure 3 visualizes the coefficient estimates for our IV specification in Equation (8) at horizons $h$ from $0$ to $6$, with shaded bands indicating 95% confidence intervals based on Newey–West standard errors with a lag length equal to the horizon $h$. Panel (a) captures the estimation for the complete sample. This specification captures how an initial funding inflow propagates through the banking system over time, including through flows of funds between banks and across institutions. We define the long-run multiplier $m_D$ as the peak cumulative response, resulting in a value of 2.06. This implies that a \$1 exogenous deposit inflow ultimately raises cumulative deposits in the banking system by approximately \$2.06. The impulse response indicates that initial deposit inflows are gradually re-deposited within the banking system, leading to additional rounds of deposit growth across local institutions. These effects level off after several weeks, suggesting that the cumulative response has converged to the overall deposit multiplier.

We next examine its time variation. We split our time sample into early and late periods, estimating the multiplier over the years 2005–2010 and 2011–2019. Panel (b) compares the dynamics of the deposit multiplier across the two samples. Comparing the magnitudes shows a marked decrease in the latter sample: the early-period response peaks at 2.38, while the late-period response peaks at 1.36.

The reduced-form multiplier embeds two distinct economic forces: a direct effect, captured by banks' marginal propensity to expand credit out of deposit inflows, and an indirect effect, captured by the share of credit proceeds that recycle back into deposit funding. The observed decline could therefore reflect changes in either channel, driven by post-crisis regulatory reforms and policy, the expanding role of non-bank intermediaries, or shifts in household behavior. Disentangling these forces and quantifying the total effect of deposit flows on credit requires additional structure, which we turn to next.

**Figure 3**: Deposit Multiplier Estimates

This figure reports local projection estimates of the dynamic response of deposits to a deposit shock at time $t$, estimated using an instrumental variables approach. Panel (a) shows the full-sample estimate from 2005 to 2019. Panel (b) compares estimates for the early (2005–2010) versus late (2011–2019) samples. Shaded bands denote 95% confidence intervals.




(a) Full Sample

(b) Early vs. Late

# 4 Model

## 4.1 Overview

The model features four types of agents: depositors, borrowers, a representative banking system, and a representative non-bank credit intermediary. The key exogenous variables are the current policy rate $r^S$, the equity capital of the banking sector $E$, and the level of central bank reserves $R$, which together characterize the state of the world. Given these variables, depositors allocate their wealth across bank deposits, market savings, and cash, while borrowers choose among bank credit, non-bank credit, and not borrowing. Banks set deposit and loan rates to maximize profits, subject to capital and liquidity constraints. The non-bank credit intermediary sets its lending rate to maximize profits. Finally, competitive funding markets clear the associated balance sheets. Money market funds allocate household market savings between short-term non-bank funding claims and safe assets, while a broader institutional sector supplies bank wholesale funding and the remaining market funding of non-bank credit.

We characterize the equilibrium in two layers. The first is a static portfolio equilibrium: taking household financial wealth as given, markets clear as a consistent allocation of financial claims, with the banking sector and the non-bank intermediary optimizing subject to their

constraints. Within this layer there are no payment flows to trace: the equilibrium is a settled allocation of stocks where all markets clear. In particular, the lending response $\lambda$ captures the banking sector's optimal reallocation of a marginal dollar of stable funding between new lending and reduced wholesale funding, given how the capital and liquidity slacks re-bind.

The second layer is the monetary-creation channel, where payment flows arise: lending generates payments that circulate and partially resettle in the banking system, and we summarize this process with a single recycling parameter $\theta$, the fraction of lending that returns as stable deposits. Because we seek the marginal response to a shock rather than the full monetary equilibrium, money creation enters the model only through $\theta$.

Our objects of interest, bank credit and total credit, are therefore governed by two objects: the lending response $\lambda$, evaluated at the equilibrium balance sheet, and the recycling rate $\theta$. We recover $\lambda$ structurally by differentiating the equilibrium, and use the data to discipline the monetary layer summarized by $\theta$.

## 4.2 Depositors

There is a mass $W^d$ of depositors. Each depositor $i$ makes a discrete choice over a menu $k \in \{D, M, C\}$ that includes a bank deposit $D$, market savings (e.g., money market fund shares) $M$, and cash $C$. These options differ in the yields they offer and the liquidity services they provide. Depositor $i$'s utility from option $k$ is

$$u_i^k \;=\; \alpha_d \, r^k + \xi_k + \varepsilon_i^k, \tag{9}$$

where $\alpha_d$ is the rate sensitivity of depositors and $\varepsilon_i^k$ is the idiosyncratic taste shock for depositor $i$ over option $k$. Bank deposits pay rate $r^d$ and provide liquidity services $\xi_D$, market savings pay the policy rate $r^S$ with liquidity services normalized to zero ($\xi_M = 0$), and cash provides liquidity services $\xi_C$ but pays no interest ($r^C = 0$).

We assume that the idiosyncratic taste shock $\varepsilon_i^k$ is i.i.d. Type-1 extreme value, the standard multinomial logit assumption, allowing for a closed form solution of the choice probabilities. The market share of option $k \in \{D, M, C\}$ is

$$s_k(r^d) \;=\; \frac{e^{\xi_k + \alpha_d \, r^k}}{e^{\xi_D + \alpha_d \, r^d} \;+\; e^{\alpha_d \, r^S} \;+\; e^{\xi_C}}, \tag{10}$$

and the demand for each instrument is given by

$$X_k(r^d) \;=\; s_k(r^d) \cdot W^d, \qquad k \in \{D, M, C\}. \tag{11}$$

Because the multinomial logit shares sum to one, depositor allocations satisfy the wealth identity

$$D + M + C \;=\; W^d. \tag{12}$$

Equation (12) closes the depositor side of the model: every dollar of household financial wealth is allocated across one of the three options, so any reallocation across categories must net to zero.

## 4.3 Borrowers

There is a mass $W^b$ of borrowers. Each borrower $j$ makes a discrete choice over a menu $k \in \{\ell, nb, \emptyset\}$ that includes bank credit $\ell$, non-bank credit $nb$, and not borrowing $\emptyset$. Borrower $j$'s utility from option $k$ is

$$u_j^k \;=\; -\alpha_\ell \, r^k + \xi_k + \eta_j^k, \tag{13}$$

where $\alpha_\ell$ is the rate sensitivity of borrowers and $\eta_j^k$ is the idiosyncratic taste shock for borrower $j$ over option $k$. Bank credit charges rate $r^\ell$ and provides non-rate services $\xi_\ell$, non-bank credit charges rate $r^{nb}$ and provides non-rate services $\xi_{nb}$, and the utility of not borrowing is normalized to zero ($r^\emptyset = \xi_\emptyset = 0$).

As with deposit demand, we assume that the idiosyncratic taste shock $\eta_j^k$ is i.i.d. Type-1 extreme value. The market share of option $k \in \{\ell, nb, \emptyset\}$ is

$$s_k(r^\ell, r^{nb}) \;=\; \frac{e^{\xi_k - \alpha_\ell \, r^k}}{1 \;+\; e^{\xi_\ell - \alpha_\ell \, r^\ell} \;+\; e^{\xi_{nb} - \alpha_\ell \, r^{nb}}}, \tag{14}$$

and credit quantities are given by[7]

$$L_k(r^\ell, r^{nb}) \;=\; s_k(r^\ell, r^{nb}) \cdot W^b, \qquad k \in \{\ell, nb\}. \tag{15}$$

[7]For ease of notation we denote the banking sector's loan demand $L_\ell(r^\ell, r^{nb})$ by $L(r^\ell, r^{nb})$ in the remainder of the paper.

Because the logit shares sum to one, borrower allocations satisfy the identity

$$L + L_{nb} + L_{\emptyset} \;=\; W^b, \tag{16}$$

where $L_{\emptyset} = s_{\emptyset}(r^{\ell}, r^{nb}) \cdot W^b$ denotes the mass of potential borrowers who choose not to borrow at prevailing rates. Equation (16) is parallel to the depositor wealth identity (12) on the savings side, with the understanding that the $\emptyset$ component reflects unrealized borrowing capacity rather than a claim held on any balance sheet.

## 4.4 Banking System

### 4.4.1 Profit Maximization

We model a representative banking sector that chooses the deposit rate $r^d$ and loan rate $r^{\ell}$ to maximize profits. It takes as given the policy rate $r^S$, equity capital $E$, and the demand functions of depositors $D(r^d)$ and borrowers $L(r^{\ell})$. The bank is a price-taker in the wholesale funding market, where it borrows at the policy rate $r^S$.[8] The banking sector also holds central bank reserves $R$, which we treat as an exogenous, system-level quantity. The balance sheet constraint is

$$L(r^{\ell}) + R = E + D(r^d) + W, \tag{17}$$

so that, for any $(r^d, r^{\ell})$, wholesale funding is determined residually as

$$W = L(r^{\ell}) + R - E - D(r^d). \tag{18}$$

The banking sector's profit maximization problem is given by[9],

$$\max_{r^d, r^{\ell}} \; (r^{\ell} - r^S - \mu_{\ell})L \;+\; (r^S - r^d - \mu_d)D \;+\; \xi \ln\big(K\big) \;+\; \tau \ln\big(\Lambda\big). \tag{19}$$

[8] We use "wholesale funding" as shorthand for non-core funding, encompassing rate-sensitive or runnable liabilities. The economically relevant distinction in the model is the funding's stability: $D$ denotes stable core deposits while $W$ absorbs the remainder.

[9] We present the bank's profit maximization after substituting out the balance-sheet identity. Reserves earn the policy rate $r^S$ and are financed at the margin by wholesale funding at the same rate, so reserve income nets out. The substitution places the opportunity cost $r^S$ inside the loan and deposit margins and leaves a term $r^s E$, which is constant with equity fixed and does not affect the banking sector's choice of rates.

In (19), the first two terms capture profits from loan and deposit markets respectively, net of per-unit operating costs $\mu_\ell$ and $\mu_d$, which capture expenses such as borrower screening and servicing on the loan side and branch operations and payment services on the deposit side. Unlike these per-unit costs of intermediation, the remaining terms capture the costs associated with capital and liquidity management, which depend on the size and composition of the bank's balance sheet and which we describe in detail below.

We model the costs of capital and liquidity management with log penalties, which reduce profits at an increasing rate as the banks' capital or liquidity slack approaches zero. These penalties are a reduced-form representation of the banking sector's revealed preference for maintaining capital and liquidity buffers. This preference may stem from regulatory requirements; from internal risk management, such as avoiding costly recapitalization, default, or liquidity shortfalls; or from market discipline, as wholesale funding spreads widen when capital and liquidity positions weaken.

First, the bank faces a risk-invariant capital constraint captured by $\xi \ln\big(K\big)$. Let $\kappa \in [0,1]$ denote the banking sectors' minimum desired capital ratio. Define the capital slack $K$ as

$$K \;=\; E - \kappa\,(L+R) \;=\; (1-\kappa)E \;-\; \kappa(D+W) \geq 0. \tag{20}$$

Thus, the capital slack $K$ is the excess of equity capital $E$ over a fixed fraction $\kappa$ of total assets, with loans and reserves weighted uniformly. This risk-invariant leverage ratio resembles the Supplementary Leverage Ratio (SLR) implemented in the Basel III regulatory framework, though $\kappa$ need not correspond to a regulatory minimum; it may equally capture an internally chosen target above any mandated level.[10] By rearranging with the balance sheet identity, we see that higher deposits $D$ and wholesale funding $W$ tighten the constraint, while greater equity $E$ relaxes it. As the balance sheet grows relative to equity, the capital slack $K$ shrinks, and the penalty $\xi \ln(K)$ reduces profits, making additional asset expansion increasingly costly.

Second, the bank faces a liquidity constraint captured by $\tau \ln\big(\Lambda\big)$. Let $\chi \in [0,1]$ denote the banking sectors' minimum desired ratio of liquid asset holdings relative to wholesale funding. Define the liquidity slack $\Lambda$ as,

$$\Lambda = R - \chi W \geq 0 \tag{21}$$

Thus, the liquidity slack $\Lambda$ is the excess of reserves $R$ over a fraction $\chi$ of wholesale fund-

[10] The model is qualitatively similar if we consider a risk-based capital requirement.

ing $W$. This formulation resembles the Liquidity Coverage Ratio (LCR) in Basel III, which requires banks to hold high-quality liquid assets against runnable short-term liabilities. More broadly, however, it captures banks' precautionary liquidity management against the withdrawal and rollover risk inherent in demandable and short-term funding, in the spirit of Diamond and Dybvig (1983). From (21), higher wholesale funding tightens the constraint, while greater reserves relax it. As wholesale funding rises without a matching increase in liquid assets, the liquidity slack $\Lambda$ shrinks, and the penalty $\tau \ln(\Lambda)$ reduces profits, making reliance on unstable funding increasingly costly.

Importantly, these frictions link deposit-taking to loan supply. Since wholesale funding is the residual funding margin, a deposit inflow reduces wholesale borrowing dollar-for-dollar for a given balance sheet. The liquidity penalty is what makes this relevant for lending: because wholesale funding requires liquid backing while deposits do not, substituting deposits for wholesale funding relaxes the liquidity constraint and lowers the marginal cost of lending. Without the liquidity penalty, marginal deposit inflows would be irrelevant for the marginal cost of lending and thus would have no effect on lending behavior. The capital penalty and market power in deposit and loan markets govern the strength of this link.[11]

### 4.4.2 First Order Conditions

Given the profit maximization problem above, we next derive the corresponding first order conditions. Throughout, starred quantities denote values at the optimum, in particular $K^\star$ and $\Lambda^\star$ are the capital and liquidity slack evaluated at the equilibrium balance sheet implied by $(r^{d\star}, r^{\ell\star})$.

We begin with the condition for the loan rate $r^{\ell\star}$,

$$r^{\ell\star} = r^S + \mu_\ell + \frac{1}{\alpha_\ell(1 - s_\ell(r^{\ell\star}))} + \chi \frac{\tau}{\Lambda^\star} + \kappa \frac{\xi}{K^\star}. \tag{22}$$

This expression shows that the banking sector's optimal loan rate $r^{\ell\star}$ can be written as the sum of five terms. The first is the policy rate, which provides the baseline cost of funds in the economy. The second is the per-unit operating cost of bank lending. The third is a demand

[11] Making these funding margins imperfect substitutes is a deliberate departure from Modigliani-Miller neutrality, and is well grounded for banks: equity is costly to raise, and deposits, which households hold for their safety and liquidity services, are more stable than wholesale funding. Because the penalties are reduced form representations of banks' revealed preferences for buffers, our analysis requires only that banks behave as if these margins are imperfect substitutes, not a stand on the deeper source of the friction.

markup, reflecting the bank's market power arising from borrowers' preference for bank loans: when loan demand is less elastic, the bank sets a higher rate above the policy rate. The fourth is a liquidity wedge, which arises because extending loans must be funded with additional wholesale liabilities and therefore tightens the liquidity slack. The fifth is a capital wedge, which reflects the marginal cost of balance sheet expansion: when capital slack is scarce and $K^\star$ is small, the marginal cost of expanding assets rises, increasing lending rates. The term $\tau/\Lambda^\star$ can be interpreted as the shadow cost of liquidity, while $\xi/K^\star$ can be interpreted as the shadow cost of capital.

Next, we solve for the optimal deposit rate $r^{d\star}$,

$$r^{d\star} \;=\; r^S \;-\; \mu_d \;-\; \frac{1}{\alpha_d\left(1 - s_d(r^{d\star} \mid r^S)\right)} \;+\; \chi\frac{\tau}{\Lambda^\star}. \tag{23}$$

The deposit first order condition shows that the banking sector's optimal deposit rate $r^{d\star}$ is determined by four terms. The first is the policy rate, which provides the baseline return on funds. The second is the per-unit operating cost of deposit provision. The third is a deposit markdown, reflecting the bank's market power arising from depositors' preference for bank deposits: when deposit demand is less elastic, the bank can offer a lower rate below the policy rate. The fourth is a liquidity wedge, which raises the deposit rate banks are willing to pay: deposits displace wholesale funding and thereby relax banks' liquidity slack.

Taken together, equations (22) and (23) reveal a direct link between deposit taking and credit provision. Deposits reduce reliance on wholesale funding and relax the liquidity requirement. The banking sector's willingness to extend credit is therefore tied to its ability to attract deposits, providing the mechanism through which the credit–deposit multiplier operates.

## 4.5 Non-bank Credit Intermediary

We model a non-bank credit intermediary that sets the non-bank credit rate $r^{nb}$ to maximize profits. It takes as given the policy rate $r^S$ and borrower demand $L_{nb}(r^{nb})$. The credit intermediary earns a spread over the policy rate and faces a per-unit cost $\mu_{nb}$ which can capture, for example, costs related to funding or monitoring. In the baseline specification this cost is constant; in Section 6 we consider an extension that allows non-bank funding costs to vary with conditions in

the short-term funding market. The non-bank's profit maximization problem is,

$$\max_{r^{nb}} \ \Pi_{nb} = (r^{nb} - r^S - \mu_{nb})\, L_{nb}(r^{nb}). \tag{24}$$

The optimal non-bank loan rate satisfies

$$r^{nb} = r^S + \mu_{nb} + \frac{1}{\alpha_\ell(1 - s_{nb})}. \tag{25}$$

This pricing condition is analogous to the bank loan pricing equation in Equation (22). The first term is the policy rate, which provides the baseline cost of funds in the economy. The second is the per-unit cost of non-bank lending. The third is a demand markup, reflecting the non-bank's market power arising from borrowers' preference for non-bank credit: when loan demand is less elastic, the non-bank sets a higher rate above the policy rate.

Because the non-bank bears no state-dependent balance-sheet costs, the bank to non-bank spread reveals the shadow price of intermediating through the banking sector's balance sheet: it widens exactly when the bank's capital and liquidity slack are scarce. To the extent that bank's wedges reflect regulation in excess of banks' internalized economic costs, this spread is the price of regulatory arbitrage.

The non-bank intermediary plays two roles in the model. First, it disciplines the elasticity of bank credit demand in equilibrium: because borrowers substitute between bank and non-bank credit, the non-bank's cost structure shapes the residual demand curve facing banks and thereby influences banks' marginal lending response $\lambda$ to deposit inflows. Second, it provides the margin along which we evaluate whether non-bank credit can offset the decline in the bank channel, as we explore in Section 6.

## 4.6 Money Markets and Institutional Market Funding

Market funding is supplied through two sectors. Households place market savings $M$ with money market funds, which allocate these balances between short-term claims that finance non-bank credit, $N$, and a residual safe asset $B_M$. The money market balance sheet is therefore

$$M = N + B_M. \tag{26}$$

Alongside money markets, a broader institutional sector supplies market funding $P$. It allocates these funds across bank wholesale claims $W$, non-bank funding $P_{nb}$, and a residual safe asset $B_P$. Non-bank credit is financed jointly by money market and institutional claims. The corresponding balance sheets are

$$P = W + P_{nb} + B_P, \tag{27}$$

$$L_{nb} = N + P_{nb}. \tag{28}$$

Both funding sectors are competitive, and all claims earn the policy rate $r^S$. The residual safe assets are supplied elastically at that rate, so the allocation of market funding is pinned down by the funding needs of banks and non-banks. Bank wholesale funding $W$ is determined residually by the bank balance sheet in (17), while non-bank lending $L_{nb}$ is determined by the intermediary's optimization in (24). The two funding sectors supply these claims and place any remaining balances in the residual safe assets.

Combining the sectoral balance sheets gives aggregate market clearing,

$$M + P = W + L_{nb} + B, \qquad B \equiv B_M + B_P. \tag{29}$$

This structure separates household market savings from the broader pool of institutional funding while allowing both sectors to finance non-bank credit. In the baseline, the funding allocation clears balance sheets without introducing an additional pricing wedge. Section 6 extends the model by allowing money-market depth to affect the marginal cost of non-bank intermediation.

## 4.7 Equilibrium

An equilibrium consists of rates $\{r^{d\star}, r^{\ell\star}, r^{\star}_{nb}\}$, quantities $\{D^{\star}, M^{\star}, L^{\star}, L^{\star}_{nb}, W^{\star}\}$, and funding allocations $\{P^{\star}, N^{\star}, P^{\star}_{nb}, B^{\star}_{M}, B^{\star}_{P}\}$, given the exogenous state $(r^S, E, R)$, such that:

1. Depositors maximize utility over bank deposits, market savings, and cash, yielding demands $D(r^{d\star})$, $M(r^{d\star})$, and $C(r^{d\star})$.

2. Borrowers maximize utility over bank loans, non-bank loans, and not borrowing, yielding loan demands $L(r^{\ell\star})$ and $L_{nb}(r^{\star}_{nb})$.

3. The banking sector chooses $(r^{d\star}, r^{\ell\star})$ to maximize its profits given by (19), subject to its balance sheet identity in (17) and the definitions of capital slack $K$ in (20) and liquidity slack $\Lambda$ in (21), which residually determines its wholesale funding demand $W^\star$.

4. The non-bank credit intermediary sets $r^{nb\star}$ to maximize its profits given by (24).

5. All markets clear. Banks and non-banks accommodate deposit and credit demand at their chosen rates, and the money-market and institutional funding balances satisfy (26)–(28).

## 4.8 Recovering the Credit-Deposit Multiplier

With the equilibrium in hand, we can now address the central question of the paper: how much credit does a deposit inflow ultimately create? We develop the answer in three steps. The first two introduce the banking sector's marginal lending response $\lambda$ and the recycling rate $\theta$; the third combines them into the bank-credit response to a deposit inflow, and extends it to the total credit response by incorporating the non-bank sector.

### 4.8.1 The Marginal Lending Response $\lambda$

Deposit inflows can arise from several distinct shocks: a depositor preference shock $\xi_D$ that shifts household allocation towards deposits, a reserve injection $dR$ that expands household financial wealth, or a loan demand shock $\xi_\ell$ that raises lending and generates deposit growth through the recycling of payment flows. Each shock produces a marginal lending response, $\lambda^0$, which depends on what else moves alongside the deposit inflow. The subsequent amplification, however, is shock-independent: at every round beyond the first, the bank faces the same inflow of recycled stable deposits regardless of which shock initiated the chain.

We therefore focus on the marginal lending response to recycled deposits, $\lambda \equiv dL/dD$: how much additional credit the banking sector extends per dollar of stable deposit inflow arriving on its own, with reserves and household wealth held fixed, taking into account the induced reduction in wholesale funding and the re-binding of the capital and liquidity slacks. Because it governs every recycling round, $\lambda$ is the object that determines amplification. Appendix A.2 derives $\lambda$, along with the shock-specific first-round responses $\lambda^0$, by differentiating the bank's balance sheet identity, the capital and liquidity slacks, and the bank and non-bank pricing con-

ditions around an interior equilibrium. For the recycled-deposit case, this yields,

$$\lambda = \frac{dL}{dD} = \frac{\dfrac{\tau\chi^2}{\Lambda^2}}{\dfrac{1}{\tilde{\kappa}_\ell^{nb}} + \dfrac{\tau\chi^2}{\Lambda^2} + \dfrac{\xi\kappa^2}{K^2}}. \tag{30}$$

Several economic forces shape $\lambda$. First, liquidity regulation plays a dual role. Stable deposits relax the liquidity constraint by reducing reliance on wholesale funding, a benefit captured by $\tau\chi^2/\Lambda^2$ in the numerator. At the same time, expanding loans tightens the liquidity constraint, a cost captured by the same term in the denominator, so the net liquidity effect depends on how binding the constraint is at the margin. Second, competition from non-banks shapes $\lambda$ through the coefficient $\tilde{\kappa}_\ell^{nb}$, which captures the residual loan-demand curvature facing banks after borrower substitution and non-bank repricing.[12] A larger $\tilde{\kappa}_\ell^{nb}$ means that banks face a more elastic residual loan-demand curve, so a relaxation of balance-sheet constraints translates into a larger lending response. Third, capital constraints discipline balance sheet expansion through $\xi\kappa^2/K^2$ in the denominator. When capital slack is limited, additional lending raises the shadow cost of leverage, reducing $\lambda$. Overall, $\lambda$ reflects the joint interaction of liquidity regulation, capital regulation, and credit-market competition in determining how strongly recycled deposit growth translates into bank lending.

### 4.8.2 The Deposit Recycling Rate $\theta$

The deposit recycling rate $\theta$ is the fraction of lending-generated payment flows that ultimately settles back into the banking system as stable deposits. When the bank extends $\Delta L$ in new loans, borrowers spend the proceeds, and the resulting payment flows accrue to the final recipients of loan-financed spending.[13] The treasurer sector allocates a fraction $\theta$ of these flows to stable bank deposits. The complementary share $1 - \theta$ flows to money market funds, yielding

$$dD = dD_0 + \theta\, dL, \tag{31}$$

$$dM = (1 - \theta)\, dL. \tag{32}$$

[12]The Appendix derives this term from the baseline non-bank pricing condition

[13]Loans are long-maturity claims. Principal is not repaid within the period in which production and payments occur, so repayment and refinancing decisions are abstracted from in the cash-flow accounting.

In the baseline, movements in $M$ do not enter any pricing condition, so equation (32) is a pure accounting record and the bank's response to each round's recycled deposits is $dL = \lambda\, dD$, with $\lambda$ as derived in Section 4.8.1. The extension in Section 6 relaxes this by letting non-bank costs depend on $M$.

Note that $\theta$ need not equal the deposit share $s^d$ that an average wealth holder would choose. The marginal recipients of lending-driven payment flows, such as firms, institutional investors, or non-bank intermediaries, hold proportionally more in short-term funding instruments than the average wealth holder, paralleling the distinction between marginal and average propensities to consume in fiscal policy.

Crucially, $\theta$ is not a margin the banking system optimizes over. From the perspective of any individual bank, it is an externality: when a bank extends a loan, the resulting payment flows circulate through the economy and may settle as stable deposits at other banks, or may not return to the banking system at all. The originating bank neither captures nor internalizes this margin. Because no individual bank optimizes over $\theta$, it does not enter the first-order conditions of the representative banking sector. Unlike $\lambda$, which emerges endogenously from the bank's profit maximization, $\theta$ is a primitive of the payment environment that we discipline using the reduced-form multiplier estimates.

### 4.8.3 The Bank-Credit and Total-Credit Responses

Equation (31) and the lending responses of Section 4.8.1 form a two-equation system in total deposits and total lending: deposits equal the initial inflow plus the recycled share of lending, while lending responds with the shock-specific $\lambda^0$ to the initial inflow and with $\lambda$ to recycled deposits at every subsequent round. Substituting one into the other gives the bank-credit response,

$$dL^{bank}_{total} = \lambda^0\, m_D\, dD_0, \qquad m_D = \frac{1}{1-\lambda\theta}, \tag{33}$$

where $m_D$ is the deposit multiplier. The first-round lending response $\lambda^0$ governs how much credit the initiating inflow creates, while $\lambda$ and $\theta$ jointly determine how strongly each round's lending sustains the next. For a pure recycled-deposit shock, $\lambda^0 = \lambda$.[14]

[14] In the baseline this also holds for a depositor preference shock: the accompanying outflow from market savings affects no pricing condition, so $\lambda_{\xi_D} = \lambda_D$. In the extension of Section 6, the two differ, as the outflow raises non-bank funding costs.

Total credit adds the induced response of non-bank credit, which in the baseline operates through borrower substitution alone: as bank lending expands and bank loan rates fall, borrowers shift away from non-bank credit. Let $\lambda_{nb,D}$ denote this non-bank response per dollar of recycled deposits. Equivalently, since each recycled dollar induces $\lambda$ of bank lending, the non-bank response per dollar of bank lending is $\lambda_{nb,D}/\lambda^D$. Because the non-bank responds only to the bank loan rate, this applies to every round of induced bank lending, so the cumulative non-bank response is $(\lambda_{nb,D}/\lambda^D)dL_{\text{bank}}$, and the total-credit response is

$$dL_{\text{total}} = \left[1 + \frac{\lambda_{nb,D}}{\lambda}\right] \frac{\lambda^0}{1-\theta\lambda} dD_0. \tag{34}$$

Because $\lambda_{nb,D} < 0$, total credit responds by less than bank credit: the deposit inflow reaches non-banks only through the fall in the bank loan rate, which crowds out non-bank lending. Section 6 introduces an additional channel by allowing non-bank costs to depend on short-term funding conditions, which are held fixed here.

One may ask whether movements in the residual safe claims $B_M$ and $B_P$ could affect the total-credit response. They could matter along two margins. First, trades in these assets could generate payment flows that settle as bank deposits; this margin is absorbed by measurement, since $\theta$ is disciplined by the observed propagation of deposit shocks and thus captures deposit creation through any channel. Second, changes in these claims could affect non-bank funding conditions; our assumption that non-bank funding is perfectly elastic shuts this channel, and the spillover exercise of Section 6 quantifies how large such effects would need to be to alter the aggregate conclusion.

# 5 Model Estimation

We estimate the model using aggregate U.S. banking data to recover the structural objects that govern the credit–deposit multiplier: the primitives underlying banks' marginal lending response $\lambda$, and the deposit recycling rate $\theta$. The estimation combines capital, liquidity, and demand parameters measured directly from balance sheet and pricing data with deterministic minimum-distance moment conditions. The banking sector's deposit and loan pricing conditions discipline the shadow costs of illiquidity and leverage, while the non-bank pricing condition disciplines the average marginal cost of non-bank credit. We jointly target these pricing relationships and the

reduced-form multiplier estimates from Section 3, which pin down the recycling rate of loan proceeds into stable deposits.

A useful feature of this approach is that the observed balance sheet quantities serve as inputs to the estimation rather than as targets of it. The model does not need to endogenously match the time-series levels of deposits, bank credit, liquid assets, money market fund assets, or non-bank credit. Instead, these balance sheet quantities are observed state variables that determine market shares, capital and liquidity slacks, and pricing wedges. Conditional on those observed states, the pricing moments discipline the shadow-cost parameters, while the reduced-form multiplier moments discipline deposit recycling. The structural model is therefore used to interpret marginal pricing and propagation around observed balance sheets, rather than to explain the full evolution of financial-sector quantities.

## 5.1 Data and Inputs to Estimation

We assemble an aggregate quarterly panel for the U.S. commercial banking sector combining Federal Reserve balance sheet quantities with FDIC Quarterly Banking Profile interest flows. All quantities are expressed relative to nominal GDP.[15] The sample spans 1999Q1–2024Q4. We measure core deposits, loans, wholesale funding, and equity using aggregate balance sheet data. Deposit and loan rates are constructed from interest flows and stocks. The policy rate is the effective federal funds rate. On the depositor side, money market funds and currency serve as liquid outside options. On the borrower side, nonbank credit is measured using corporate debt securities from the Financial Accounts. Details of variable construction are provided in Appendix A.3.1.

Two groups of parameters serve as inputs to the estimation, as reported in Table 3. First, the balance-sheet-implied capital and liquidity ratios $\kappa$ and $\chi$ define the pre- and post-reform regimes. Rather than imposing statutory values, we infer the banking sector's desired capital and liquidity ratios from observed balance sheet positions. In each regime $j \in \{pre, post\}$, let $b = 0.01$ denote a one-percentage-point ratio buffer. We set the capital ratio to the minimum

[15] Scaling by nominal GDP renders quantities stationary and comparable across the sample. It also embed an assumption: $\tau$ and $\xi$ are constant in GDP units, so a given slack provides the same comfort whenever it is the same share of the economy. This is natural, since the shocks that buffers absorb scale with the banking system.

observed capital ratio less this buffer,[16]

$$\kappa_j = \min_{t \in \mathcal{T}_j} \frac{E_t}{L_t + R_t} - b. \tag{35}$$

We set the liquidity ratio analogously, using the minimum observed reserve-to-wholesale-funding ratio less the same one-percentage-point buffer,

$$\chi_j = \min_{t \in \mathcal{T}_j} \frac{R_t}{W_t} - b. \tag{36}$$

This procedure pins the ratios to the tightest observed balance-sheet positions in each regime while allowing a small positive buffer at the binding observation. We plot the full quarterly series of the quantities $K_t$ and $\Lambda_t$ in Appendix Figure A.3.

Because the ratios are inferred from revealed balance-sheet behavior rather than imposed from statute, the regime shift in $\kappa$ and $\chi$ absorbs not only regulatory reform but any structural change in the banking sector's desired buffers over the period. For example, consolidation that improves the average bank's access to wholesale funding would lower the sector's desired liquidity ratio and be reflected in $\chi_{post}$. The estimated regimes should therefore be read as summarizing the banking sector's effective balance-sheet preferences in each era, whatever their source.

Second, the deposit and loan demand elasticities determine the demand markups that enter the pricing first-order conditions. These elasticities govern the responsiveness of deposits and loans to interest rates and therefore shape equilibrium funding pass-through and lending competition. We take these elasticities from the literature (Xiao, 2019; Diamond et al., 2024).[17]

## 5.2 Minimum-Distance Estimation and Parameter Estimates

Given these inputs, we estimate the parameter vector

$$\Theta = \{\, \tau,\ \xi,\ \theta_{pre},\ \theta_{post},\ \mu_{nb} \,\}, \tag{37}$$

[16]This follows the spirit of Begenau (2020), who disciplines regulatory requirements using observed bank balance-sheet ratios. The one-percentage-point buffer prevents the tightest observed balance sheet in a regime from having exactly zero slack; because it is expressed in ratio points rather than as a fixed level, the implied slack at the binding observation scales with the balance sheet, $b(L_t + R_t)$ for capital and $bW_t$ for liquidity.

[17]The exact values of these elasticities do not qualitatively affect our results.

**Table 3**: Model Inputs

Panel A lists the balance-sheet-implied capital and liquidity parameters for the pre- and post-reform regimes. Panel B lists deposit and loan rate sensitivites.

A. Balance Sheet-Implied Parameters

| Symbol | Description | Pre-Reform | Post-Reform |
|---|---|---|---|
| $\kappa$ | Capital requirement | 0.027 | 0.083 |
| $\chi$ | Liquidity ratio | 0.368 | 0.703 |

B. Price Elasticities

| Symbol | Description | Value |
|---|---|---|
| $\alpha_d$ | Depositor rate sensitivity | 1.270 |
| $\alpha_\ell$ | Borrower rate sensitivity | 5.560 |

by deterministic minimum distance, choosing parameters so that the model jointly matches the pricing first-order conditions and the reduced-form credit-deposit multipliers in the early and late periods. This follows the logic of structural banking estimation in which demand elasticities are disciplined outside the supply-side estimation, and supply-side primitives are recovered from equilibrium pricing and quantity moments (Wang et al., 2022); In our aggregate setting, identification comes from time-series variation in observed balance-sheet slacks, ratios, and spreads, together with the external multiplier moments, rather than from cross-sectional bank heterogeneity.

The shadow-cost parameters $(\tau, \xi)$ and the non-bank cost $\mu_{nb}$ are common across regimes; the recycling rates $\theta_{pre}$ and $\theta_{post}$ are regime-specific. Time variation in the liquidity and capital wedges, $\chi_t\tau/\Lambda_t$ and $\kappa_t\xi/K_t$, therefore comes only from measured sources: the slacks $\Lambda_t$ and $K_t$, which evolve quarterly with bank balance sheets, and the ratios $\chi_t$ and $\kappa_t$, which vary in the pre- and post-period as inferred from balance sheet positions. Because the same $(\tau, \xi)$ must price both regimes, the regime-split pricing moments act as overidentifying restrictions: the estimated tightening of the wedges is identified by measured movements in slacks and ratios rather than by freely shifting parameters.

The estimation targets two sets of moments, summarized in Table 4: the banking and non-bank pricing conditions, and the reduced-form multipliers. The banking sector's loan and deposit pricing first-order conditions discipline the shadow costs of liquidity $\tau$ and leverage $\xi$.

**Table 4**: Targeted Moments and Parameter Discipline

The table summarizes the variation that disciplines each estimated object.

| Object | Moment or variation | Economic role |
|---|---|---|
| $\tau, \xi$ | Bank deposit and loan pricing residuals, together with time-series variation in $\Lambda_t$, $K_t$, $\chi_t$, and $\kappa_t$ | Liquidity and leverage shadow-cost scales |
| $\theta_{pre}, \theta_{post}$ | Early and late reduced-form credit–deposit multiplier targets | Recycling of loan proceeds into stable deposits |
| $\mu_{nb}$ | Average residual non-bank pricing wedge after the demand markup | Non-bank marginal-cost component |

The deposit pricing residual follows from the model's first order condition (23) directly,

$$\varepsilon_t^d(\Theta) = (r_t^d - r_t^S) + \mu_d + \frac{1}{\alpha_d(1 - s_t^d)} - \chi_t \frac{\tau}{\Lambda_t}. \tag{38}$$

The loan pricing residual adjusts the model's condition (22) for two features of the data. First, measured loan yields are contractual rates on risky credit, so we net out expected losses using the loan-loss provision rate $p_t^\ell$; this measures the time-varying component of the per-unit lending cost $\mu_\ell$ directly so that its cyclicality does not load onto the wedges. Second, measured bank credit $L_t$ comprises loans and securities; we assume that securities are competitively priced and carry no default risk, thus both the markup and the expected-loss adjustment apply only to the loan component and are scaled by the loan share $\omega_t^\ell$. The loan pricing residual is thus,

$$\varepsilon_t^\ell(\Theta) = (r_t^\ell - r_t^S) - \omega_t^\ell p_t^\ell - \mu_\ell - \frac{\omega_t^\ell}{\alpha_\ell(1 - s_t^L)} - \chi_t \frac{\tau}{\Lambda_t} - \kappa_t \frac{\xi}{K_t}. \tag{39}$$

The baseline estimation then imposes $\mu_d = \mu_\ell = 0$ for the remaining operating-cost components, so that the pricing moments discipline $\tau$ and $\xi$ directly. Any constant cost level absorbed by the shadow-cost parameters affects both regimes symmetrically and largely differences out of the cross-regime comparison that is our focus.

The baseline non-bank pricing condition disciplines $\mu_{nb}$, the average per-unit cost of non-bank credit,

$$\varepsilon_t^{nb}(\Theta) = (r_t^{nb} - r_t^S) - \mu_{nb} - \frac{1}{\alpha_\ell(1 - s_t^{nb})}. \tag{40}$$

The identification is the mirror image of the bank side. Because the non-bank bears no balance-

sheet wedges, its spread net of the demand markup identifies the cost level $\mu_{nb}$ directly. We later extend the model to consider how $\mu_{nb}$ may be affected by funding market conditions in Section 6.

The reduced-form multiplier estimates from Section 3 discipline the deposit recycling rates $\theta_{pre}$ and $\theta_{post}$. Both moments are instances of the deposit multiplier, with the first-round response $\lambda^0$ set to match each period's Treasury settlement institutions. In the pre-period, settlements move deposits without reserves, so $\lambda^0 = \lambda_t$ and the multiplier moment is,

$$g^m_{pre}(\Theta) = \hat{m}^{pre} - \frac{1}{T_{pre}} \sum_{t \in \mathcal{T}pre} \frac{1}{1 - \theta pre\, \lambda_t(\Theta)}. \tag{41}$$

In the post-period, after the Treasury General Account became a Federal Reserve account, settlements increase deposits and reserves together. The first round therefore propagates with the reserve-backed response $\lambda^R_t$, while the subsequent rounds of lending-generated deposits propagate with the recycled-deposit response $\lambda_t$,

$$g^m_{post}(\Theta) = \hat{m}^{post} - \frac{1}{T_{post}} \sum_{t \in \mathcal{T}post} \left(1 + \frac{\theta post, \lambda^R_t(\Theta)}{1 - \theta_{post}, \lambda_t(\Theta)}\right). \tag{42}$$

The minimum-distance objective stacks eight moment gaps: the pre- and post-period averages of the deposit, bank-loan, and non-bank pricing residuals, and the pre- and post-period multiplier gaps,

$$g(\Theta) = \left(\bar{\varepsilon}^d_{pre}, \bar{\varepsilon}^d_{post}, \bar{\varepsilon}^\ell_{pre}, \bar{\varepsilon}^\ell_{post}, \bar{\varepsilon}^{nb}_{pre}, \bar{\varepsilon}^{nb}_{post}, g^m_{pre}, g^m_{post}\right)'. \tag{43}$$

The estimator solves

$$\hat{\Theta} = \arg\min_{\Theta} g(\Theta)' W g(\Theta), \tag{44}$$

where $W$ is diagonal and held fixed during optimization: for the pricing moments, inverse Newey-West long-run variances of preliminary regime-average residuals with four quarterly lags; for the multiplier moments, the reduced-form IV variances from Section 3. Standard errors use the minimum-distance sandwich covariance matrix, with the covariance between the pricing block and the externally estimated multiplier block set to zero. Table 5 compares each empirical

**Table 5**: Targeted Moment Fit

The table reports empirical targeted moments and their model-implied counterparts. For pricing moments, the t-statistic is the model-implied moment minus the empirical moment divided by the Newey–West standard error of the corresponding regime-average pricing residual, using four quarterly lags. For multiplier moments, the t-statistic uses the reduced-form IV standard error.

| Moment | Actual | Model | t-stat |
|---|---|---|---|
| Pricing moments (annual percentage points) | | | |
| Deposit spread, pre | 0.886 | 0.115 | -0.864 |
| Deposit spread, post | 0.794 | 1.110 | 0.656 |
| Bank-loan spread, pre | 2.456 | 2.459 | 0.004 |
| Bank-loan spread, post | 3.025 | 3.025 | -0.001 |
| Non-bank spread, pre | 3.155 | 2.764 | -0.733 |
| Non-bank spread, post | 2.573 | 2.807 | 0.568 |
| Reduced-form multiplier moments | | | |
| $m^D_{pre}$ | 2.38 | 2.38 | -0.000 |
| $m^R_{post}$ | 1.36 | 1.36 | -0.000 |

target with its model-implied counterpart, with t-statistics scaling each model-data gap by the corresponding moment standard error.

## 5.3 The Total Bank-Credit Effect of Deposit Flows

The estimated model decomposes the credit–deposit multiplier into its two constituent forces: banks' marginal lending response $\lambda$ and the deposit recycling rate $\theta$. With both in hand, we can now answer the central question of the paper: how much credit does a deposit inflow ultimately create, and how has this changed over the past two decades?

Table 6 reports the model-implied values for $\lambda$ and $\theta$. Banks' marginal lending response $\lambda$ has declined from 0.88 to 0.58. This represents a decline of over 30% in the share of each marginal deposit dollar that is channeled into new lending. The remainder is absorbed by reductions in wholesale funding. Deposits have moved from being primarily a source of credit creation to also substantially being a tool of balance sheet management.

The deposit recycling rate $\theta$ falls from 0.65 to 0.47. This also represents a decline of almost 30% in the share of loan-generated payment flows that ultimately settle as stable deposits within the banking system. This decline is consistent with two complementary forces.

**Table 6**: Model Estimates

Panel A reports the main estimated parameters and model-implied recycled-deposit lending responses. Standard errors are reported in parentheses. Panel B reports the matched credit–deposit multipliers and the model-implied effects of an initial recycled-deposit inflow on bank credit and total credit. Total credit includes the induced response of non-bank credit. The post-period multiplier target is reserve-backed.

A. Estimated Parameters

| | Estimate | SE |
|---|---|---|
| Shadow costs | | |
| $\tau$ Illiquidity | 0.073 | (0.023) |
| $\xi$ Leverage | 0.336 | (0.086) |
| Deposit recycling | | |
| $\theta_{pre}$ | 0.647 | (0.112) |
| $\theta_{post}$ | 0.465 | (0.401) |
| Recycled-deposit lending response | | |
| $\lambda^{rec}_{pre}$ | 0.882 | (0.032) |
| $\lambda^{rec}_{post}$ | 0.583 | (0.073) |

B. Model-Implied Objects

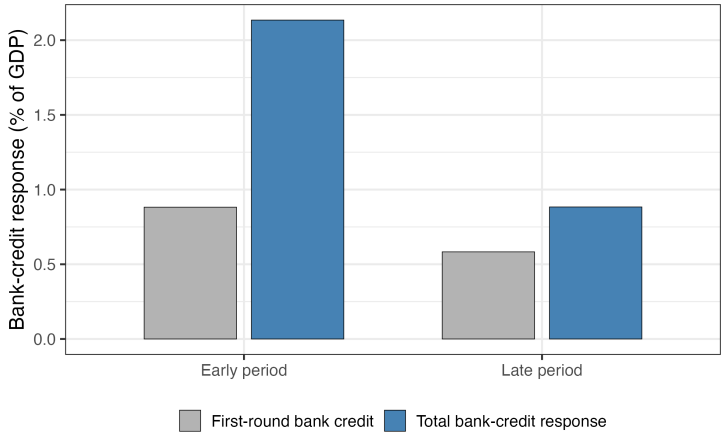
| | Model | Data |
|---|---|---|
| Credit-Deposit Multiplier | | |
| $m^{D}_{pre}$ | 2.38 | 2.38 |
| $m^{R}_{post}$ | 1.36 | 1.36 |
| Effect on Bank Credit | | |
| Pre period | 2.135 | |
| Post period | 0.884 | |
| Effect on Total Credit | | |
| Pre period | 1.140 | |
| Post period | 0.497 | |

First, bank deposits have become less special as a savings instrument: households and firms increasingly allocate funds toward non-bank alternatives such as money market funds. Second, the composition of bank assets has shifted from loans toward securities, and the two generate different payment flows: loan proceeds are spent and resettle broadly across households and firms, whereas security purchases route funds to sellers—often institutional investors—whose marginal savings vehicle is less likely to be a stable bank deposit.

As a result, the total bank-credit response to a recycled deposit inflow, $\Delta L^{bank}_{total} = \lambda/(1 - \theta\lambda)$ per dollar of initial deposits, has declined from $2.13$ to $0.88$. Each dollar of deposits today generates less than half of the bank credit it once did. This decline is driven both by the weaker lending response and diminished deposit recycling, which weakens the amplification from one round to the next. Figure 4 illustrates this by comparing the effect from banks' initial lending response to the total bank-credit effect following a one percent of GDP deposit inflow in the early and late periods. In the earlier period, the multiplier more than doubles the initial credit response. By the later period, amplification is much smaller. Taking the multiplier into account is crucial to understanding the magnitude of the decline in credit creation in the late period relative to the early period. Comparing only banks' marginal lending responses suggests a roughly one-third decline. But the total bank-credit effect has fallen by nearly 60 percent. The difference

**Figure 4**: First-Round and Total Bank-Credit Effects

The figure compares the first-round bank-credit effect and the total bank-credit effect of a one percent of GDP deposit inflow in the early and late periods. The first-round effect corresponds to banks' initial lending response, $\lambda^D \Delta D_0$, abstracting from subsequent recycling. The total effect incorporates full amplification through endogenous deposit re-depositing and multiplier dynamics, $\lambda^D \Delta D_0/(1-\theta\lambda^D)$.




reflects the compounding nature of the multiplier: when $\theta$ weakens, the amplification mechanism contracts disproportionately.

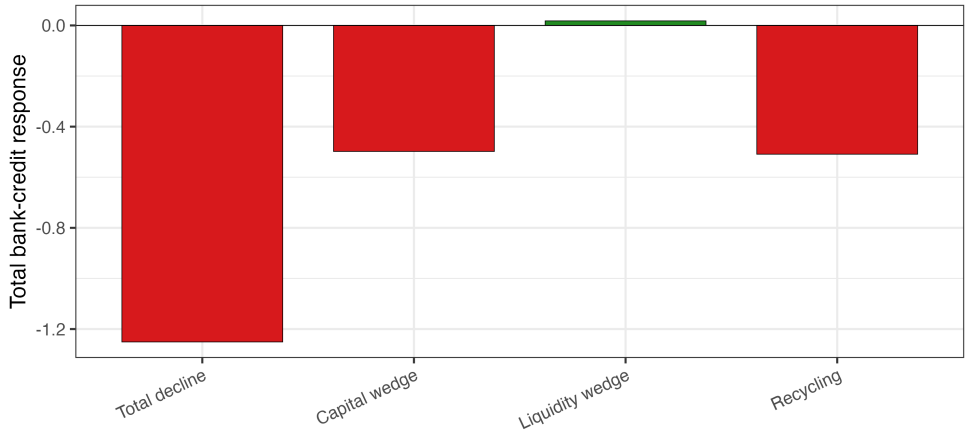
# 6 Decomposing the Decline in Credit Creation

## 6.1 Drivers of the Decline in Bank-Credit Creation

We next use the model to study the drivers of the decline in the credit impact of deposit flows. We decompose the decline into two channels. The first varies the structural primitives governing bank behavior, including costs of capital and liquidity, and the extent of deposit recycling. The second varies balance-sheet quantities, asking how Quantitative Easing and Tightening affects the banking sector's capacity to intermediate deposits into credit. Together, the two exercises reveal that the decline reflects both an increase in banks' balance sheet costs and a decline in the specialness of deposits, as well as a transformation of the balance sheets themselves.

**Figure 5**: Drivers of the Decline in Bank-Credit Creation

The figure reports the early-to-late change in the total bank-credit response, $\lambda^D/(1-\theta\lambda^D)$. The first bar reports the total change. The remaining bars hold the late-period balance sheet fixed and successively move the capital wedge, the liquidity wedge, and deposit recycling to their pre-period historical average values, each bar building on the previous one.




### 6.1.1 Changing Primitives: Capital, Liquidity, and Deposit Recycling

We decompose the early-to-late decline in the total bank-credit response, $\Delta L^{bank}_{total} = \lambda/(1-\theta\lambda)$, using a sequence of regime-average counterfactuals in Figure 5. Starting from the late-period economy, we successively return each driver to its pre-period average value, each step building on the previous one: first the capital wedge $\xi\kappa^2/K^2$, then the liquidity wedge $\tau\chi^2/\Lambda^2$, then the recycling rate $\theta$.[18] The late-period balance sheet is held fixed throughout.

Because the steps are layered rather than added one at a time, the three bars in Figure 5 are additive: they sum to the effect of returning all three primitives to pre-period values. The gap between this combined effect and the total decline is a residual reflecting the change in the balance sheet's own composition and scale, which we explore in the next subsection. Figure 5 shows that the capital wedge and reduced recycling are the two first-order forces: each lowers the total bank-credit response by about half a dollar per deposit dollar. The liquidity wedge is small and works in the opposite direction, slightly offsetting the decline by increasing the liquidity benefit of the marginal deposit dollar.

[18]Each late-period quarter's wedge is rescaled proportionally so that its regime average matches the pre-period average, preserving the late period's quarter-to-quarter variation. We focus on the wedge terms rather than the scale parameters $\tau$ and $\xi$ which are not interpretable in isolation: the economically relevant objects are the wedges, which combine the estimates with the observed ratios and slacks.

### 6.1.2 Changing Quantities: Quantitative Easing and Tightening

We next turn from primitives to quantities. Two major post-crisis policies transformed the size and composition of bank balance sheets: Quantitative Easing (QE), which flooded the banking system with reserves, and Quantitative Tightening (QT), which drained reserves from the banking system. The exercises in this section ask how these balance-sheet changes alter the marginal bank-credit response to a new deposit dollar. They should therefore be distinguished from estimates of the direct effect of QE on the level of bank lending, such as Diamond et al. (2024). In the model, QE and QT operate on the same balance-sheet frictions in opposite directions. QE adds reserves and deposits together, relaxing the liquidity position but expanding the balance sheet against fixed equity, which tightens the capital position and lowers the marginal value of a deposit. QT can instead tighten the liquidity constraint by removing reserves without a commensurate decline in deposit claims, raising the marginal value of stable funding.

**Quantitative Easing.** We ask how QE changed the transmission of deposit inflows into bank credit. To answer this, we construct a counterfactual banking system in which QE-created reserves and the associated funding are removed, and compare the marginal bank-credit response to a dollar inflow across the observed and counterfactual balance sheets. Because balance-sheet policy injects money in reserve-backed form, we evaluate the response to such an inflow, with first-round response $\lambda^R$.

Constructing the counterfactual requires taking a stand on how QE-created reserves are funded. When the Federal Reserve purchases Treasury securities and MBS from non-bank counterparties, it creates reserves on bank balance sheets, and the sellers' bond holdings are replaced by claims on the banking system.[19] These QE-created balances land disproportionately with institutional holders, asset managers, pension funds, and dealers' clients, whose bank claims are rate-sensitive and functionally closer to wholesale funding than to the stable core deposits that relax banks' liquidity management (Blickle et al., 2025). In the model, the parameter $\zeta_{QE}$ governs this composition. In the baseline, we set $\zeta_{QE} = 0.5$, which means that half of the created wealth is held as core deposits, while the remainder appears on bank balance sheets as wholesale funding. QE therefore supplies reserves while diluting the funding mix toward flightier claims.

[19] Household bond holdings lie outside the depositor menu by design: the menu $\{D, M, C\}$ models the allocation of liquid wealth, taking the division of portfolios between liquid and longer-duration assets as given. QE moves wealth across this boundary, converting bonds into money-like claims, and thus enters as an expansion of $W^d$.

The counterfactual removes QE-created reserves and the associated deposit and wholesale funding balances, holding bank and non-bank credit fixed, as detailed in Appendix A.4.2.

Figure 6 reports the result, which nets opposing balance-sheet margins. On the liquidity side, QE's reserves relax the liquidity position, though the wholesale funding QE arrives with demands liquid backing of its own, absorbing part of the relief. On the capital side, the balance-sheet expansion against fixed equity tightens the capital position. In our calibration the capital effect dominates: QE reduces the bank-credit response to a reserve-backed marginal dollar inflow by \$1.56.

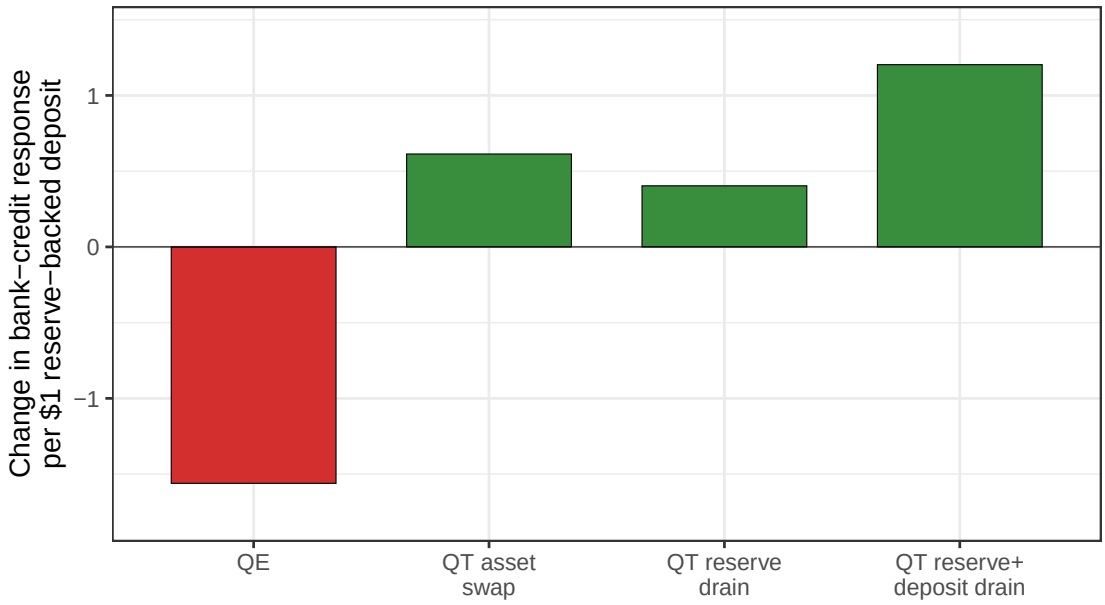
**Quantitative Tightening.** We next ask how QT changes the transmission of deposit inflows into bank credit. Beginning in mid-2022, the Federal Reserve allowed its securities holdings to run off; as Treasury and MBS holdings mature without reinvestment, the Treasury reissues the debt to the market. Acharya and Rajan (2024) emphasize that this shrinkage can reduce bank-held reserves without a commensurate reduction in deposit claims. This observation disciplines our exercises: the relevant question is how a smaller stock of liquid assets, relative to deposit claims, changes the marginal value of new stable deposit funding. We emphasize that these exercises do not measure the effect of QT on the level of bank lending; they ask whether, after the QT balance-sheet adjustment, an additional deposit dollar has greater multiplier power.

Because how QT's runoff lands on bank balance sheets is uncertain, we report three short-run counterfactuals, as detailed in Appendix A.4.2. The first treats QT as an asset swap: banks absorb one half of the runoff by exchanging reserves for less liquid securities, so total assets, deposits, wholesale funding, and credit are unchanged and only the liquidity buffer falls; isolating a pure liquidity tightening. The second is a reserve drain: bank credit and deposits are held fixed, reserves fall, and wholesale funding adjusts residually; in this case the reserve reduction deteriorates liquidity while the diminished reliance on wholesale funding reduces the liquidity burden of funding. The third is a reserve-and-deposit drain: reserves and deposits fall one-for-one with bank credit fixed. The second and third counterfactuals also shrink capital-relevant assets, so they combine the liquidity tightening with a loosening of the capital position.

Figure 6 reports the results. All three variants raise the marginal bank credit response: the asset swap by $0.61$, the short-run reserve drain by $0.40$, and the reserve-and-deposit drain by $1.20$. The ordering follows the mechanics. The asset swap is a pure liquidity tightening on the asset side. The reserve drain adds capital relief, but its liquidity tightening is partially

**Figure 6**: Effects of QE and QT

The figure reports the change in the bank-credit response per dollar of reserve-backed deposit inflow, $\lambda^0/(1-\theta\lambda^D)$, induced by balance-sheet policy in the post-period banking sector. The QE bar compares the observed post-period balance sheet to a no-QE counterfactual. The QT bars use the decline in Federal Reserve Treasury and MBS holdings from 2022Q2 to 2024Q4, applying one half of the runoff to bank balance sheets. The asset-swap bar holds total bank assets, credit, deposits, and wholesale bank funding fixed but reduces liquid assets. The reserve-drain bar holds bank credit and deposits fixed while liquid assets other than bank credit fall and wholesale bank funding adjusts residually. The reserve-and-deposit-drain bar lets reserves and deposits fall one-for-one while holding bank credit fixed.




offset by the accompanying decline in wholesale funding, which reduces the liquid backing the banking requires; on net it delivers the smallest effect. The reserve-and-deposit drain combines the full liquidity tightening, since wholesale funding is unchanged and no offset arises, with the capital relief, and therefore delivers the largest effect. Together with the QE result, the exercises show balance-sheet policy moving the same frictions in opposite directions: QE dampens the transmission of new deposits into credit, while QT amplifies it.

These exercises highlight that QE and QT operate through the same balance-sheet frictions in opposite directions. This carries a direct implication for the deposits channel of monetary policy in Drechsler et al. (2017): policy actions that operate through the supply or attractiveness of bank deposits have smaller credit effects in a QE-expanded banking system, where an additional deposit dollar commands less multiplier power. More broadly, the deposit multiplier is not a fixed structural constant but a policy-dependent object, lowered by balance-sheet expansion and partially restored by runoff.

## 6.2 Implications for Aggregate Credit

The total effect of deposit inflows on bank credit has declined sharply, from $2.13$ to $0.88$. A natural question is whether non-bank credit intermediation, through bond markets, private credit, and other market-based channels, can compensate for this weakened deposit channel. We approach this question in two steps. We first examine the baseline equilibrium response of bank, non-bank, and aggregate credit to a deposit shock when non-bank marginal costs are unrelated to funding market conditions. We then ask how large the spillover from funds flowing into short-term funding markets to non-bank intermediation costs would need to be to restore the aggregate credit response to its early-period level.

### 6.2.1 Non-Bank Competitive Response

We begin with the direct margin: the equilibrium credit response to a deposit inflow when non-bank costs do not depend on funding market conditions. Figure 7 reports the baseline responses of bank, non-bank, and total credit to a deposit inflow of one percent of GDP in each period. Bank credit responses far less in the late period, and the total credit response falls by more than half. Non-bank credit contracts in both periods, the deposit inflow reaches non-banks only through the fall in bank loan rates as banks expand, which pulls borrowers away from non-bank credit.

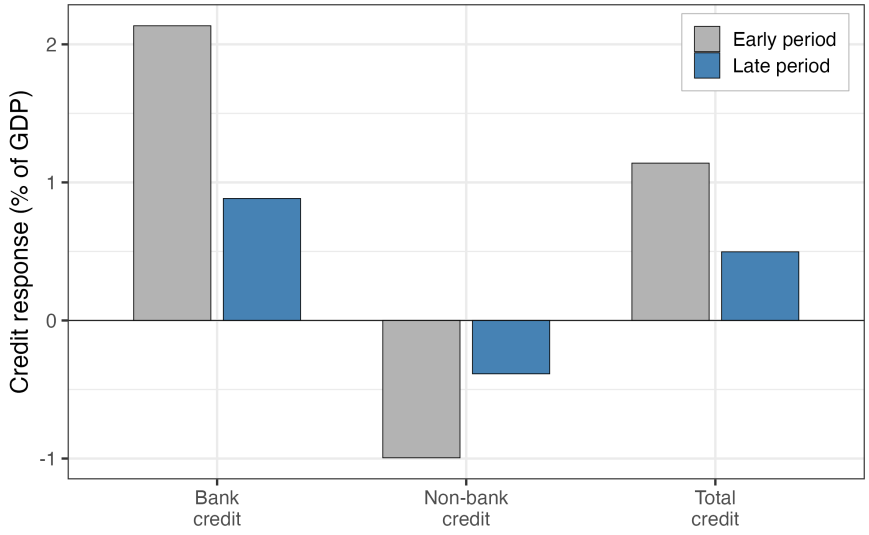
This contraction clarifies why the competitive channel cannot close the gap between the early and late periods. Our motivating question is whether a more developed non-bank sector can deliver the credit creation that the bank channel no longer provides. Through borrower substitution alone, the answer is no, and the reason is one of sign: non-banks respond to a deposit inflow by contracting, so a more elastic or more substitutable non-bank sector deepens the contraction and widens the gap rather than narrowing it. For non-banks to compensate for the lost amplification, they must operate through a different margin, one that lowers their cost of intermediation as funds flow into short-term funding markets. We turn to that margin next.

### 6.2.2 How Large Must the Funding-Market Spillover Be?

We now ask how strong the spillover margin would need to be to matter. To do so, we extend the baseline model to let non-bank funding conditions depend specifically on money-market depth

**Figure 7**: Baseline Total Effect on Aggregate Credit

The figure reports the baseline equilibrium credit response to a one percent of GDP deposit inflow in the early and late periods. Non-bank marginal costs are held fixed with respect to money-market depth, so the non-bank response reflects borrower substitution only. Bars show levels, not late-minus-early differences.




$M$. Money market funds use these balances to hold short-term non-bank funding claims $N$ and safe assets $B_M$, so a deeper pool of money-market balances may lower the marginal cost of non-bank intermediation. We capture this reduced-form effect with a single loading, $\kappa_M$:

$$\mu_{nb}(M) = \bar{\mu}_{nb} - \kappa_M M. \tag{45}$$

Here $\kappa_M$ measures the sensitivity of non-bank marginal costs to money-market depth. When lending-generated payment flows do not recycle into stable deposits, let $\omega_M$ denote the share that enters money market funds. Evaluated at the no-spillover total-credit response, the resulting inflow is

$$dM^{sel} = \omega_M(1-\theta)dL^0_{\text{total}}, \tag{46}$$

where the superscript $0$ denotes the baseline with $\kappa_M = 0$. Appendix A.4.3 derives the total-credit response with this leakage channel, combining the bank response, borrower substitution, and the effect of money-market depth on non-bank funding costs.

We use the extension to ask what cost sensitivity would restore the late-period aggregate credit response to its early-period level. The baseline total-credit response falls from $1.14$ in

the early period to $0.50$ in the late period, leaving a gap of $0.64$ per dollar of initial deposits. Under full leakage, $\omega_M = 1$, closing this gap requires $\kappa_M = 2.21$. Thus, a one-percent-of-GDP increase in money-market depth must lower non-bank marginal intermediation costs by $2.21$ basis points. With half leakage, $\omega_M = 0.5$, the required loading doubles to $4.42$. More generally, the required cost sensitivity scales inversely with the share of non-recycled proceeds that reaches money markets.

The full-leakage case is the most favorable configuration for the non-bank funding channel and therefore provides the smallest required value of $\kappa_M$. Even in this case, borrower substitution and improved funding conditions do not replace the lost bank amplification unless money-market inflows pass through materially to non-bank intermediation costs. More broadly, because non-bank intermediation chains are longer, neither recycled non-bank proceeds nor leaked bank flows are likely to compensate for the loss in bank recycling. On the direct margin, non-banks contract rather than substitute; on the funding margin, they do not replace the multiplier.

# 7 Conclusion

The literature on banking has long debated whether deposits fund loans or loans create deposits, but has paid relatively little attention to the feedback loop between the two. This paper shows that this feedback loop is central to understanding the declining role of deposits in credit provision. A dollar of deposits today generates less than half of the credit it did two decades ago, and accounting for the amplification channel is crucial: focusing on banks' marginal lending response alone understates the decline in total credit creation. The diminished specialness of bank deposits means that fewer credit proceeds return as stable funding to support further credit creation. As recycling weakens, the compounding nature of the multiplier causes the amplification mechanism to contract.

Our findings point to an implicit narrowing of banking, driven not by institutional separation but by regulation and non-bank competition. As a result, both money and credit have become less responsive to deposit inflows, dampening the amplification of the credit cycle. These changes have implications for the transmission of monetary and fiscal policy, the cyclicality of credit, and the consequences of further migration of funds from banks to non-bank alternatives such as money market funds and stablecoins. Concerns about such migration may have more

muted credit effects today than they would have had in earlier periods, but the weakening of the bank channel itself represents a meaningful and difficult-to-replace loss of amplification in the financial system.

# References


**Acharya, Viral V. and Raghuram Rajan**, "Liquidity, Liquidity Everywhere, Not a Drop to Use: Why Flooding Banks with Central Bank Reserves May Not Expand Liquidity," *The Journal of Finance*, 2024, *79* (5), 2943–2991.

**Begenau, Juliane**, "Capital requirements, risk choice, and liquidity provision in a business-cycle model," *Journal of Financial Economics*, 2020, *136* (2), 355–378.

_ **and Tim Landvoigt**, "Financial regulation in a quantitative model of the modern banking system," *The Review of Economic Studies*, 2022, *89* (4), 1748–1784.

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

**Bernanke, Ben S, Mark Gertler, and Simon Gilchrist**, "The financial accelerator in a quantitative business cycle framework," *Handbook of macroeconomics*, 1999, *1*, 1341–1393.

**Blickle, Kristian, Jian Li, Xu Lu, and Yiming Ma**, "The dynamics of deposit flightiness and its impact on financial stability," Technical Report, National Bureau of Economic Research 2025.

**Brunnermeier, Markus K and Yuliy Sannikov**, "A macroeconomic model with a financial sector," *American Economic Review*, 2014, *104* (2), 379–421.

**Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru**, "The secular decline of bank balance sheet lending," Technical Report, National Bureau of Economic Research 2024.

**Corbae, Dean and Pablo D'Erasmo**, "Capital buffers in a quantitative model of banking industry dynamics," *Econometrica*, 2021, *89* (6), 2975–3023.

**DeMarzo, Peter M, Arvind Krishnamurthy, and Stefan Nagel**, "Interest rate risk in banking," Technical Report, National Bureau of Economic Research 2024.

**Denbee, Edward, Christian Julliard, Ye Li, and Kathy Yuan**, "Network risk and key players: A structural analysis of interbank liquidity," *Journal of Financial Economics*, 2021, *141* (3), 831–859.

**Diamond, Douglas W and Philip H Dybvig**, "Bank runs, deposit insurance, and liquidity," *Journal of political economy*, 1983, *91* (3), 401–419.

**Diamond, William, Zhengyang Jiang, and Yiming Ma**, "The reserve supply channel of unconventional monetary policy," *Journal of Financial Economics*, 2024, *159* (C).

**Donaldson, Jason Roderick, Giorgia Piacentino, and Anjan Thakor**, "Warehouse banking," *Journal of Financial Economics*, 2018, *129* (2), 250–267.

**Drechsler, Itamar, Alexi Savov, and Philipp Schnabl**, "The Deposits Channel of Monetary Policy*," *The Quarterly Journal of Economics*, 05 2017, *132* (4), 1819–1876.

**_ , _ , and _** , "Banking on deposits: Maturity transformation without interest rate risk," *The Journal of Finance*, 2021, *76* (3), 1091–1143.

**Egan, Mark, Ali Hortaçsu, Nathan Kaplan, Aditya Sunderam, and Vincent Yao**, "Dynamic Competition for Sleepy Deposits," *NBER Working Paper*, 2025, (w34267).

**Friedman, Milton and Anna J. Schwartz**, "Money and Business Cycles," *The Review of Economics and Statistics*, 1963, *45* (1), 32–64.

**Gertler, Mark and Nobuhiro Kiyotaki**, "Chapter 11 - Financial Intermediation and Credit Policy in Business Cycle Analysis," in Benjamin M. Friedman and Michael Woodford, eds., *Benjamin M. Friedman and Michael Woodford, eds.*, Vol. 3 of *Handbook of Monetary Economics*, Elsevier, 2010, pp. 547–599.

**_ and Peter Karadi**, "A model of unconventional monetary policy," *Journal of Monetary Economics*, 2011, *58* (1), 17–34. Carnegie-Rochester Conference Series on Public Policy: The Future of Central Banking April 16-17, 2010.

**Greenwood, Robin, Samuel G. Hanson, and Jeremy C. Stein**, "A Comparative-Advantage Approach to Government Debt Maturity," *Journal of Finance*, 2015, *70* (4), 1683–1722.

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

**Holmstrom, Bengt and Jean Tirole**, "Financial intermediation, loanable funds, and the real sector," *the Quarterly Journal of economics*, 1997, *112* (3), 663–691.

**Jamilov, Rustam and Tommaso Monacelli**, "Bewley banks," *Review of Economic Studies*, 2025, p. rdaf062.

**Kashyap, Anil K., Raghuram Rajan, and Jeremy C. Stein**, "Banks as Liquidity Providers: An Explanation for the Coexistence of Lending and Deposit-taking," *The Journal of Finance*, 2002, *57* (1), 33–73.

**Kashyap, Anil K, Raghuram Rajan, and Jeremy C Stein**, "Banks as liquidity providers: An explanation for the coexistence of lending and deposit-taking," *The Journal of finance*, 2002, *57* (1), 33–73.

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

**Koont, Naz, Tano Santos, and Luigi Zingales**, "Destabilizing digital" bank walks"," Technical Report, National Bureau of Economic Research 2024.

**Kundu, Shohini, Seongjin Park, and Nishant Vats**, "The Geography of Bank Deposits and the Origins of Aggregate Fluctuations," 2021. SSRN Working Paper. Available at SSRN: https://ssrn.com/abstract=3883605 or http://dx.doi.org/10.2139/ssrn.3883605.

**Li, Lei, Elena Loutskina, and Philip E. Strahan**, "Deposit market power, funding stability and long-term credit," *Journal of Monetary Economics*, 2023, *138*, 14–30.

**Lu, Xu, Yang Song, and Yao Zeng**, "Tracing the impact of payment convenience on deposits: Evidence from depositor activeness," *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper, The Wharton School Research Paper*, 2024.

**Piazzesi, Monika and Martin Schneider**, "Payments, credit and asset prices," 2018.

**Schneider, Thomas, Philip Strahan, and Jun Yang**, *Banking in the United States* 04

**Thakor, Anjan and G Yu Edison**, "Funding liquidity creation by banks," *Journal of Financial Stability*, 2024, *73*, 101295.

**Tobin, James**, "Commercial banks as creators of 'money," 1963.

**Wang, Yifei, Toni M. Whited, Yufeng Wu, and Kairong Xiao**, "Bank Market Power and Monetary Policy Transmission: Evidence from a Structural Estimation," *Journal of Finance*, 2022, *77* (4), 2093–2141.

**Werner, Richard A**, "Can banks individually create money out of nothing?—The theories and the empirical evidence," *International review of financial analysis*, 2014, *36*, 1–19.

**Whited, Toni M, Yufeng Wu, and Kairong Xiao**, "Will central bank digital currency disintermediate banks?," *Available at SSRN 4112644*, 2022.

**Xiao, Kairong**, "Monetary Transmission through Shadow Banks," *The Review of Financial Studies*, 10 2019, *33* (6), 2379–2420.

# Appendix

## Contents

**A.1 Institutional Details and Additional Empirical Results** **2**

A.1.1 Non-Bank Multiplier . . . . . 2

A.1.2 First-Stage Estimates . . . . . 4

A.1.3 Subsamples Estimation with Both Instruments . . . . . 5

A.1.4 Estimates of the Credit-Deposit Multiplier . . . . . 6

A.1.5 Robustness Tests . . . . . 7

**A.2 Model Derivations** **8**

A.2.1 Bank First Order Conditions . . . . . 8

A.2.2 Local Pass-Through Derivations . . . . . 8

A.2.3 Spillovers from Money Markets to Non-banks . . . . . 14

A.2.4 How Deposit Inflows Propagate into Credit . . . . . 20

A.2.5 Average and Marginal Deposit Recycling . . . . . 21

A.2.6 Profits and Credit Creation at Longer Horizons . . . . . 23

**A.3 Details on Model Calibration** **24**

A.3.1 Data Construction . . . . . 24

A.3.2 Demand System Details . . . . . 25

A.3.3 Implied Balance Sheet Slack Variables . . . . . 26

**A.4 Additional Counterfactual Results** **27**

A.4.1 Decomposition . . . . . 27

A.4.2 Counterfactual Construction: QE and QT . . . . . 29
A.4.3 Money-Market-to-Non-bank Spillover Calculation . . . . . 30

# A.1 Institutional Details and Additional Empirical Results

## A.1.1 Non-Bank Multiplier

Having established the link between deposits and bank credit, and the forces that give rise to the credit-deposit multiplier, it is useful to consider whether an analogous multiplier operates for non-bank intermediaries.

**Funding composition and credit provision.** Like a bank, a non-bank intermediary's funding composition can shape its willingness to extend credit. More stable funding (longer-maturity debt, committed facilities, or equity-like claims) relaxes rollover and liquidity management frictions and supports additional lending at the margin. In this sense, the marginal lending response to a funding inflow is positive for non-banks as well, so that $\lambda_{nb} > 0$. The microfoundation differs from the bank case: non-banks typically do not face LCR-style liquidity regulation, and the relevant friction operates through funding costs and rollover risk rather than through a regulatory liquidity wedge, but the qualitative channel is the same.

**Equilibrium determination of liabilities.** Like a bank, a non-bank's liabilities are jointly determined in equilibrium by the supply of claims the intermediary is willing to issue and the demand for those claims from investors. Shifts on either side move both the quantity of non-bank funding and its price, just as in the bank deposit market. A non-bank sector facing an inflow of investor demand for its liabilities expands its balance sheet and extends additional credit, through the same logic that governs bank balance sheet expansion.

**The settlement asset and deposit recycling.** The key distinction between banks and non-banks lies in the recycling parameter $\theta$. The credit-deposit feedback loop requires that loan-generated payment flows return to the issuing sector as stable funding. For banks, this recycling is substantial because bank deposits serve as the economy-wide settlement asset: nearly all payments in the real economy clear through deposits, so a large fraction of lending-generated spending mechanically returns to the banking system as deposits. Non-bank liabilities, by contrast, function as settlement media only within narrow subsystems (e.g., money market fund shares in institutional cash management or stablecoins in crypto markets). Payment flows that cross the

boundary of these subsystems leak into bank deposits and do not return as non-bank funding.

As a result, non-bank recycling rates are small in absolute terms, and the limited recycling that does occur within non-bank subsystems typically does not support long-term credit provision to the real economy. Money market funds hold short-dated instruments such as commercial paper and Treasuries rather than extending long-term credit. Stablecoin-based lending finances leveraged positions in other crypto assets rather than real investment, and tri-party repo channels credit to dealers and hedge funds for securities financing rather than to households and firms. Each non-bank subsystem either fails to extend the relevant type of credit, or channels it to leveraged financial intermediaries rather than to the real economy.

**Non-bank Multiplier.** For long-term credit provision to the real economy, the object of interest in this paper, we therefore treat $\theta_{nb} \approx 0$, so that the non-bank multiplier $m_{nb} = 1/(1 - \lambda_{nb}\theta_{nb}) \approx 1$. The credit-deposit feedback loop is a feature specific to the banking system's combination of settlement-asset liabilities and long-term lending to households and firms, and has no close non-bank analog along the dimensions that matter for aggregate credit creation. We therefore focus on banks' credit-deposit multiplier in the main text and abstract from non-bank multipliers.

## A.1.2 First-Stage Estimates

**Table A.1**: First-Stage Estimates

This table reports first-stage estimates from regressions of weekly core deposit growth on high-frequency Treasury cash-flow innovations. The dependent variable is weekly log core deposit growth, winsorized at the 1st and 99th percentiles. $Z_t$ denotes the contemporaneous Treasury-flow innovation normalized by lagged core deposits, and $Z_{t-1}$ denotes its one-week lag. The Treasury-flow residual is also winsorized at the 1st and 99th percentiles before scaling. Columns (1)–(3) include both timing instruments in the full, early, and late samples. Column (4) reports the preferred late-period first stage using only the contemporaneous innovation. All specifications include lagged deposit growth, exclude January weeks and October 2008, and report standard errors in parentheses. The first-stage $F$-statistic tests the excluded instrument set. One, two, and three stars indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

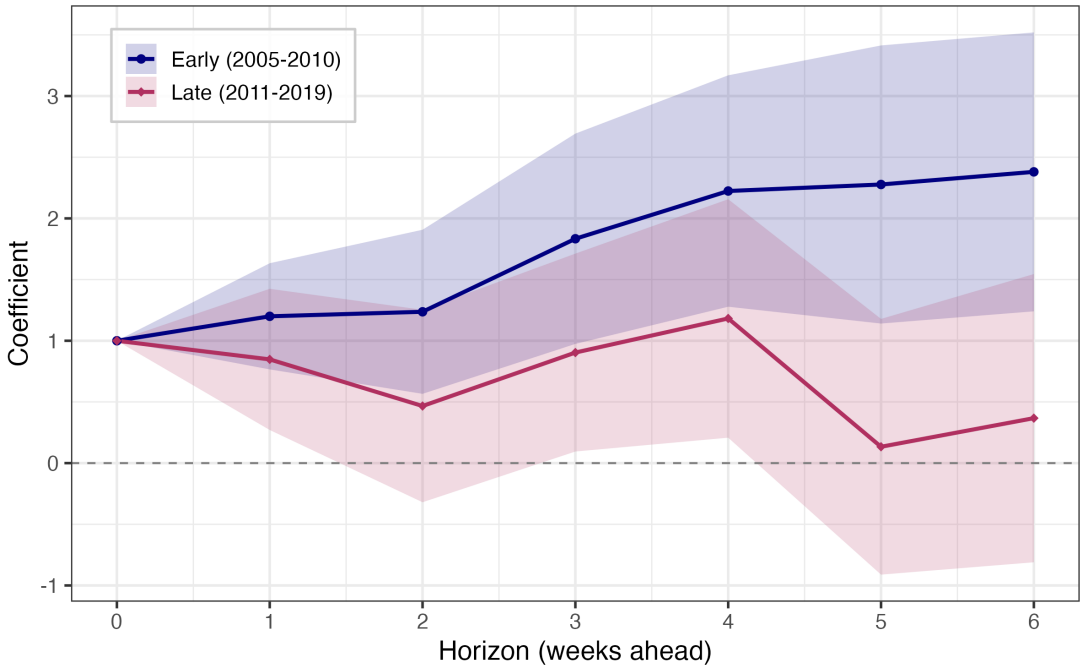
| | (1) Full | (2) Early | (3) Late | (4) Late |
|---|---|---|---|---|
| Lagged shock | $0.19^{***}$ | $0.28^{***}$ | $0.05$ | |
| | $(0.04)$ | $(0.06)$ | $(0.06)$ | |
| Current shock | $0.14^{***}$ | $0.09$ | $0.20^{***}$ | $0.20^{***}$ |
| | $(0.04)$ | $(0.06)$ | $(0.06)$ | $(0.06)$ |
| $\Delta D_{t-1}$ | $-0.29^{***}$ | $-0.30^{***}$ | $-0.27^{***}$ | $-0.26^{***}$ |
| | $(0.04)$ | $(0.06)$ | $(0.05)$ | $(0.05)$ |
| First-stage F | 16.08 | 11.01 | 5.64 | 10.53 |
| $R^2$ | 0.12 | 0.16 | 0.09 | 0.09 |
| Num. obs. | 663 | 234 | 429 | 429 |

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors in parentheses.

### A.1.3 Subsamples Estimation with Both Instruments

**Figure A.1**: Deposit Multiplier Estimates With Both Instruments

This figure reports local projection estimates of the dynamic response of deposits to a deposit shock at time $t$ for the early (2005-2010) and late (2011-2019) samples, estimated using an instrumental variables approach. Shaded bands denote 95% confidence intervals.



### A.1.4 Estimates of the Credit-Deposit Multiplier

**Table A.2**: Estimates of the Credit-Deposit Multiplier

The table reports local projection estimates of the dynamic response of deposits to a deposit shock at time $t$, estimated using an instrumental variables approach. Panel (a) shows the full-sample estimate from 2005 to 2019, panel (b) the estimates for the early period (2005–2010), and panel (c) the estimates for the late period (2011–2019). One, two, and three stars indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The first-stage F statistic for the excluded instruments is reported for each specification. All specifications include lagged deposit growth, exclude January weeks and October 2008, and use deposit-growth variables and Treasury-flow residuals winsorized at the 1st and 99th percentiles.

(a) Full Sample: 2005–2019

| | $h=1$ | $h=2$ | $h=3$ | $h=4$ | $h=5$ | $h=6$ |
|---|---|---|---|---|---|---|
| Deposit Growth | $1.15^{***}$ | $1.14^{***}$ | $1.79^{***}$ | $2.04^{***}$ | $1.95^{***}$ | $2.06^{***}$ |
| | (0.21) | (0.25) | (0.33) | (0.38) | (0.43) | (0.44) |
| First-stage F | 16.08 | 16.08 | 16.08 | 16.08 | 16.08 | 16.08 |
| $R^2$ | 0.25 | 0.09 | −0.07 | −0.11 | −0.25 | −0.11 |
| Num. obs. | 663 | 663 | 663 | 663 | 663 | 663 |

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors in parentheses.

(b) Early Sample: 2005–2010

| | $h=1$ | $h=2$ | $h=3$ | $h=4$ | $h=5$ | $h=6$ |
|---|---|---|---|---|---|---|
| Deposit Growth | $1.20^{***}$ | $1.24^{***}$ | $1.83^{***}$ | $2.22^{***}$ | $2.28^{***}$ | $2.38^{***}$ |
| | (0.27) | (0.32) | (0.42) | (0.51) | (0.60) | (0.61) |
| First-stage F | 11.01 | 11.01 | 11.01 | 11.01 | 11.01 | 11.01 |
| $R^2$ | 0.21 | 0.01 | −0.03 | −0.12 | −0.41 | −0.24 |
| Num. obs. | 234 | 234 | 234 | 234 | 234 | 234 |

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors in parentheses.

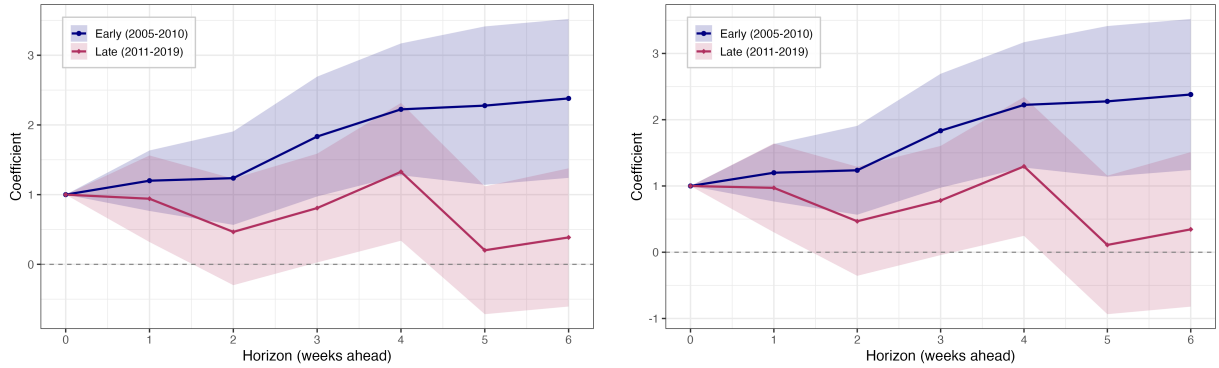
(c) Late Sample: 2011–2019

| | $h=1$ | $h=2$ | $h=3$ | $h=4$ | $h=5$ | $h=6$ |
|---|---|---|---|---|---|---|
| Deposit Growth | $0.95^{***}$ | 0.49 | $0.84^{*}$ | $1.36^{**}$ | 0.25 | 0.45 |
| | (0.32) | (0.38) | (0.43) | (0.53) | (0.58) | (0.63) |
| First-stage F | 10.53 | 10.53 | 10.53 | 10.53 | 10.53 | 10.53 |
| $R^2$ | 0.35 | 0.23 | 0.24 | 0.13 | 0.08 | 0.13 |
| Num. obs. | 429 | 429 | 429 | 429 | 429 | 429 |

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors in parentheses.

**Figure A.2**: Deposit Multiplier Estimates: Robustness to Changes in Reserves

This figure reports local projection estimates of the dynamic response of deposits to a deposit shock at time $t$, estimated using an instrumental variables approach. The early-period estimates use contemporaneous and lagged Treasury-flow innovations as instruments, while the late-period reserve-robustness estimates use only the contemporaneous innovation. It compares estimates for the early (2005–2010) versus late (2011–2019) samples. Shaded bands denote 95% confidence intervals.




(a) Level of Reserves

(b) Changes in Reserves

## A.1.5 Robustness Tests

For our Treasury cash flow innovation shock for deposits, a concern may be that in the post-crisis period, Treasury cash flows may affect reserves through movements in the Treasury's account at the Federal Reserve. This could generate co-movement between deposits and reserves that is not driven by deposit inflows. To test whether reserve movements are a driver of our results, we re-estimate the late period estimate while controlling for the lagged level of aggregate reserves, $R_{t-1}$, in Panel (a) of Figure A.2, and the simultaneous change in aggregate reserves, $\Delta R_t$, in Panel (b) of Figure A.2. The results remain very similar, indicating that the estimates primarily reflect deposit inflows rather than reserve co-movement at weekly frequency.

We also check the preferred Treasury-flow innovation for serial correlation. After residualizing weekly cash flows on week-of-year and quarter-year fixed effects, lags 1–4, and lags 12 and 13, and then winsorizing the residual at the 1st and 99th percentiles, Ljung–Box tests through six lags do not reject the null of no serial correlation at the 10 percent level. The $p$-values are 0.166 in the full sample, 0.412 in the early sample, and 0.123 in the late sample.

## A.2 Model Derivations

### A.2.1 Bank First Order Conditions

The bank chooses loan and deposit rates taking reserves $R$ and equity $E$ as state variables. With wholesale funding determined residually by $W = L + R - E - D$, profits are

$$\Pi = (r^\ell - r^S - \mu_\ell)L + (r^S - r^d - \mu_d)D + \xi \log K + \tau \log \Lambda, \tag{1}$$

where

$$K = E - \kappa(L + R), \qquad \Lambda = R - \chi W = \chi E + (1 - \chi)R - \chi L + \chi D.$$

Using the logit demand derivatives

$$\frac{dL}{dr^\ell} = -\alpha^\ell(1 - s^\ell)L, \qquad \frac{dD}{dr^d} = \alpha^d(1 - s^d)D,$$

the loan and deposit first-order conditions are

$$r^\ell = r^S + \mu_\ell + \frac{1}{\alpha^\ell(1 - s^\ell)} + \chi\frac{\tau}{\Lambda} + \kappa\frac{\xi}{K}, \tag{2}$$

$$r^d = r^S - \mu_d - \frac{1}{\alpha^d(1 - s^d)} + \chi\frac{\tau}{\Lambda}. \tag{3}$$

These are the pricing equations used in estimation. The pass-through derivations below differentiate them around an interior equilibrium.

### A.2.2 Local Pass-Through Derivations

This section derives the local pass-through objects used in the model by totally differentiating the equilibrium system around an interior optimum. We hold equity and policy rates fixed and determine wholesale funding residually from the bank balance sheet. Note that while the main text writes the recycled-deposit pass-through as $\lambda$, here we write it as $\lambda_D$ to distinguish it from the first-round responses to other shocks.

We study small shocks around an interior optimum $x^* = (r_d^*, r_\ell^*, r_{nb}^*, W^*, D^*, L^*, L_{nb}^*, M^*)$.

Given our functional forms, the objective and demand maps are $C^1$ on the domain $\{K > 0, \Lambda > 0\}$, and we assume two regularity conditions at $x^*$:

(i) **Interiority:** the optimum lies in the smooth region, with $K > 0$, $\Lambda > 0$, $W > 0$, and all shares strictly in $(0, 1)$;

(ii) **Nonsingular Jacobian:** the Jacobian of the stacked equilibrium system with respect to $x$ is invertible.

Under (i)–(ii), the Implicit Function Theorem delivers a unique, differentiable local equilibrium map. Its $C^1$ smoothness permits total differentiation of each equilibrium identity, which we carry out in turn for the model's balance-sheet definitions and optimality conditions to obtain the linearized relations among $(d\Lambda, dK, dL, dD, dW, dL_{nb}, dM)$.

The equations to be differentiated are

$$L + R = E + D + W, \qquad \text{(balance sheet)} \tag{4}$$

$$\Lambda = R - \chi W, \qquad \text{(liquidity slack)} \tag{5}$$

$$K = E - \kappa(L + R), \qquad \text{(capital slack)} \tag{6}$$

$$r^{\ell} = r^S + \mu_{\ell} + \frac{1}{\alpha^{\ell}(1 - s_{\ell})} + \chi\frac{\tau}{\Lambda} + \kappa\frac{\xi}{K}, \qquad \text{(bank loan pricing)} \tag{7}$$

$$r^{nb} = r^S + \mu_{nb} + \frac{1}{\alpha^{\ell}(1 - s_{nb})}, \qquad \text{(non-bank pricing)} \tag{8}$$

**(a) Balance sheet identity.** Differentiating (4) gives

$$dW = dL + dR - dD. \tag{9}$$

Thus, a loan expansion must be financed by deposits, reserves, or wholesale funding through the balance sheet.

**(b) Regulatory slacks.** Differentiating (5) and (6), and substituting (9), gives

$$d\Lambda = (1 - \chi)dR - \chi dL + \chi dD, \qquad dK = -\kappa(dL + dR). \tag{10}$$

The corresponding differential of the regulatory part of the loan-pricing condition is

$$d\left(\chi\frac{\tau}{\Lambda}+\kappa\frac{\xi}{K}\right)=\left(\frac{\tau\chi^2}{\Lambda^2}+\frac{\xi\kappa^2}{K^2}\right)dL-\frac{\tau\chi^2}{\Lambda^2}dD + \left(\frac{\xi\kappa^2}{K^2}-\frac{\tau\chi(1-\chi)}{\Lambda^2}\right)dR. \tag{11}$$

The $dL$ term is the marginal tightening from balance-sheet expansion. The $dD$ term is the liquidity relief from replacing wholesale funding with deposits. The $dR$ term is the net reserve effect: reserves relax the liquidity constraint but expand assets relevant for the capital constraint.

**(c) Bank loan pricing.** Differentiating the bank loan-pricing condition (7), holding $r^S$ and $\mu_\ell$ fixed, gives

$$dr^\ell=\frac{1}{\alpha^\ell(1-s_\ell)^2}ds_\ell+d\left(\chi\frac{\tau}{\Lambda}+\kappa\frac{\xi}{K}\right). \tag{12}$$

With primitive borrower demand shifters fixed, the logit share derivatives are

$$ds_\ell=-\alpha^\ell s_\ell(1-s_\ell)dr^\ell+\alpha^\ell s_\ell s_{nb}dr^{nb}, \tag{13}$$

The corresponding bank loan quantity derivative is

$$dL=-\alpha^\ell L\big[(1-s_\ell)dr^\ell-s_{nb}dr^{nb}\big]. \tag{14}$$

Substituting (13) into (12) yields

$$dr^\ell=-\frac{s_\ell}{1-s_\ell}dr^\ell+\frac{s_\ell s_{nb}}{(1-s_\ell)^2}dr^{nb}+d\left(\chi\frac{\tau}{\Lambda}+\kappa\frac{\xi}{K}\right), \tag{15}$$

or, collecting the $dr^\ell$ terms,

$$dr^\ell=\frac{s_\ell s_{nb}}{1-s_\ell}dr^{nb}+(1-s_\ell)d\left(\chi\frac{\tau}{\Lambda}+\kappa\frac{\xi}{K}\right). \tag{16}$$

**(d) Non-bank pricing.** Differentiating the non-bank pricing condition (8), holding $r^S$ and $\mu_{nb}$ fixed, gives

$$dr^{nb}=\frac{1}{\alpha^\ell(1-s_{nb})^2}ds_{nb}. \tag{17}$$

With primitive borrower demand shifters fixed, the non-bank share derivative is

$$ds_{nb} = \alpha^{\ell} s_{nb}\big[s_{\ell} dr^{\ell} - (1 - s_{nb}) dr^{nb}\big]. \tag{18}$$

Substituting (18) into (17) gives

$$dr^{nb} = \frac{s_{nb}}{(1 - s_{nb})^2}\big[s_{\ell} dr^{\ell} - (1 - s_{nb}) dr^{nb}\big].$$

Solving for the non-bank rate response,

$$\frac{dr^{nb}}{dr^{\ell}} = \frac{s_{\ell} s_{nb}}{1 - s_{nb}}. \tag{19}$$

**(e) Residual loan-demand curvature.** Combining bank loan demand (14) with (19) gives

$$dr^{\ell} = -\frac{1}{\alpha^{\ell} L\left[(1 - s_{\ell}) - s_{nb}\frac{dr^{nb}}{dr^{\ell}}\right]} dL.$$

The bank loan-pricing condition equates the loan rate to the demand markup plus regulatory wedges. To isolate the movement in those wedges, we subtract the change in the demand markup from $dr^{\ell}$. Since the markup is $1/[\alpha^{\ell}(1 - s_{\ell})]$, its derivative along the same loan-demand curve is

$$d\left[\frac{1}{\alpha^{\ell}(1 - s_{\ell})}\right] = \frac{s_{\ell}}{\alpha^{\ell} L (1 - s_{\ell})^2} dL.$$

The left-hand side of the differentiated bank loan-pricing condition is therefore

$$\begin{aligned} dr^{\ell} - d\left[\frac{1}{\alpha^{\ell}(1 - s_{\ell})}\right] &= -\left[\frac{1}{\alpha^{\ell} L\left((1 - s_{\ell}) - s_{nb}\frac{dr^{nb}}{dr^{\ell}}\right)} + \frac{s_{\ell}}{\alpha^{\ell} L (1 - s_{\ell})^2}\right] dL \\ &= -\frac{1}{\tilde{\kappa}_{\ell}^{nb}} dL. \end{aligned}$$

Equivalently,

$$\tilde{\kappa}_{\ell}^{nb} = \alpha^{\ell} L (1 - s_{\ell})^2 \frac{(1 - s_{\ell}) - s_{nb}\frac{dr^{nb}}{dr^{\ell}}}{(1 - s_{\ell}) - s_{\ell} s_{nb}\frac{dr^{nb}}{dr^{\ell}}}. \tag{20}$$

$$dr^{\ell} - d\left[\frac{1}{\alpha^{\ell}(1 - s_{\ell})}\right] = -\frac{1}{\tilde{\kappa}_{\ell}^{nb}} dL. \tag{21}$$

This is the effective residual loan-demand curvature facing banks after borrower substitution and non-bank repricing.

Using (21), the primitive loan-pricing condition becomes

$$-\frac{1}{\tilde{\kappa}_\ell^{nb}} dL = d\left(\chi \frac{\tau}{\Lambda} + \kappa \frac{\xi}{K}\right). \tag{22}$$

Substituting (11) into (22) gives

$$\begin{aligned}\left[1 + \tilde{\kappa}_\ell^{nb}\left(\frac{\tau\chi^2}{\Lambda^2} + \frac{\xi\kappa^2}{K^2}\right)\right] dL = \tilde{\kappa}_\ell^{nb}\frac{\tau\chi^2}{\Lambda^2} dD \\ + \tilde{\kappa}_\ell^{nb}\left(\frac{\tau\chi(1-\chi)}{\Lambda^2} - \frac{\xi\kappa^2}{K^2}\right) dR.\end{aligned} \tag{23}$$

The $dD$ term is the liquidity relief from deposit funding. The $dR$ term is the net reserve effect.

**Recycled-deposit pass-through.** For a recycled deposit inflow, reserves are fixed. Setting $dR = 0$ in (23) gives

$$\left[1 + \tilde{\kappa}_\ell^{nb}\left(\frac{\tau\chi^2}{\Lambda^2} + \frac{\xi\kappa^2}{K^2}\right)\right] dL = \tilde{\kappa}_\ell^{nb}\frac{\tau\chi^2}{\Lambda^2} dD.$$

Hence

$$\lambda_D \equiv \frac{dL}{dD} = \frac{\tilde{\kappa}_\ell^{nb}\frac{\tau\chi^2}{\Lambda^2}}{1 + \tilde{\kappa}_\ell^{nb}\left(\frac{\tau\chi^2}{\Lambda^2} + \frac{\xi\kappa^2}{K^2}\right)} = \frac{\frac{\tau\chi^2}{\Lambda^2}}{1/\tilde{\kappa}_\ell^{nb} + \frac{\tau\chi^2}{\Lambda^2} + \frac{\xi\kappa^2}{K^2}}. \tag{24}$$

With $dR = 0$, wholesale funding moves one-for-one with the part of the deposit inflow not used for bank credit:

$$\frac{dW}{dD} = \lambda_D - 1. \tag{25}$$

The associated non-bank credit response through borrower substitution follows from the same borrower-demand derivatives. Holding borrower demand shifters fixed,

$$\begin{aligned} dL &= -\alpha^\ell L\left[(1 - s_\ell)dr^\ell - s_{nb}dr^{nb}\right], \\ dL_{nb} &= \alpha^\ell L_{nb}\left[s_\ell dr^\ell - (1 - s_{nb})dr^{nb}\right]. \end{aligned}$$

Using the local non-bank repricing response (19), the bank-loan derivative implies

$$dr^{\ell} = -\frac{1}{\alpha^{\ell} L\left[(1-s_{\ell}) - s_{nb}\frac{dr^{nb}}{dr^{\ell}}\right]} dL.$$

Substituting this expression into the non-bank quantity derivative gives

$$dL_{nb} = -\frac{L_{nb}}{L}\frac{s_{\ell} - (1-s_{nb})\frac{dr^{nb}}{dr^{\ell}}}{(1-s_{\ell}) - s_{nb}\frac{dr^{nb}}{dr^{\ell}}} dL.$$

For a marginal recycled-deposit inflow, $dL = \lambda_D dD$. Therefore

$$\lambda_{nb,D} \equiv \frac{\partial L_{nb}}{\partial D} = -\lambda_D \frac{L_{nb}}{L}\frac{s_{\ell} - (1-s_{nb})\frac{dr^{nb}}{dr^{\ell}}}{(1-s_{\ell}) - s_{nb}\frac{dr^{nb}}{dr^{\ell}}}. \tag{26}$$

This is the local non-bank response per dollar of realized recycled deposits. The total response to an initial deposit impulse is obtained below after distinguishing the shock-specific first-round lending response from the recycled-deposit response that governs subsequent rounds.

**Reserve-backed deposit shock.** For a reserve-backed deposit injection, reserves move with deposits. Returning to (23) and dividing by the coefficient on $dL$ gives

$$dL = \lambda_D dD - \frac{\frac{\xi\kappa^2}{K^2} - \frac{\tau\chi(1-\chi)}{\Lambda^2}}{1/\tilde{\kappa}_{\ell}^{nb} + \frac{\tau\chi^2}{\Lambda^2} + \frac{\xi\kappa^2}{K^2}} dR. \tag{27}$$

The second term is the net balance-sheet-expansion cost of reserves: the accompanying asset expansion tightens the capital constraint, while reserves partly offset this by relaxing the liquidity constraint. A reserve-backed deposit injection sets $dR = dD$, so the first-round reserve-backed lending response is

$$\lambda^R \equiv \left.\frac{dL}{dD}\right|_{dR=dD} = \lambda_D - \frac{\frac{\xi\kappa^2}{K^2} - \frac{\tau\chi(1-\chi)}{\Lambda^2}}{1/\tilde{\kappa}_{\ell}^{nb} + \frac{\tau\chi^2}{\Lambda^2} + \frac{\xi\kappa^2}{K^2}} = \frac{\frac{\tau\chi}{\Lambda^2} - \frac{\xi\kappa^2}{K^2}}{1/\tilde{\kappa}_{\ell}^{nb} + \frac{\tau\chi^2}{\Lambda^2} + \frac{\xi\kappa^2}{K^2}}. \tag{28}$$

This object is a first-round lending response to a deposit inflow that arrives with reserves. Unlike $\lambda_D$, it need not be positive: if the capital cost from the accompanying asset expansion is sufficiently large, $\xi\kappa^2/K^2 > \tau\chi/\Lambda^2$, then the numerator in (28) is negative and reserve-backed deposits reduce bank lending at the margin. It is distinct from the recycled-deposit response $\lambda_D$,

which applies to later loan-generated deposits that do not come with new reserves.

**First-round and recycled-round accounting.** Let $dD_0$ denote the initiating deposit impulse, and let $\lambda^0$ denote the bank lending response that applies to that first impulse. For a pure recycled-deposit shock, $\lambda^0 = \lambda_D$. For a reserve-backed deposit shock, $\lambda^0 = \lambda^R$. Subsequent loan-generated deposits are ordinary recycled deposits, so those rounds propagate with $\lambda_D$, not necessarily with $\lambda^0$.

The bank-credit recursion is therefore

$$dL_{\text{tot}}^{\text{bank}} = \lambda^0 dD_0 + \lambda_D \theta dL_{\text{tot}}^{\text{bank}}. \tag{29}$$

Solving,

$$\frac{dL_{\text{tot}}^{\text{bank}}}{dD_0} = \frac{\lambda^0}{1 - \theta\lambda_D}. \tag{30}$$

The corresponding deposit multiplier is

$$\frac{dD_{\text{tot}}}{dD_0} = 1 + \theta \frac{dL_{\text{tot}}^{\text{bank}}}{dD_0} = 1 + \frac{\theta\lambda^0}{1 - \theta\lambda_D}. \tag{31}$$

When $\lambda^0 = \lambda_D$, this collapses to the usual $(1 - \theta\lambda_D)^{-1}$ multiplier. When the first round is reserve-backed, $\lambda^0 = \lambda^R$, giving the post-period multiplier target used in (42).

The non-bank borrower-substitution response scales with total bank credit, not directly with the initial deposit impulse. Using (26),

$$dL_{nb,\text{tot}}^{\text{sub}} = \frac{\lambda_{nb,D}}{\lambda_D} dL_{\text{tot}}^{\text{bank}} = \frac{\lambda_{nb,D}}{\lambda_D} \frac{\lambda^0}{1 - \theta\lambda_D} dD_0. \tag{32}$$

Thus, holding money-market depth fixed, the total aggregate-credit response to the initial deposit impulse is

$$\frac{dL_{\text{total}}}{dD_0} = \left(1 + \frac{\lambda_{nb,D}}{\lambda_D}\right) \frac{\lambda^0}{1 - \theta\lambda_D}. \tag{33}$$

### A.2.3 Spillovers from Money Markets to Non-banks

The local pass-through derivation above uses the baseline non-bank pricing condition in (8). This subsection adds the spillover from money markets to non-banks by allowing short-term

funding depth $M$ to shift non-bank marginal intermediation costs directly. Let

$$\mu_{nb}(M) = \bar{\mu}_{nb} - \kappa_M M, \qquad \kappa_M > 0,$$

so that deeper money markets lower non-bank marginal costs. Taking $M$ as given when setting the loan rate, the extended non-bank profit problem is

$$\max_{r^{nb}} \ \Pi_{nb}^{M} = (r^{nb} - r^{S} - \mu_{nb}(M))L_{nb}(r^{nb}). \tag{34}$$

The first-order condition is

$$0 = L_{nb} + \left(r^{nb} - r^{S} - \mu_{nb}(M)\right) \frac{\partial L_{nb}}{\partial r^{nb}}. \tag{35}$$

Using the logit own-rate derivative

$$\frac{\partial L_{nb}}{\partial r^{nb}} = -\alpha^{\ell}(1 - s_{nb})L_{nb},$$

(35) gives the extended non-bank pricing condition:

$$r^{nb} = r^{S} + \mu_{nb}(M) + \frac{1}{\alpha^{\ell}(1 - s_{nb})} \tag{36}$$

Its differential is

$$dr^{nb} = \frac{1}{\alpha^{\ell}(1 - s_{nb})^2} ds_{nb} + \mu'_{nb}(M)dM. \tag{37}$$

The same borrower-substitution term moves both the non-bank share and non-bank loan quantity:

$$ds_{nb} = \alpha^{\ell} s_{nb}\left[s_{\ell} dr^{\ell} - (1 - s_{nb})dr^{nb}\right],$$
$$dL_{nb} = \alpha^{\ell} L_{nb}\left[s_{\ell} dr^{\ell} - (1 - s_{nb})dr^{nb}\right].$$

Substituting the share derivative into (37) gives

$$dr^{nb} = \frac{s_{nb}}{(1 - s_{nb})^2}\left[s_{\ell} dr^{\ell} - (1 - s_{nb})dr^{nb}\right] + \mu'_{nb}(M)dM.$$

Solving for the non-bank rate response,

$$dr^{nb} = \frac{s_{\ell} s_{nb}}{1 - s_{nb}} dr^{\ell} + (1 - s_{nb})\mu'_{nb}(M)dM = \frac{s_{\ell} s_{nb}}{1 - s_{nb}} dr^{\ell} - (1 - s_{nb})\kappa_M dM. \tag{38}$$

Thus the money-market spillover does not change the local non-bank repricing response used in the recycled-deposit pass-through derivation: when $dM = 0$,

$$\frac{dr^{nb}}{dr^{\ell}} = \frac{s_{\ell} s_{nb}}{1 - s_{nb}},$$

the same expression as in (19). The additional term in (38) matters only when the shock moves money-market depth.

Substituting (38) into the non-bank quantity derivative gives

$$dL_{nb} = \alpha^{\ell} L_{nb}(1 - s_{nb}) s_{\ell} dr^{\ell} - \alpha^{\ell} L_{nb}(1 - s_{nb})^2 \mu'_{nb}(M) dM. \tag{39}$$

Holding the bank loan rate fixed, the direct non-bank response to money-market depth is therefore

$$\lambda^{direct}_{nb,M} \equiv \left. \frac{\partial L_{nb}}{\partial M} \right|_{dr^{\ell}=0} = -\alpha^{\ell} L_{nb}(1 - s_{nb})^2 \mu'_{nb}(M) = \alpha^{\ell} L_{nb}(1 - s_{nb})^2 \kappa_M. \tag{40}$$

Money-market depth also shifts bank credit through borrower substitution. Bank loan demand and (38) imply

$$dr^{\ell} = -\frac{1}{\alpha^{\ell} L \left[ (1 - s_{\ell}) - s_{nb} \frac{s_{\ell} s_{nb}}{1 - s_{nb}} \right]} dL + \frac{s_{nb}(1 - s_{nb}) \mu'_{nb}(M)}{(1 - s_{\ell}) - s_{nb} \frac{s_{\ell} s_{nb}}{1 - s_{nb}}} dM.$$

Subtracting the bank loan-markup derivative, as in (21), and substituting into the differentiated loan-pricing condition yields

$$\begin{aligned} \left[ 1 + \tilde{\kappa}^{nb}_{\ell} \left( \frac{\tau \chi^2}{\Lambda^2} + \frac{\xi \kappa^2}{K^2} \right) \right] dL = \tilde{\kappa}^{nb}_{\ell} \frac{\tau \chi^2}{\Lambda^2} dD \\ + \tilde{\kappa}^{nb}_{\ell} \left( \frac{\tau \chi (1 - \chi)}{\Lambda^2} - \frac{\xi \kappa^2}{K^2} \right) dR \\ + \tilde{\kappa}^{nb}_{\ell} \frac{s_{nb}(1 - s_{nb}) \mu'_{nb}(M)}{(1 - s_{\ell}) - s_{nb} \frac{s_{\ell} s_{nb}}{1 - s_{nb}}} dM. \end{aligned} \tag{41}$$

The first two terms are the baseline recycled-deposit and reserve effects. The final term is the money-market spillover. Since $\mu'_{nb}(M) < 0$, an increase in $M$ lowers non-bank loan rates and

shifts some credit away from banks:

$$\lambda_{L,M} \equiv \frac{\partial L}{\partial M} = \frac{\tilde{\kappa}_\ell^{nb} \dfrac{s_{nb}(1-s_{nb})\mu'_{nb}(M)}{(1-s_\ell) - s_{nb}\dfrac{s_\ell s_{nb}}{1-s_{nb}}}}{1 + \tilde{\kappa}_\ell^{nb}\left(\dfrac{\tau\chi^2}{\Lambda^2} + \dfrac{\xi\kappa^2}{K^2}\right)} = -\frac{\tilde{\kappa}_\ell^{nb} \dfrac{s_{nb}(1-s_{nb})\kappa_M}{(1-s_\ell) - s_{nb}\dfrac{s_\ell s_{nb}}{1-s_{nb}}}}{1 + \tilde{\kappa}_\ell^{nb}\left(\dfrac{\tau\chi^2}{\Lambda^2} + \dfrac{\xi\kappa^2}{K^2}\right)}. \tag{42}$$

**Aggregate-credit accounting.** The bank response in (41) is only one part of the total-credit response. Substituting the bank loan-demand derivative into (39) gives the non-bank quantity response as a function of the bank-loan quantity and money-market depth:

$$dL_{nb} = \underbrace{-\frac{L_{nb}}{L}\frac{s_\ell(1-s_{nb})}{(1-s_\ell) - s_{nb}\frac{s_\ell s_{nb}}{1-s_{nb}}}}_{\text{borrower substitution per dollar of } dL} dL + \lambda^{cost}_{nb,M} dM, \tag{43}$$

The cost-channel coefficient is the coefficient on $dM$ in (39) after holding fixed the bank-loan quantity $dL$. It has a direct non-bank marginal-cost piece and an indirect bank-rate piece:

$$\begin{aligned}\lambda^{cost}_{nb,M} &= -\alpha^\ell L_{nb}(1-s_{nb})^2\mu'_{nb}(M) \quad + \alpha^\ell L_{nb}(1-s_{nb})s_\ell \frac{s_{nb}(1-s_{nb})\mu'_{nb}(M)}{(1-s_\ell) - s_{nb}\frac{s_\ell s_{nb}}{1-s_{nb}}} \\ &= -\alpha^\ell L_{nb}(1-s_{nb})^2\mu'_{nb}(M)\left[1 - \frac{s_\ell s_{nb}}{(1-s_\ell) - s_{nb}\frac{s_\ell s_{nb}}{1-s_{nb}}}\right].\end{aligned} \tag{44}$$

The first addend is the direct $dM$ term in (39); the second substitutes the $dM$-component of $dr^\ell$ from the bank loan-demand derivative into the $dr^\ell$ term in (39). The first term in (43) is ordinary borrower substitution: when bank lending rises and bank loan rates fall, non-bank credit falls. The second term is the remaining money-market cost channel after this borrower-substitution response to $dL$ has been separated out. With $\mu'_{nb}(M) = -\kappa_M$, this term is positive when deeper money markets lower non-bank marginal costs.

Combining (41) with (43), the local change in bank and non-bank credit, before closing

the deposit-recycling loop, can be written as

$$\begin{aligned} dL + dL_{nb} &= \underbrace{dL}_{\text{bank credit}} + \underbrace{\frac{\lambda_{nb,D}}{\lambda_D} dL}_{\text{non-bank borrower substitution}} + \underbrace{\lambda^{cost}_{nb,M} dM}_{\text{money-market cost channel}} \\ &= \left(1 + \frac{\lambda_{nb,D}}{\lambda_D}\right)(\lambda_D dD + \lambda_{L,M} dM) + \lambda^{cost}_{nb,M} dM, \end{aligned} \tag{45}$$

where $\lambda_{nb,D}/\lambda_D$ is the borrower-substitution term from (26). The total-credit response to an initial deposit impulse is obtained only after adding the $\theta$ accounting below.

**Loan-proceed leakage into money markets.** The one-step expression in (45) takes $dM$ as an external change in money-market depth. To keep the money-market spillover as a diagnostic extension rather than a second fixed point, we select the no-spillover total-credit response as the allocation that generates payment flows. From (33),

$$\frac{dL^0_{\text{total}}}{dD_0} = \left(1 + \frac{\lambda_{nb,D}}{\lambda_D}\right) \frac{\lambda^0}{1 - \theta\lambda_D}, \tag{46}$$

where the superscript $0$ denotes $\kappa_M = 0$. Under full leakage of non-recycled proceeds into money-market funds, the selected money-market inflow is

$$dM^{sel} = (1 - \theta) dL^0_{\text{total}}. \tag{47}$$

Evaluating the money-market spillover once around this selected allocation gives

$$\begin{aligned} \frac{dL^{spill}_{\text{total}}}{dD_0} &= \frac{dL^0_{\text{total}}}{dD_0} + \left[\left(1 + \frac{\lambda_{nb,D}}{\lambda_D}\right)\lambda_{L,M} + \lambda^{cost}_{nb,M}\right] \frac{dM^{sel}}{dD_0} \\ &= \frac{dL^0_{\text{total}}}{dD_0} \left\{1 + (1 - \theta)\left[\left(1 + \frac{\lambda_{nb,D}}{\lambda_D}\right)\lambda_{L,M} + \lambda^{cost}_{nb,M}\right]\right\}. \end{aligned} \tag{48}$$

This selected-allocation calculation accounts for the fact that non-bank credit also generates payment flows, while avoiding an additional money-market fixed point. If only a fraction of non-recycled proceeds enters money-market funds, replace $1-\theta$ in (47) and (48) by that fraction times $1 - \theta$.

**Deposit-preference shock.** A pure preference shift from money market funds into deposits has $dR = 0$ and $dM = -dD$. From (41),

$$dL = \lambda_D dD + \lambda_{L,M} dM.$$

Setting $dM = -dD$ gives

$$\lambda_{\xi_D} \equiv \left.\frac{dL}{dD}\right|_{dR=0, dM=-dD} = \lambda_D - \lambda_{L,M}. \tag{49}$$

Because $\lambda_{L,M} < 0$, a shift out of money market funds and into deposits raises bank lending both through the usual deposit-liquidity channel and by increasing non-bank marginal costs. Since reserves are fixed, $dW/dD = \lambda_{\xi_D} - 1$. For the corresponding local change in bank and non-bank credit, (45) gives

$$\left.\frac{dL + dL_{nb}}{dD}\right|_{dR=0, dM=-dD} = \left(1 + \frac{\lambda_{nb,D}}{\lambda_D}\right)(\lambda_D - \lambda_{L,M}) - \lambda_{nb,M}^{cost}. \tag{50}$$

For a reserve injection that also changes depositor wealth, such as the QE counterfactual, the imposed split determines $dD$, and money-market-fund assets absorb the remaining reserve-created wealth:

$$dM = dR - dD.$$

Equations (41) and (45) combine the local bank-credit response, non-bank borrower substitution, and the direct money-market cost channel at the resulting balance-sheet state. For loan-proceed spillovers, the leakage accounting in (48) additionally links the change in $M$ to the non-recycled share of bank lending.

**Wholesale funding.** All wholesale-funding responses come from the balance sheet:

$$dW = dL + dR - dD.$$

Thus $dW/dD = \lambda_D - 1$ for recycled deposits, $dW/dD = \lambda_{\xi_D} - 1$ for deposit-preference shocks, and $dW/dD = \lambda^R$ for reserve-backed deposit injections with $dR = dD$.

### A.2.4 How Deposit Inflows Propagate into Credit

To build intuition for how deposit inflows generate credit, consider an exogenous stable deposit inflow $dD_0$.[20] In equilibrium, total deposits $dD$ and total loans $dL$ are jointly determined by two conditions,

$$dL = \lambda dD \tag{51}$$

$$dD = dD_0 + \theta dL. \tag{52}$$

The first, Equation (51), is the bank's optimal lending response which states that the bank extends $\lambda$ dollars of new loans for each dollar of deposits. The second, Equation (52), is the deposit recycling condition which states that total deposits equal the exogenous inflow plus the endogenous deposits created by lending: when the bank extends $dL$ in loans, borrowers spend those funds, generating payment flows of which a fraction $\theta$ is retained within the banking system as stable deposits. Solving this system yields directly the deposit multiplier and the corresponding increase in bank loans,

$$dD = \frac{dD_0}{1 - \lambda\theta}, \quad \text{and} \quad dL = \lambda \frac{dD_0}{1 - \lambda\theta}. \tag{53}$$

To see how this equilibrium quantity of deposits and loans arises, consider the round-by-round balance sheet dynamics. Since no new reserves accompany the deposit inflows, the initial effect is a pure liability recomposition: the stable deposit $dD_0$ replaces the wholesale funding one-for-one, with no change in balance sheet size. Because stable deposits are less runnable than wholesale, this recomposition relaxes the bank's liquidity constraint, creating room for balance sheet expansion.

The bank responds by originating $\lambda dD_0$ in new loans. When the bank lends, it simultaneously creates a new liability: the borrower's account is credited. This new liability is initially flighty, because the borrower may intend to spend the funds immediately. The balance sheet has expanded by $\lambda dD_0$ on both sides, with new loans on the asset side, and new wholesale funding on the liability side. The original stable deposit $dD_0$ remains in place, backing existing reserves.

[20] Such an inflow arises naturally from a shift in depositor preferences $\xi_d$ that reallocates household savings from non-bank vehicles into bank deposits, or from a shift in borrower demand $\xi_\ell$ that stimulates lending and generates payment flows that settle as stable deposits. In either case, the inflow represents a change in the composition of bank liabilities rather than an injection of outside reserves.

It was not used to fund the loan, rather it relaxed the liquidity constraint that made the bank willing to expand.[21]

As the borrowers spend the loan proceeds and the resulting payment flows circulate through the economy, a fraction $\theta$ of the initially flighty liabilities settle into stable core deposits. This adds $\theta\lambda dD_0$ in stable funding to the banking system without changing balance sheet size. It is a liability recomposition from flighty to stable funding. The remaining $(1-\theta)\lambda dD_0$ stay as flighty liabilities that do not support further expansion. The new stable deposits relax the liquidity constraint further, enabling the bank to originate another $\lambda^2\theta dD_0$ in loans, again creating flighty liabilities that partially stabilize through $\theta$. The process continues with diminishing rounds, converging to the fixed point derived above.

### A.2.5 Average and Marginal Deposit Recycling

The baseline propagation system treats the recycling rate $\theta$ as locally constant. This is the relevant object for the multiplier exercise above, which asks how an additional deposit inflow propagates around a given balance sheet and market environment. It is useful, however, to distinguish this marginal recycling rate from the average recycling embedded in the stock equilibrium.

Let stable deposits generated by the outstanding loan book be summarized by an aggregate recycling function $\Theta(L)$, so that

$$D = D^0 + \Theta(L), \tag{54}$$

where $D^0$ denotes deposits generated outside the loan-payment feedback loop. The average recycling rate associated with the stock of bank credit is

$$\theta^{avg}(L) = \frac{\Theta(L)}{L}, \tag{55}$$

whereas the marginal recycling rate relevant for the propagation of an additional dollar of lending

[21] An equivalent accounting decomposition is to view the deposit inflow $dD_0$ as directly funding the loan: a fraction $\lambda$ finances new lending while the remaining $1-\lambda$ retires wholesale. Both decompositions produce identical balance sheet outcomes: assets increase by $\lambda dD_0$, deposits increase by $dD_0$, and wholesale falls by $(1-\lambda)dD_0$. However, the split decomposition obscures the role of $\theta$, since the stable funding appears to leave when the borrower spends, with only $\theta\lambda dD_0$ returning. The self-funding interpretation avoids this confusion: the stable deposit remains in place, lending creates flighty liabilities, and $\theta$ unambiguously adds stable funding each round.

is

$$\theta^{marg}(L) = \Theta'(L). \tag{56}$$

The constant-$\theta$ case used in the baseline corresponds to the linear special case $\Theta(L) = \theta L$, in which average and marginal recycling coincide.

In general, the two objects can differ. If the composition of borrowers, payees, or cash managers changes with the scale of lending, the fraction of loan-generated payment flows that settles as stable deposits need not be constant. For example, inframarginal lending may generate payments to households and firms that hold balances as stable bank deposits, while marginal lending may increasingly generate payments to institutional cash managers that sweep funds into money market funds, repo, Treasury bills, or other short-term funding vehicles. In that case $\Theta(L)$ is concave and $\theta^{marg}(L) < \theta^{avg}(L)$. Conversely, in a flight-to-safety episode or a shift toward deposit-special recipients, marginal payment flows may be more likely to remain in stable deposits than the existing stock, so $\theta^{marg}(L) > \theta^{avg}(L)$.

This distinction does not alter the bank's first-order conditions. Individual banks do not internalize the aggregate recycling function when they originate credit: loan proceeds may settle at other banks or outside the banking system, and the originating bank does not control the fraction that returns as stable funding. Recycling is therefore an aggregate payment-system object rather than a private choice variable. The bank FOCs determine the marginal lending response $\lambda$ to a realized deposit inflow, while the recycling function determines how lending-generated payment flows feed back into the banking system.

Our empirical multiplier and minimum-distance estimation discipline the local marginal object, $\theta^{marg}$, because the question is how an identified deposit inflow propagates through additional rounds of lending and deposit creation. Recovering an average recycling rate would require decomposing the stock of deposits into balances generated by outstanding bank lending versus balances generated by outside sources such as household portfolio demand, government transfers, central bank asset purchases, foreign inflows, or other non-lending channels. Since these balances are not separately tagged in aggregate data, the paper focuses on the marginal recycling rate that governs deposit-shock transmission.

### A.2.6 Profits and Credit Creation at Longer Horizons

At longer horizons, the credit-creating capacity of deposit inflows is further strengthened by profits generated through lending activity. When borrowers use credit to finance production, they generate output $Y$ and earn profits $\Pi = Y - wN - r^{\ell}L - r^{nb}L_{nb}$, while banks and nonbank lenders earn intermediation profits from interest spreads. These profits ultimately accrue to households as wages, dividends, and retained earnings, raising depositor wealth $W_d$ and shifting out deposit demand. To the extent that bank profits are retained rather than distributed, they also raise bank equity $E$, relaxing the capital constraint and increasing the bank's willingness to expand its balance sheet.

The first channel operates through $\theta$. Note that while the total pool of cash injected into the payment system by lending remains $\Delta L$, higher household wealth $W_d$ from positive profits shifts out deposit demand, increasing the fraction of circulating balances that households choose to hold as stable deposits, effectively raising $\theta$. The second channel operates through $\lambda$. Retained bank profits raise equity $E$, which relaxes the capital constraint $K = E - \kappa(L + R)$ and lowers the shadow cost of balance sheet expansion, increasing the lending response $\lambda$ in response to any given deposit inflow.

Both channels operate at business cycle frequency, since profits must be realized, distributed or retained, and reinvested before they affect deposit demand or bank equity, whereas the credit-deposit multiplier that we study propagates through bank balance sheets at a high frequency horizon over weeks. We therefore hold $E$ fixed and evaluate $\theta$ at prevailing deposit demand conditions, so that the multiplier captures short-run balance sheet propagation in isolation.

The profit-driven amplification that we abstract from is closely related to the financial accelerator studied in the macro-finance literature (Bernanke et al., 1999; Gertler and Kiyotaki, 2010; He and Krishnamurthy, 2013), in which higher intermediary net worth relaxes borrowing constraints and enables further credit expansion. Our credit-deposit multiplier operates before and independently of this channel. At the same time, the two mechanisms would interact at longer horizons: profits raise depositor wealth and bank equity, which strengthen the credit-deposit feedback loop through both $\theta$ and $\lambda$, generating more lending and further profits.

## A.3 Details on Model Calibration

### A.3.1 Data Construction

This appendix describes the construction of the aggregate quantity and price series used in the calibration and estimation of the model.

We assemble a quarterly panel for the U.S. commercial banking sector spanning 1999Q1 to 2024Q4. Balance sheet quantities are obtained from the Federal Reserve's weekly series for all commercial banks, including total assets, total liabilities, total loans, total deposits, and large time deposits. Weekly observations are converted to quarterly frequency by taking the last available observation within each calendar quarter. Core deposits are defined as total deposits net of large time deposits. Wholesale funding is defined as total liabilities net of core deposits, so large time deposits enter residual wholesale bank funding rather than the stable deposit measure. Equity is constructed residually as the difference between assets and liabilities. The model's liquid asset position $R$ is measured as total bank assets net of bank credit, where bank credit includes loans and securities. This measure includes central bank reserves together with other liquid claims such as fed funds sold, reverse repo, and cash items, which provide effective liquidity alongside reserves. It is therefore larger than the stock of reserve balances alone, and the liquidity slack $\Lambda$ and the QE and QT exercises should be read with this broader liquid asset measure in mind.

Deposit and loan rates are constructed using interest income and expense data from the FDIC Quarterly Banking Profile (QBP). Specifically, deposit and loan rates are measured as annualized ratios of trailing four-quarter interest flows to trailing four-quarter average stocks. Because the bank-credit quantity includes loans and securities, the baseline uses the QBP loan yield as the gross return proxy for bank credit and applies provision and demand-markup terms only to the loan share in the pricing equation. Using trailing windows smooths high-frequency variation in flows and ensures comparability with the quarterly balance sheet quantities. The policy rate is measured by the effective federal funds rate, averaged over the same trailing four-quarter window.

Market credit rates are measured using Moody's Aaa corporate bond yields. On the borrower side, nonbank credit is proxied by the stock of outstanding U.S. corporate debt securities from the Financial Accounts of the United States. On the depositor side, money market fund

assets and currency in circulation are used as alternatives to deposits. All quantities are scaled by nominal GDP to create consistent time series.

Regulatory quantities entering the model are constructed from aggregate balance sheet objects. The capital requirement in each regime is set to the minimum observed ratio $E_t/(L_t + R_t)$ less a one-percentage-point buffer. The liquidity requirement in each regime is set to the minimum observed ratio $R_t/W_t$ less the same one-percentage-point buffer. The buffer is therefore measured in ratio points rather than as a common GDP-scaled level of slack. Capital slack and liquidity slack are computed as functions of equity, loans, reserves, core deposits, and wholesale funding, and therefore vary over time with the evolution of bank balance sheets.

### A.3.2 Demand System Details

This appendix provides additional details on the construction of market sizes, market shares, and demand objects used in the calibration.

On the depositor side, the market size in quarter $t$ is defined as

$$W_t^d \equiv D_t + M_t + C_t,$$

where $D_t$ denotes core deposits, $M_t$ denotes money market fund assets, and $C_t$ denotes currency in circulation. Depositor market shares are given by $s_t^d \equiv D_t/W_t^d$, $s_t^M \equiv M_t/W_t^d$, and $s_t^C \equiv C_t/W_t^d$.

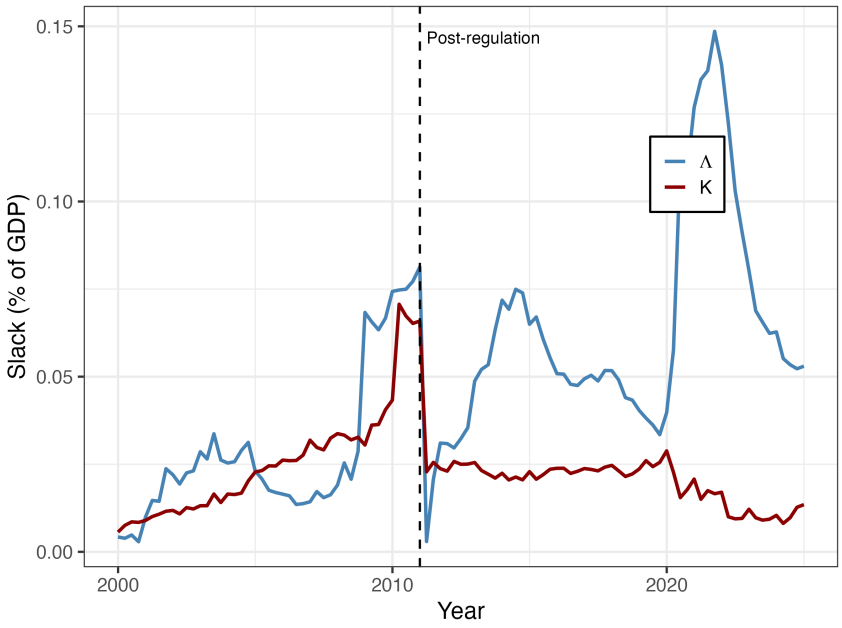
On the borrower side, bank credit is measured by aggregate bank credit $L_t$, and nonbank credit is measured by the stock of outstanding corporate debt securities $L_{nb,t}$. To allow for an outside option of not borrowing, total borrower market size is defined as

$$W_t^b \equiv \eta_b(L_t + L_{nb,t}),$$

where $\eta_b > 1$ pins down the average share of non-borrowers. Borrower market shares are $s_t^L \equiv L_t/W_t^b$ and $s_t^{nb} \equiv L_{nb,t}/W_t^b$, with the no-borrowing share given by $s_t^0 \equiv 1 - s_t^L - s_t^{nb}$. In the baseline calibration we set the no-borrowing share to $s^0 = 0.20$, so $\eta_b = 1/(1 - s^0) = 1.25$.

**Figure A.3**: Model Implied Structural Parameters

The figure plots the implied slacks $(K_t, \Lambda_t)$. The vertical line marks the start of the post-reform regime. Lower values indicate tighter constraints.




### A.3.3 Implied Balance Sheet Slack Variables

Next we provide additional details on the evolution of the capital and liquidity slack variables used in the model. Capital slack is defined as the difference between available equity and required equity under the balance-sheet-implied capital requirement. Liquidity slack is defined as the difference between reserves and wholesale bank liabilities scaled by the balance-sheet-implied liquidity requirement. Both slack variables are functions of balance sheet quantities and therefore vary over time with the evolution of deposits, loans, wholesale bank funding, reserves, and equity. Figure A.3 plots the implied quarterly series for capital slack and liquidity slack over the sample period. The figure illustrates a pronounced tightening of both constraints following the financial crisis and the implementation of post-crisis regulatory reforms. These movements play a central role in shaping pricing behavior and the propagation of deposit shocks in the model.

## A.4 Additional Counterfactual Results

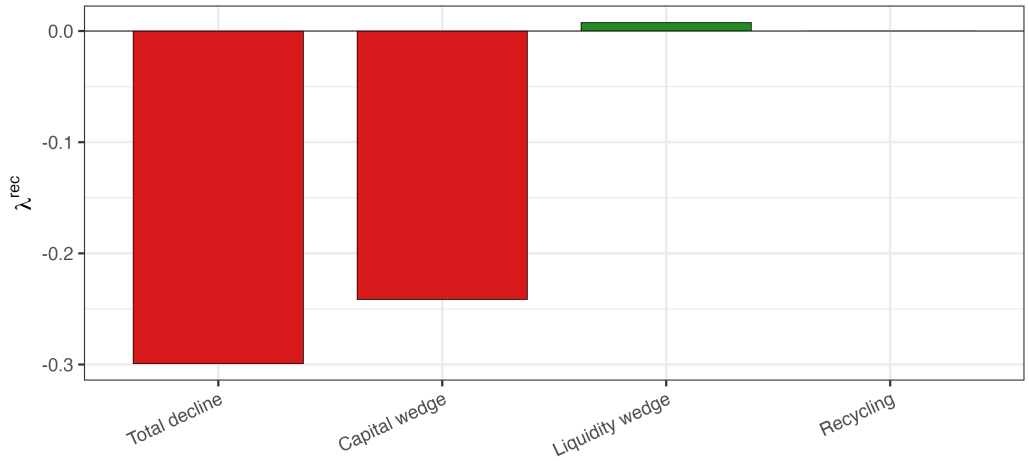
### A.4.1 Decomposition

Panel (a) reports banks' marginal lending response $\lambda$. The first bar reports the total change between the early and late periods. The remaining bars successively move the capital wedge, the liquidity wedge, and deposit recycling to their pre-period historical average values, each building on the last, holding the late-period balance sheet fixed. The capital wedge accounts for most of the decline in $\lambda$ ($0.24$ of the $0.30$ total); the liquidity wedge contributes almost nothing on net, since its historical average magnitude barely differs between the two periods. Recycling has no direct effect on $\lambda$, since it affects amplification conditional on the lending response rather than the lending response itself, and its bar is accordingly exactly zero.

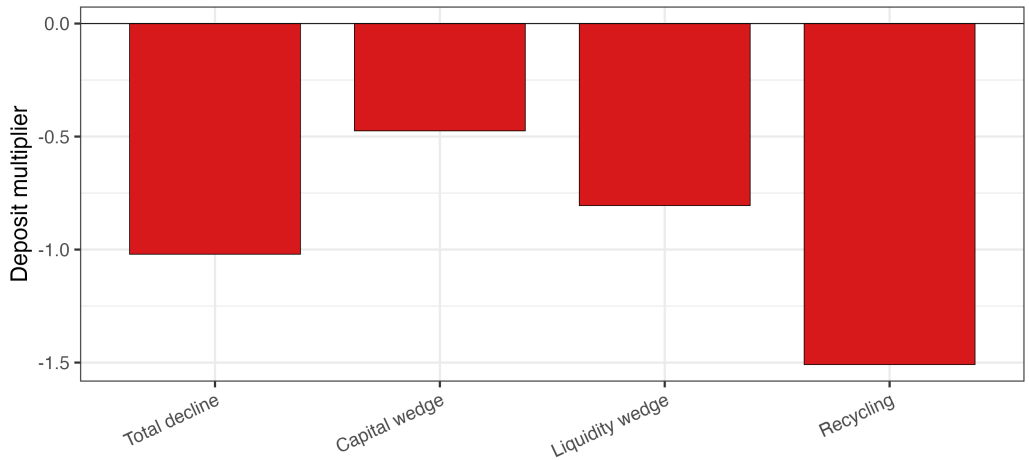
Panel (b) reports the deposit multiplier $m$, which combines the recycling amplification with the reserve-backed first-round response $\lambda^R$. The multiplier falls by $1.02$ in the regime-average comparison. Successively moving the capital wedge, then the liquidity wedge, then recycling to their pre-period values would, on their own, imply a decline of $2.79$, more than twice the actual decline, because $\lambda^R$ is highly sensitive to capital slack near the thin late-period capital position. This is offset by the change in the late-period balance sheet's own composition, which pushes in the opposite direction once the primitives are held fixed at their pre-period values.

**Figure A.4**: Decomposing Decline in Credit Creation

Each panel reports the regime-average change from the early-period baseline to the late-period baseline. The first bar shows the total change. The subsequent bars hold the late-period balance sheet fixed and successively move the capital wedge, the liquidity wedge, and deposit recycling to their pre-period historical average values, each bar building on the previous one.




(a) Banks' Marginal Lending Response $\lambda$




(b) Credit-Deposit Multiplier $m$

### A.4.2 Counterfactual Construction: QE and QT

#### A.4.2.1 Quantitative Easing

We measure QE-created depositor wealth as the fraction of reserve growth attributable to the Fed's asset purchases. For each quarter $t$, we define

$$Q_{QE,t} = \phi \cdot \max(R_t - \bar{R}_{pre}, 0), \tag{57}$$

where $\bar{R}_{pre}$ is the average reserve-to-GDP ratio over 2005–2007, before QE began, and $\phi$ is the share of reserve growth attributed to asset purchases that create money-like wealth for the non-bank public. Our baseline sets $\phi = 1$. We assume QE-created wealth is initially split between deposits and short-term funding markets, with no expansion of cash. Let $\zeta_{QE}$ denote the deposit share of this initial split. The baseline sets $\zeta_{QE} = 0.5$:

$$D_{QE,t} = \zeta_{QE} Q_{QE,t}, \qquad M_{QE,t} = (1 - \zeta_{QE}) Q_{QE,t}, \qquad C_{QE,t} = 0. \tag{58}$$

We construct a counterfactual balance sheet by removing QE-created reserves and the associated depositor-wealth components:

$$\tilde{D}_t = D_t - D_{QE,t}, \qquad \tilde{M}_t = M_t - M_{QE,t}, \qquad \tilde{C}_t = C_t, \qquad \tilde{R}_t = R_t - Q_{QE,t}. \tag{59}$$

We evaluate the counterfactual at the observed 2021Q4 balance sheet, the last quarter before the QT window begins. Given the smaller deposit base $\tilde{D}$, money market fund depth $\tilde{M}$, and reserves $\tilde{R}$, we hold bank and non-bank credit fixed and let wholesale funding adjust residually to satisfy the bank balance sheet. We then evaluate the recycled-deposit response $\tilde{\lambda}_D$ and the reserve-backed first-round response $\tilde{\lambda}^R$ at the counterfactual balance sheet and compute the bank-credit response per dollar of reserve-backed deposit inflow, $\tilde{\lambda}^R/(1 - \theta_{post}\tilde{\lambda}_D)$, the object reported in Figure 6. Holding $\theta_{post}$ fixed in both the observed and no-QE economies isolates the balance-sheet effect of QE on the marginal bank-credit response, rather than adding a separate counterfactual change in the deposit-recycling rate.

#### A.4.2.2 Quantitative Tightening

We measure the total QT volume as the decline in the Fed's combined Treasury and MBS holdings from their peak in 2022Q2 to 2024Q4, expressed as a fraction of GDP. Let this quantity be $X_{QT}$. A pass-through parameter $\omega$ captures the share of the runoff absorbed outside bank balance sheets, for example through non-bank or money market fund adjustment. The bank balance-sheet shock is therefore $(1-\omega)X_{QT}$, and the baseline sets $\omega = 0.5$.

We consider three short-run implementations. In the asset-swap experiment, banks absorb $(1-\omega)X_{QT}$ of the runoff by replacing reserves with less liquid securities. Total bank assets other than bank credit, deposits, wholesale funding, and credit are held fixed, so the capital-relevant balance sheet is unchanged. The liquidity asset entering the liquidity constraint falls by $(1-\omega)X_{QT}$, capturing a tighter liquidity buffer without an immediate decline in deposit claims. In the reserve-drain experiment, liquid assets other than bank credit fall by $(1-\omega)X_{QT}$. Bank credit and deposits are held fixed, and wholesale funding adjusts residually to satisfy the bank balance-sheet identity,

$$W = L + R - E - D.$$

In the reserve-and-deposit-drain experiment, liquid assets other than bank credit and deposits both fall by $(1-\omega)X_{QT}$. Bank credit and non-bank credit are held fixed, and wholesale funding is again determined by the bank balance-sheet identity. Because assets and stable deposit liabilities fall one-for-one, this short-run experiment isolates the effect of lower liquid asset buffers without the Acharya–Rajan wedge between reserves and deposit claims.

### A.4.3 Money-Market-to-Non-bank Spillover Calculation

This appendix records the simple calculation behind the non-bank offset exercise in Section 6. In the baseline figure, non-bank marginal costs are unrelated to money-market depth, so $\kappa_M = 0$. The money-market spillover calculation asks how large $\kappa_M$ would have to be to restore the late-period aggregate credit response to its early-period level.

Let $C_{pre}$ and $C_{post}$ denote the baseline total-credit responses to a one percent of GDP deposit inflow in the early and late periods, combining the bank-credit response with non-bank borrower substitution. In the current calibration, $C_{pre} = 1.14$ and $C_{post} = 0.50$, so the aggregate

gap to be closed is

$$\Delta C = C_{pre} - C_{post} = 0.64.$$

For reference, the corresponding bank-credit responses are $B_{pre} = 2.13$ and $B_{post} = 0.88$. The aggregate gap is about half the bank-credit decline because non-bank crowding-out through borrower substitution was larger in the early period, when the bank response itself was stronger.

Let a share $\omega_M$ of loan proceeds that do not recycle into stable deposits flow into money market funds. Then money-market depth is evaluated at the selected no-spillover total-credit response:

$$dM^{sel} = \omega_M(1 - \theta_{post})dL^0_{\text{total},post}.$$

Here $dL^0_{\text{total},post}$ is the no-spillover total-credit response, including the deposit-recycling multiplier and baseline non-bank borrower substitution. Using (48), define $\Delta L^{spill}_{\text{total},post}(\kappa_M)$ as the selected-allocation total-credit response under a money-market spillover. The terms $\lambda_{L,M,post}(\kappa_M)$ and $\lambda^{cost}_{nb,M,post}(\kappa_M)$ are both linear in $\kappa_M$ under the parameterization $\mu_{nb}(M) = \bar{\mu}_{nb} - \kappa_M M$, and the spillover increment is linear in the leakage share $\omega_M$. The offset loading is the value that solves

$$\Delta L^{spill}_{\text{total},post}(\kappa_M^{offset}) - \Delta L^{spill}_{\text{total},post}(0) = \Delta C.$$

Under full leakage, $\omega_M = 1$, the required loading is $\kappa_M^{offset} = 2.21$. Under half leakage, $\omega_M = 0.5$, the required loading doubles to $\kappa_M^{offset} = 4.42$.