Who Bears the Burden of Bank Capital Regulation?

Buffers, Risk Weights, and the Supply of Credit

Juan Zurita*

This draft: October 23, 2026

Job Market Paper

Click here for the latest version

Abstract

*Who bears the burden of bank capital regulation? Risk weights determine which loans contract; voluntary capital buffers determine which banks contract most; and the statutory size tiers regulators actually set identify the exposed banks poorly. Using narratively identified capital-requirement shocks and a quarterly panel of U.S. banks (1994–2019), we document the first two facts directly: commercial lending contracts about a third more than real estate lending within banks (*1.31*), and banks with the least room above their own requirement cut C&I lending* 1.32 *times more than banks with the most. The buffer gradient extends well beyond the two percent of bank-quarters at the regulatory frontier—which a capital constraint alone cannot produce—and so identifies a cost of regulatory capital that is not risk-weighted. A heterogeneous macro-banking model built on that mechanism reproduces the gradient as an* untargeted *moment, 1.32 against 1.32 in the data. The statute sorts banks by size; the mechanism sorts them by buffer, and the two are close to orthogonal: statutory tier explains* 1.8% *of the cross-sectional variance in regulatory slack, and* 86% *of thin-buffer banks face the* lowest *statutory requirement. Risk weights and size tiers redistribute the burden of a tightening but barely change its aggregate magnitude—a gross transition cost excluding the stability benefits of capital.*

JEL classification: E44, E58, G21, G28.

*School of Economics, University of Edinburgh. Email: juan.zurita@ed.ac.uk. This paper has greatly benefited from discussions with Corrado Di Guilmi, Mengheng Li, Peter Docherty, Jorge Miranda Pinto, Mario Fiorini, Adeline Delavande, Peter Siminski, David Goldbaum, Timothy Kam, Mei Dong, Shawn Chen, James Graham, Rustam Jamilov, and Marlon Azinovic-Yang, as well as participants at the Econometric Society Australasian Meeting 2023, the 2022 Financial Markets and Corporate Governance at Monash University and the WEAI Conference Melbourne 2023. I am grateful to Urfat Isayev for excellent research assistance. Any remaining errors are my own responsibility.

Keywords: Capital Requirements, Capital Buffers, Risk Weights, Credit Supply, Banking Heterogeneity, Macroprudential Policy

1 Introduction

Since the global financial crisis, bank capital requirements have risen substantially, and the debate over how far they should rise has been conducted in aggregates: the financial stability benefits of a higher requirement weighed against the output cost of getting there. The regime that resulted, however, is not an aggregate lever. It is deliberately differentiated—business loans carry higher regulatory risk weights than residential mortgages, and statutory requirements step up at the $10 billion and $50 billion asset thresholds. Where those thresholds sit has been revised repeatedly—raised in 2018, recast as supervisory categories in 2019, revisited after the regional-bank failures of 2023, and under revision again in 2026—but every revision has redrawn the line on the same dimension, asset size. Both margins presume that regulators know where the burden of capital regulation falls. This paper asks who in fact bears it—a question about the transmission of macroprudential policy whose answer determines both the sectoral composition of any credit contraction and the aggregate cost of the regime. Risk weights determine which loans contract; voluntary capital buffers determine which banks contract most; and the statutory size tier identifies neither. The first two appear on every bank's balance sheet and in no statute; the third, the only margin regulators directly set, explains under 2% of the cross-sectional variance in regulatory slack, so the margin available to policy and the margin that determines incidence are close to orthogonal. The aggregate cost of a tightening, by contrast, is governed almost entirely by the level of requirements rather than by their design.

We document the first two directly, and state them before the design that delivers them. Within banks, commercial and industrial (C&I) lending—which carries the higher regulatory risk weight—contracts about a third more than real estate lending after a tightening (a ratio of 1.31, s.e. 0.07), with hump-shaped responses that trough three to four years out; consumption and credit-card lending do not respond at any horizon. Across banks, transmission is governed by the *distance to the regulatory frontier*: sorting institutions each quarter into terciles of their lagged regulatory slack, banks in the thinnest-buffer tercile cut C&I lending 1.32 times more than banks in the thickest. The gradient survives within size classes, and it *reverses* when tercile membership is frozen at each bank's pre-sample average—what matters is the buffer a bank holds when the shock arrives, not a fixed bank type. Both facts are stated as ratios, so neither depends on the scale of the underlying innovation.

We establish both facts in two steps. Empirically, we assemble a quarterly panel of roughly 4,800 U.S. banks from the Call Reports (1994–2019) and trace the narratively identified capital-requirement shocks

of Drechsel & Miura (2025) through the cross-section of lending using local projections. Because the identified innovation is an aggregate surprise, our regressions exploit within-bank variation: bank fixed effects and macroeconomic controls absorb the common cycle, lending shows no pre-trends ahead of the identified innovations, and consumer credit—fully risk-weighted but funded against high interest margins—serves as a placebo throughout. Structurally, we develop a heterogeneous macro-banking model in which banks facing funding-market frictions hoard precautionary capital buffers, so the cross-sectional distribution of regulatory slack is an equilibrium object rather than a primitive. The regulatory constraint binds only occasionally and policy functions are kinked at the frontier, so we solve the model globally with deep-learning methods and estimate the requirement process by simulated method of moments on the sixteen within-bank impulse-response moments. Scaled through the estimated model, one unit of the identified innovation moves required capital by about four basis points, and the innovations are effectively permanent, consistent with post-crisis regulation being a sequence of regime changes rather than transitory shocks.

The economics behind both facts is a single mechanism. A bank that cannot cheaply replace funds at the margin—because its deposit franchise is exhausted and wholesale markets penalise it—self-insures against regulatory breach by holding equity above its requirement, so the cross-sectional distribution of buffers mirrors the distribution of funding frictions. When a tightening arrives, the regulatory constraint binds only for banks whose buffer it exhausts: for them the shadow cost of capital spikes, and the cheapest source of regulatory relief is to shed the assets that carry the highest risk weight—C&I loans—because each dollar cut frees the most required capital. Transmission is therefore state-dependent across banks and asymmetric across sectors for the same underlying reason: the price of regulatory capital differs across balance sheets and across asset classes.

The estimated model reproduces the sign of all sixteen targeted moments and the sectoral ordering at every horizon, though this fit—like the impulse-response magnitudes though not the welfare level—is sensitive to which solved policy represents the model; the buffer gradient reported below is computed from the model's static solution, which is exact in each regime (Online Appendix OA.G). Its transmission dynamics differ in shape: the simulated response is largest on impact and then deepens without reverting, whereas the empirical responses build to a trough at three to four years and return toward zero. More demandingly, it delivers a low-to-high buffer-tercile response ratio of 1.32 against 1.32 in the data—an *untargeted* moment that validates the buffer mechanism at the model's core.

The estimated model then serves as a laboratory for regulatory experiments that the historical record

cannot identify. Converted to a permanent one-percentage-point increase in required capital, lending contracts by 3.5% in C&I, 3.0% in real estate, and 3.1% in total—inside the range of quasi-experimental estimates from UK, euro-area, and Belgian settings. The welfare cost of that tightening is 1.4% of steady-state credit-market surplus—the sum of borrower and intermediary surplus in the loan market—and essentially all of it flows through the contraction in the scale of intermediation. The credit magnitudes carry the sensitivity noted above; the welfare level does not.

Counterfactuals that flatten risk weights eliminate the sectoral asymmetry between C&I and real estate lending, while replacing the size-tiered schedule with a uniform requirement raising the same aggregate capital reshuffles the burden across banks by buffer position; in both cases the aggregate cost is nearly unchanged. The model also speaks to the question in the title, and the answer is about targeting rather than magnitude. Statutory size tiers identify the exposed population poorly: tier explains under 2% of the cross-sectional variance in regulatory slack, and 86% of thin-buffer banks face the *lowest* statutory requirement (Section 5.3). Sorting banks by distance to the frontier instead, thin-buffer institutions contract lending about 13% more per dollar lent than well-buffered ones; because they are smaller, their share of the aggregate contraction (34%) exceeds their share of lending (32%) only mildly, so the burden is disproportionate per unit of credit rather than concentrated on a small group. Finally, transmission is non-linear in the state: the buffers that insulated most banks against the moderate innovations of 1994–2019 are a finite resource, and shocks large enough to erode them trigger contractions disproportionate to the shock.

The policy content is immediate. Debates over the architecture of capital requirements—risk weights, size tiers—are, to first order, debates about incidence rather than aggregate cost. Supervisory attention to the distribution of voluntary buffers identifies where a tightening will bind more accurately than statutory size categories do, and the same logic implies that buffer releases in downturns are most potent for institutions operating near the frontier. Because the framework is partial equilibrium and prices no stability benefits of capital, our welfare figures are gross transition costs—excluding borrower substitution to non-bank intermediaries—informative about the burden of tightening and its distribution, not about optimal requirement levels.

Our results connect three literatures. First and closest, a quantitative macro-banking literature studies capital regulation in economies with dynamic, heterogeneous banks: Corbae & Erasmo (2021), Begenau (2020), Elenev *et al.* (2021), and Mendicino *et al.* (2020). These papers characterise optimal requirements in general equilibrium; our contribution is to discipline the cross-section directly—the model is estimated

on within-bank impulse responses to identified policy innovations and validated on an untargeted cross-sectional moment—which is what allows us to speak to the *incidence* of a given tightening, at the cost of abstracting from default and its stability benefits. The difference is one of object rather than of ambition: where that literature asks what the aggregate cost of a given requirement level is, we ask how a given tightening distributes, and answering it requires carrying the distribution of regulatory slack as a state rather than reducing the banking sector to a small number of types.[1] Relative to Corbae & Erasmo (2021), who model a single loan sector in a two-bank-type economy, we introduce two sectors carrying distinct risk weights (mortgages at $\omega_M = 0.50$ versus C&I at $\omega_B = 1.00$) so that the model endogenously generates within-bank asset reallocation, estimate the requirement process from narratively identified policy innovations rather than holding the requirement fixed and evaluating tightenings as counterfactual regimes, and solve the model globally with neural-network function approximation, which allows us to preserve the full cross-sectional distribution of buffers that drives incidence.[2] Second, a quasi-experimental literature measures the credit-supply effects of requirement changes: Aiyar *et al.* (2016) for UK bank-specific requirements, Gropp *et al.* (2020) for the EBA capital exercise, De Jonghe *et al.* (2020a) for Pillar-2 decisions, Jimenez *et al.* (2017) for dynamic provisioning, and Mésonnier & Monk (2015) and Conti *et al.* (2023) at the aggregate level; Acharya *et al.* (2022) provide a particularly granular anatomy of macroprudential transmission using Irish mortgage-level data. Relative to this work, we hold the policy event fixed and trace how the same identified innovation is absorbed across asset classes and across the buffer distribution, and we provide the structural mapping from those responses to aggregate welfare. Third, our funding-capacity evidence connects to the deposits-channel literature of Drechsler *et al.* (2017): banks with strong deposit franchises face little funding risk, and we show that they optimally hold thinner regulatory buffers, so the deposit franchise shapes not only monetary transmission but also the incidence of capital regulation. Methodologically, the global deep-learning solution follows Maliar *et al.* (2021).

[1] The contrast is sharpest with Elenev *et al.* (2021), whose intermediary sector is summarised by a single net-worth state. Idiosyncratic failure risk there is i.i.d. across intermediaries and failed institutions are recapitalised within the period, so the cross-section does not persist across periods and enters the aggregate conditions only through a survival share and the corresponding truncated mean of the shock distribution. The incidence of a requirement change is then a sector-level object by construction. That representation is precisely what makes their general-equilibrium block tractable, and we do not improve on it: we give up general equilibrium in order to retain a persistent buffer distribution, which is the state on which our cross-sectional moment is measured.

[2] The contrast is sharpest in how each paper disciplines its cross-section. Corbae & Erasmo (2021) target the buffer distribution at two group means—the average risk-weighted capital ratio of the largest banks and of the fringe—and validate transmission on a size-sorted lending regression driven by deposit-rate variation. The gradient reported above is untargeted, is sorted on regulatory slack rather than size, and is identified off requirement innovations; because it survives within size classes and reverses when tercile membership is frozen at pre-sample averages, the two sorts are not interchangeable.

The remainder of the paper proceeds as follows. Section 2 presents the data, the identification strategy, and the two cross-sectional facts. Section 3 develops the stylised model and Section 4 the quantitative framework. Section 5 reports the estimation, fit, and counterfactuals, and Section 6 draws the policy implications. Section 7 concludes.

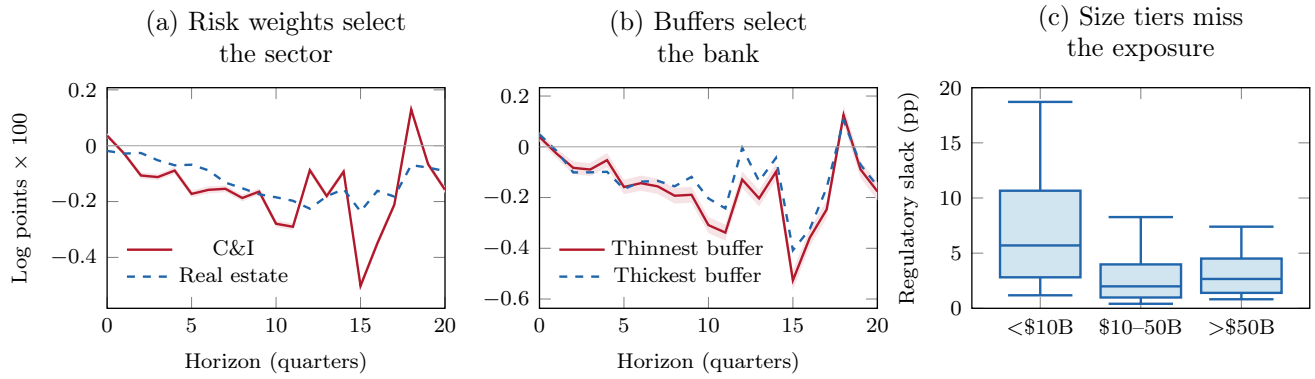


FIGURE 1.—Risk weights select the sector; buffers select the bank; size tiers miss the exposure

Notes: Panel (a): response of C&I and real estate lending to a one-unit capital-requirement innovation, within-bank local projections on the community cohort; averaged over $h = 1$–20 the C&I response is 1.28 times the real estate response. Panel (b): the C&I response re-estimated within terciles of lagged regulatory slack; averaged over the same horizons the thinnest-buffer tercile contracts 1.29 times as much as the thickest, and the thick tercile contracts significantly—which a capital constraint alone cannot produce, since fewer than three percent of bank-quarters sit at or inside the frontier. Panel (c): distribution of regulatory slack by statutory size cohort, box at p25–p75 with whiskers at p10 and p90; the tiers overlap heavily and statutory tier explains 1.8% of the cross-sectional variance in slack. Shaded bands are confidence intervals for the C&I and thinnest-tercile series.

2 Empirical Evidence

The mechanism this paper studies is balance-sheet-driven. Unlike monetary policy, which influences credit conditions by shifting the base cost of funding, changes in capital requirements alter the quantity of equity required to support a marginal unit of risk-weighted assets, creating a direct link between a bank's capital position and its shadow cost of lending (Van den Heuvel *et al.*, 2002). A tightening forces a bank into a three-margin adjustment: raising new equity, shrinking the asset portfolio, or reallocating it toward lower-risk-weight assets. Because business loans carry higher regulatory weights than mortgages, the last margin makes high-weight sectors the efficient place to seek regulatory relief per dollar of credit curtailed (De Jonghe *et al.*, 2020a; Sivec & Volk, 2021; De Jonghe *et al.*, 2020b)—and because the ability to use the first two margins depends on a bank's funding structure, the incidence of a given tightening depends on the cross-sectional distribution of funding frictions and voluntary buffers. Isolating this balance-sheet

transmission from broader monetary cycles is the task of the empirical design below; formalising it is the task of the model in Section 3.

2.1 Data and Stylized Facts

Our primary data source is the Consolidated Reports of Condition and Income (Call Reports) provided by the FDIC. The sample spans from 1994 to 2019 and includes quarterly balance sheet data for approximately 4,800 U.S. commercial banks—effectively the *population* of FDIC-insured institutions, not a sample drawn from it. This granular dataset allows us to decompose total loans and leases into specific categories: real estate, commercial and industrial (C&I), consumption, and credit card loans. We complement these micro-data with macroeconomic indicators from the Federal Reserve Bank of St. Louis (FRED), specifically real GDP growth and the effective federal funds rate.

When two institutions merge, we retain only the surviving entity's post-merger balance sheet; the acquired bank's final pre-merger observation is dropped so that the mechanical jump in the survivor's assets does not contaminate growth rates. All bank-level variables entering regressions are winsorised at the 1st and 99th percentiles of their quarterly cross-sectional distributions, which further attenuates any residual discontinuities from mergers. As a robustness check, we re-estimate the headline local projections after excluding all bank-quarters within one year of a completed acquisition; the point estimates are virtually unchanged (Online Appendix).

To analyze the transmission of policy across the banking hierarchy, we partition our sample into three cohorts based on total asset size, following the regulatory thresholds established by the Dodd-Frank Act (2010) and the EGRRCPA (2018). We define Small (Community) Banks as those with assets below $10 billion, Medium (Regional) Banks as those between $10 billion and $50 billion, and Large (Systemically Important) Banks as those exceeding $50 billion. These thresholds are not merely statistical conveniences; they reflect the points at which banks become subject to increasingly stringent stress testing, enhanced prudential standards, and closer capital oversight by the Federal Reserve.

For the time-series analysis, we aggregate individual bank-level data to these categories by summing the levels of loans and equity. For regulatory ratios, such as the total risk-based capital ratio, we compute weighted averages where each institution's share of total equity serves as the weight. This ensures that our aggregate measures reflect the systemic capitalization of each sector and that the behavior of systemically important institutions is accurately captured, rather than being obscured by an unweighted average of

thousands of smaller entities.

Table I provides summary statistics for our main variables. The high standard deviation in Total Assets ($1.4 billion mean vs. $29.5 billion SD) underscores the extreme heterogeneity in the U.S. banking sector, necessitating our size-based cross-sectional analysis.

Central to the mechanism this paper studies is the role of asset composition and risk-weighting: the "regulatory wedge" created by risk-weighted assets (RWA). Acharya *et al.* (2014) document that regulatory risk weights can diverge substantially from market-implied measures of risk, making the mapping from balance-sheet composition to effective capital charges a first-order determinant of bank behaviour. Figure 2 shows that the RWA-to-total-asset ratio rises systematically with bank equity—larger banks hold portfolios with higher average risk weights, while small banks' regulatory position is driven by portfolio composition (Le Leslé & Avramova, 2012). Following the Global Financial Crisis this wedge has widened, with smaller banks increasingly concentrating in real estate while larger banks maintain diversified exposures. The stylized facts below unpack the components of this wedge.

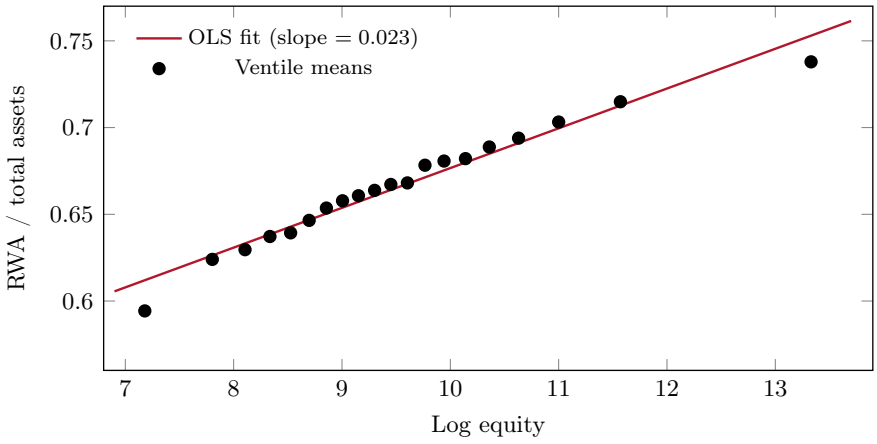


FIGURE 2.—The role of weight effects

Notes: Binned scatter: each point is the mean risk-weighted-assets-to-total-assets ratio within a ventile of log equity (899,168 bank-quarter observations, 1994–2020); the line is the OLS fit on the underlying bank-quarter data (slope 0.023). The positive slope is the "regulatory wedge": larger banks hold portfolios with higher average risk weights, while small banks' regulatory position is driven by portfolio composition. Source: US Call Reports.

Stylized Fact 1: Market Concentration. Figure 3 illustrates a significant structural shift in the U.S. credit market. While the number of institutions has declined, market participation has become increasingly concentrated among large banks. This suggests that the aggregate response to capital requirements is increasingly dominated by the behavior of a few systemically important lenders.

Figure 3 reveals a crucial link between bank size and asset composition. Small banks exhibit high

TABLE I
SUMMARY STATISTICS (1994–2019)

Variable	Source	N	Mean	Std. Dev.
Panel A: Balance Sheet (Millions $)				
Total Assets	Call Reports	996 754	1446.19	29 500.00
Total Liabilities	Call Reports	996 754	1320.56	27 000.00
Total Equity	Call Reports	994 569	151.18	2949.80
Panel B: Regulatory Ratios and Asset Shares				
Capital / RWA (%)	Call Reports	948 614	22.61	243.41
Mortgage Loan Share (0-1)	Own Calc.	996 754	0.69	0.23
Business Loans Share (0-1)	Own Calc.	996 754	0.14	0.18

loan-type concentration, with mortgage lending constituting over 90% of their total portfolios. Conversely, large banks maintain significantly more diversified balance sheets, with real estate representing only 40% of their lending. As mortgages and business loans carry different regulatory risk weights (ω), this specialization implies that macroprudential shocks will have inherently asymmetric effects across the banking distribution. Banks with specialized portfolios have fewer margins of adjustment when capital constraints bind, a feature we explore in our empirical identification and quantitative model.

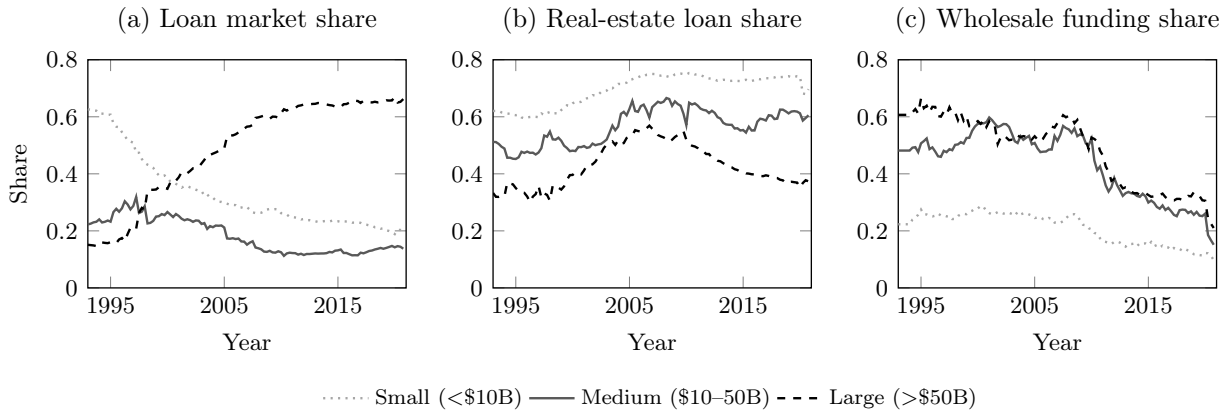


FIGURE 3.— Market concentration, asset specialization and funding structure

Notes: Panel (a) shows each size cohort's share of total bank loans; panel (b) the real-estate share of each cohort's loan portfolio; panel (c) wholesale funding as a share of non-equity funding by cohort. Cohorts split at $10B and $50B of lagged total assets. Quarterly, 1993–2020. Source: US Call Reports.

Stylized Fact 2: Asset Specialization. The importance of this sectoral divide is further illuminated by the structural divergence in bank portfolios over the last three decades. Figure 4 illustrates the "Asset Gap"—defined as the difference between real estate and business credit as a proportion of total lending—across the banking hierarchy.

Since the mid-1990s, a trend that accelerated sharply following the Global Financial Crisis, small and medium-sized banks have experienced a persistent widening of this gap. This indicates an increasing concentration in real estate at the expense of commercial and industrial (C&I) lending. In contrast, large banks have maintained a significantly narrower gap, reflecting more diversified balance sheets and a broader range of lending activities.

From a regulatory standpoint, this specialization is a critical determinant of policy transmission. Because real estate and business loans carry distinct risk weights (ω_{RE} vs. ω_{BIZ}), a bank's capacity to buffer a capital shock depends heavily on its portfolio flexibility and funding capacity. While a diversified large institution can pivot between asset classes to optimize its risk-weighted assets (RWA), specialized small banks face narrower margins of adjustment when capital constraints bind. This empirical specialization provides the foundational rationale for the heterogeneous "sectoral lending channel" that we explore in both our empirical identification and quantitative model.

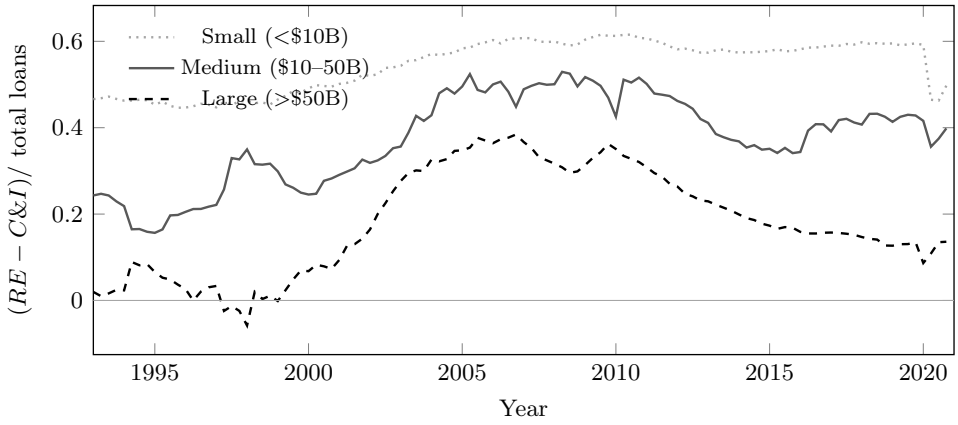


FIGURE 4.— Difference between real estate and business credit

Notes: This figure shows the difference between real estate and business credit as a proportion of total lending for small, medium and large banks, 1993–2020. Cohorts split at $10B/$50B of lagged total assets. Source: US Call Reports.

Stylized Fact 3: The Regulatory Wedge and Capital Volatility. The distinction between a bank's accounting leverage and its regulatory capital position is central to understanding the impact of macroprudential shocks. Figure 5 compares the equity-to-asset ratio (the leverage ratio, panel a) with the equity-to-risk-weighted asset ratio (the regulatory capital ratio, panel b) for the three bank cohorts.

Two key observations emerge from this comparison. First, while small banks often appear more highly capitalized than large banks on an unweighted basis (panel a), this advantage is compressed when accounting for risk-weighting (panel b). This confirms that the regulatory burden is not uniformly

distributed: the transition from total assets to risk-weighted assets (RWA) creates a "regulatory wedge" that is most binding for institutions with specialized, high-risk-weight portfolios.

Second, the equity-to-RWA ratio exhibits significantly higher volatility than the simple leverage ratio. This discrepancy underscores the significance of the risk-weighting mechanism (ω) as a driver of capital-requirement shocks. Because the denominator of the regulatory ratio is sensitive to shifts in both asset composition and regulatory risk-classification, banks must manage their portfolios dynamically to maintain compliance. The higher sensitivity of the weighted ratio suggests that banks with smaller equity or less diversified portfolios have fewer degrees of freedom to absorb unanticipated changes in capital requirements, a mechanism we formalize in our theoretical framework.

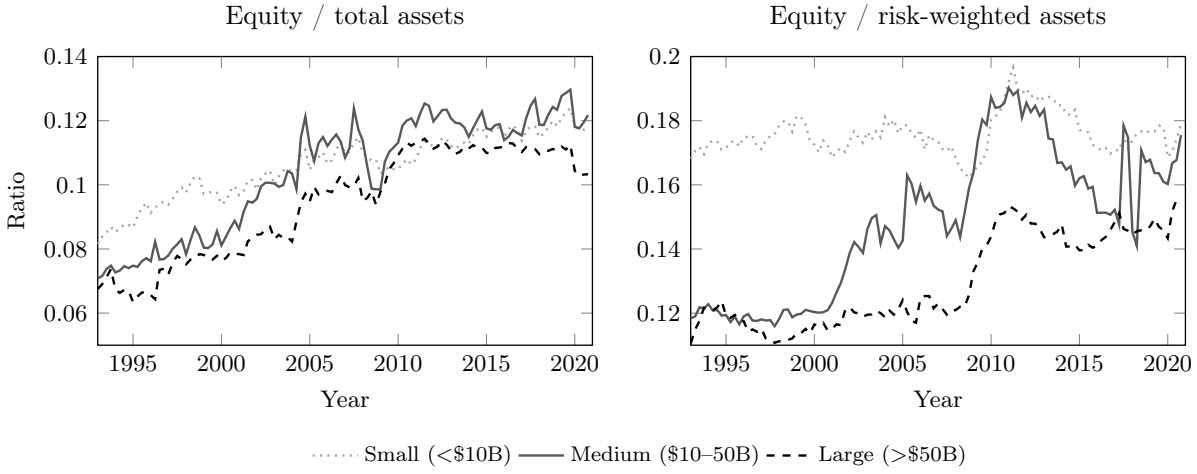


FIGURE 5.—Leverage and regulatory capital ratios by bank size

Notes: Panel (a) plots the equity-to-total-asset ratio (leverage ratio) and panel (b) the equity-to-risk-weighted-asset ratio (equity-weighted within cohort) for small, medium and large banks, 1993–2020. Source: US Call Reports.

2.2 The Evolution of Capital Regulation in the United States

Capital regulation in the United States has undergone a structural transformation over the past four decades, shifting from a microprudential focus on individual bank solvency to a macroprudential framework centered on systemic stability. This evolution provides a rich source of variation for identifying the impact of regulatory shocks on the lending channel.

The modern era of U.S. capital regulation began with the 1992 implementation of the Basel I Accord, which established uniform minimum capital ratios based on credit-risk-weighted assets. However, Basel I's coarse risk buckets eventually proved insufficient for capturing the complexities of off-balance-sheet

exposures and market risk. While the Basel II framework attempted to introduce more granular, internal ratings-based (IRB) sensitivities in the mid-2000s, its implementation was interrupted by the 2008 financial crisis. The crisis revealed fundamental fragilities in bank capital structures-namely, an over-reliance on short-term wholesale funding and insufficient high-quality loss-absorbing capital.

In the wake of the crisis, the regulatory landscape shifted toward the *quantity* and *quality* of capital. The Dodd-Frank Act (2010) and the subsequent U.S. implementation of Basel III (2013) introduced the Common Equity Tier 1 (CET1) requirements, capital conservation buffers, and a supplementary leverage ratio. Most importantly for our study, these reforms introduced "tailored" regulation: systemically important financial institutions (SIFIs) became subject to annual stress tests and higher capital surcharges, while the Economic Growth, Regulatory Relief, and Consumer Protection Act (EGRRCPA, 2018) later eased the compliance burden for community banks.

Table II summarizes the major regulatory events used in our empirical analysis. Following the identification strategy of Drechsel & Miura (2025), we focus on the "first mention" of these legislative acts in official speeches rather than their final enactment dates. This approach captures the pre-emptive behavior of banks as they adjust their balance sheets in anticipation of new requirements. As shown in the final column of Table II, banks exhibit significant equity capital accumulation between the initial announcement and the formal passing of the acts. For instance, equity capital grew by an average of 39% following the first mentions of the Riegle-Neal Act and by 27% surrounding the Dodd-Frank Act.

This pre-emptive accumulation of equity capital is central to our mechanism. It suggests that regulatory announcements trigger a significant balance sheet reorganization long before the legal enforcement date. In the following sections, we utilize these announcement-day surprises to identify the causal impact of capital requirement shifts on the cross-section of bank lending.

TABLE II
FINANCIAL REGULATION AND EQUITY CAPITAL ACCUMULATION

Legislative Act	**Passed**	**1st Mention**	**Mean Δ Equity**
Riegle-Neal Interstate Banking Act	09/13/1994	09/15/1992	+39%
Gramm-Leach-Bliley Act	11/12/1999	04/19/1998	+26%
Dodd-Frank Wall Street Reform Act	07/21/2010	07/24/2008	+27%
EGRRCPA (Regulatory Relief Act)	05/24/2018	09/28/2016	+7%

Notes: This table reports key U.S. legislative acts and the dates of the speeches in which they were first publicly introduced, following Drechsel & Miura (2025). Mean Δ Equity represents the average growth in bank equity between the first mention and enactment.

2.3 The Anatomy of the Mechanism: Funding Capacity and Regulatory Slack

While the macroeconomic history in the previous section establishes that regulatory announcements induce equity accumulation, the cross-sectional capacity to respond to these shifts depends fundamentally on a bank's liability structure. To open the black box of this mechanism before estimating our causal macroeconomic models, we utilize our quarterly bank-level data to examine the joint distribution of regulatory cushions, funding alternatives, and dynamic adjustment choices.

Stylized Fact 4: The Funding-Buffer Trade-Off. Our theoretical framework posits that banks hold voluntary capital buffers as a precautionary cushion against regulatory breaches. Institutions with deep, flexible funding sources can afford to hold minimal excess capital because they can seamlessly raise external funding or issue equity under stress. Conversely, banks facing significant funding market frictions must hoard larger precautionary capital buffers to self-insure.

To document this in the data, we define a bank's continuous *Funding Capacity* as the share of total liabilities funded by core retail deposits and non-interest-bearing (NIB) deposits; its complement is the share of wholesale funding and other managed liabilities.[3] We define *Regulatory Slack* (D_{it}) as the distance between a bank's total risk-based capital ratio and the prevailing regulatory minimum requirement ($r_{it} - \kappa_t$), measured in percentage points. We formalize this relationship using the following descriptive panel specification:

$$D_{it} = \alpha_i + \gamma_t + \beta_1 \, Funding \; Capacity_{it} + \beta_2 \; \log(\text{Assets}_{it}) + \varepsilon_{it} \tag{1}$$

where γ_t are quarter fixed effects and α_i are bank fixed effects, included in the most demanding variant. Because a bank's funding model is a slow-moving structural characteristic, the two sets of fixed effects answer different questions: without α_i, β_1 compares banks with different liability structures; with α_i, it is identified only from within-bank changes in the funding mix.

Table III reports both, together with a pure between-banks specification estimated on bank-level time

[3]The economic rationale behind this metric follows the deposit-franchise view of bank funding: core retail and non-interest-bearing deposits are sticky and withdrawal-insensitive, so a high funding-capacity ratio indicates a liability structure that is largely insulated from funding-market stress. When subjected to a capital or macroprudential shock, such institutions face little rollover risk and can absorb the adjustment without triggering escalating marginal funding penalties. Conversely, a low ratio indicates reliance on wholesale and other managed liabilities, which are rate-sensitive and must be rolled over in markets whose access can deteriorate precisely when capital is scarce. Facing this funding risk—and a deeply costly recourse to friction-laden equity markets—such institutions optimally self-insure by maintaining larger precautionary capital buffers ex-ante. When those buffers erode, the liability-side rigidity forces the adjustment onto the asset side, leaving their aggregate lending supply highly sensitive to unexpected contractions.

averages. The funding-buffer trade-off is a sharp *cross-sectional* regularity: in the pooled and between specifications, β_1 is negative and highly significant—a one-standard-deviation increase in the core-deposit share of liabilities (≈ 0.10) is associated with 2.0–2.6 percentage points less regulatory slack, roughly 40% of the median buffer of 5.7 percentage points. Banks whose liabilities are dominated by sticky retail and non-interest-bearing deposits—a stable funding franchise that provides insurance in itself—operate with structurally thinner voluntary buffers, while banks exposed to wholesale funding and its rollover risk self-insure by carrying larger cushions. Within banks, the gradient is far smaller ($\beta_1 = -2.3$), precisely because the funding mix barely moves at the quarterly frequency: the buffer is an equilibrium object shaped by the bank's structural funding model, not a margin adjusted quarter to quarter—the interpretation our model formalises. The size gradient, by contrast, is robustly negative in every specification: larger institutions, with diversified funding and equity-market access, allow their regulatory slack to compress toward the frontier.

Table III
Distance to Constraint: Panel Estimates

Dependent Variable: *Regulatory Slack (pp)*	**(1)** Pooled	**(2)** Between banks	**(3)** Within
Funding Capacity	-19.44***	-25.97***	-2.27**
	(1.73)	(1.95)	(1.05)
Log Assets	-2.80***	-2.70***	-6.27***
	(0.09)	(0.10)	(0.16)
Quarter fixed effects	Yes	—	Yes
Bank fixed effects	No	—	Yes
Observations	935,234	16,117	935,056
R^2	0.104	0.124	0.693

Notes: Estimates of equation (1). The dependent variable is regulatory slack in percentage points, winsorised at the 1st/99th percentiles. Funding capacity is the core-plus-NIB deposit share of total liabilities. Column (2) is estimated on bank-level time averages (one observation per bank) with robust standard errors; columns (1) and (3) cluster standard errors by bank. Sample: quarterly bank panel, 1993Q2–2019Q4. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

This systematic variation in buffer hoarding constitutes the critical pre-condition driving the asymmetric aggregate transmission of macroprudential policy. Consequently, the aggregate credit elasticity of a policy innovation is determined not merely by its nominal magnitude, but by whether the innovation is sufficiently large to bind the endogenous constraint of the marginal institution. Highly capitalized institutions operating deep within their respective regulatory frontiers absorb the policy shock via liability-side reconfigurations, leaving credit supply unaltered. Conversely, institutions operating in close proximity to the regulatory frontier-predominantly small and mid-tier lenders bound by localized

retail deposit constraints-experience binding asset-side adjustments. Ultimately, it is this cross-sectional heterogeneity in the distance-to-constraint distribution, rather than a uniform balance-sheet exposure, that generates the size-dependent and sector-specific credit supply contractions documented in Section 2.4. In the quantitative framework of Section 4, funding capacity maps inversely into the funding-market friction that generates buffers endogenously: banks facing steeper marginal costs of replacing funds hold more precautionary equity, so the empirical funding-capacity gradient is the cross-sectional fingerprint of the same mechanism the model formalises.

Stylized Fact 5: Buffer Heterogeneity Cuts Across Size Tiers. Figure 6 displays the distribution of regulatory slack within each of the three statutory size cohorts. Small banks—the most numerous group—span the full range of buffer positions, from near-zero slack to buffers exceeding 20 percentage points; medium and large institutions cluster closer to the regulatory frontier but still exhibit meaningful dispersion. The key implication is that statutory size is a poor proxy for regulatory tightness: many small banks operate with ample buffers and are effectively insulated from moderate capital shocks, while some medium-tier institutions sit close to the constraint. This within-tier heterogeneity motivates our empirical strategy of conditioning on the buffer distribution directly, rather than on asset-size thresholds alone.

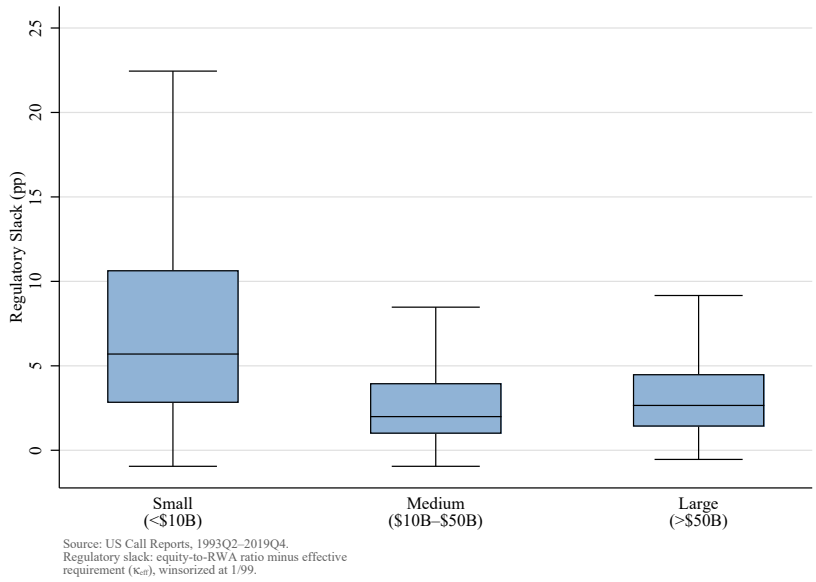


FIGURE 6.—Regulatory slack spans 20+ pp among small banks but compresses near the frontier for larger institutions

Notes: Each box plot displays the interquartile range, median, and 10th/90th percentile whiskers of regulatory slack ($D_{it} = r_{it} - \kappa_t$, in percentage points) within the small (<$10B; ≈880,000 bank-quarters), medium ($10B–$50B; ≈3,200 bank-quarters), and large (>$50B; ≈1,500 bank-quarters) cohorts. The wide dispersion among small banks reflects both economic heterogeneity in funding models and the depth of the community-bank panel; the narrow boxes for medium and large institutions confirm that these cohorts cluster near the regulatory frontier. Sample: quarterly bank panel, 1993Q2–2019Q4. Source: US Call Reports.

2.4 Empirical Strategy

Guided by the cross-sectional regularity documented in Section 2.3-namely, that precautionary capital buffers are systematically larger for institutions facing funding market frictions, and that the distance to the regulatory frontier, rather than bank size per se, dictates the intensity of credit adjustments-we now implement a causal macroeconomic framework. To capture how lending responds to an unexpected tightening of capital requirements across heterogeneous intermediaries, we identify exogenous shifts in the regulatory capital target and trace their dynamic effects on the cross-section of credit supply. U.S. banking law, including the Dodd-Frank Act and subsequent tailoring legislation, applies asset thresholds of $10B and $50B to distinguish community, regional, and systemically important institutions; the cross-sectional dispersion in precautionary buffers documented in Table III, however, cuts across these statutory tiers, which motivates conditioning the analysis on the buffer distribution itself. The two sorts are close to orthogonal: statutory tier explains only 1.8% of the cross-sectional variance in regulatory slack; 86% of the banks in the thinnest-buffer tercile—47% by assets—face the *lowest* statutory requirement, while

conversely 84% of institutions above the $50 billion threshold already sit in that tercile, because the size add-ons raise the requirement against which their buffer is measured (Online Appendix OA.F). Our subsequent empirical analysis decomposes these credit supply paths across residential real estate, commercial business, consumption, and credit card lending segments.

We employ the Local Projection (LP) method (Jordà, 2005). LPs are particularly well-suited for our analysis as they accommodate exogenously identified shocks and provide a flexible framework for examining heterogeneous responses across different credit types and bank sizes without imposing the rigid dynamic structure of a standard VAR (Plagborg-Møller & Wolf, 2021).

Our benchmark system is defined by the vector $x_t = [l_t^{RE}, l_t^{CI}, e_t, g_t, i_t]'$, where l_t^j represents log lending in sector j (both sectors enter every equation, so each response is conditioned on the other sector's history), e_t is log equity, g_t is real GDP growth, and i_t is the federal funds rate. We estimate the impulse responses at horizon h using the following specification:

$$x_{t+h} - x_{t-1} = \alpha_h + \beta_h \text{shock}_t^{\text{Cap}} + \sum_{p=1}^{6} \Psi_{h,p} x_{t-p} + \varepsilon_{t+h} \tag{2}$$

where $\text{shock}_t^{\text{Cap}}$ is the unanticipated capital requirement surprise identified via the narrative and sign-restriction approach of Drechsel & Miura (2025). The shock series is available from 2002Q1, so the effective estimation sample is 2002Q1–2019Q4; lending and control histories extend back to 1994. Following the AIC criterion, we include six lags of the control variables to account for macroeconomic momentum; the results are robust to four-lag and eight-lag alternatives (Online Appendix). Aggregate specifications are estimated by OLS with Newey–West standard errors (lag $h + 6$); bank-level specifications include bank fixed effects, winsorise log lending at the 1st/99th percentiles, and cluster standard errors by bank. Because the identified shock is an aggregate series, the bank-level regressions exploit cross-sectional heterogeneity in exposure, with the macroeconomic controls absorbing the common cycle.

Why local projections rather than a vector autoregression. We estimate (2) horizon by horizon rather than iterating an estimated system forward, for three reasons specific to this application. First, the shock is identified *outside* the system: $\text{shock}_t^{\text{Cap}}$ is the narrative and sign-restriction series of Drechsel & Miura (2025), so the identification a vector autoregression would supply internally—through a recursive ordering or sign restrictions on reduced-form innovations—is already in hand, and we project directly onto the shock (Jordà, 2005). Second, the object of interest is a *conditional* impulse response: the

paper's central result is how the response varies across the buffer distribution, which local projections deliver by estimating (2) separately within each tercile, whereas recovering the same object from a vector autoregression requires an interacted or regime-switching system in which the horizon-h response is a non-linear function of the entire coefficient matrix. Third, the data are a large unbalanced panel with entry and exit, which a panel vector autoregression would have to force into a balanced system with dynamics common to every bank.

The cost is efficiency, and it is visible in our confidence bands. A vector autoregression estimates one-step-ahead dynamics once and iterates them forward, so every horizon inherits the same estimated structure—efficient when that structure is correct. Local projections instead run a separate regression at each horizon, imposing nothing on the shape of the response and paying for that freedom in variance, increasingly so as h grows. We regard the trade as favourable here, because the restriction a vector autoregression would impose—a common propagation structure across banks and horizons—is precisely what the cross-sectional heterogeneity of Section 2.3 calls into question. The two approaches do not estimate different objects: Plagborg-Møller & Wolf (2021) show that local projections and vector autoregressions recover the same population impulse responses, and that the two coincide as the lag order grows with the sample. Our choice is therefore one of finite-sample robustness, not of identification.

Three checks support the identification. First, a pre-trend placebo re-estimating equation (2) at horizons $h \in \{-4, \ldots, -1\}$ yields coefficients indistinguishable from zero ($|t| < 1.5$ throughout): lending does not anticipate the identified innovations. Second, the shock series itself is serially uncorrelated at the quarterly frequency (first-order autocorrelation -0.47, mean ≈ 0), consistent with its construction as a sequence of surprises; the near-unit-root persistence estimated in Section 5 is a property of the requirement *level* that the surprises cumulate into, not of the surprises themselves. Third, consumption and credit-card lending—asset classes carrying full risk weights but funded against high interest margins—show no systematic response at any horizon, indicating that the shock transmits through the regulatory valuation of the balance sheet rather than through aggregate credit demand.

2.5 Results: Credit Type and Bank Size Heterogeneity

Figure 7 presents the local projection impulse response functions (IRFs) of four distinct credit categories following a positive, one-standard-deviation shock to macroprudential capital requirements. A clear hierarchy of structural responses emerges across credit instruments, ordered by regulatory risk weight.

Business lending (C&I) exhibits the most pronounced and persistent contraction, consistent with the hypothesis that regulatory tightening induces banks to compress exposures to assets carrying higher capital charges.[4]

Conversely, real estate lending displays notable structural resilience, remaining relatively stable over the forecast horizon. This trajectory suggests a classic portfolio "flight to quality," wherein institutions preferentially reallocate or maintain capital allocations in asset classes characterized by historically lower regulatory capital charges and stable collateral foundations.

Meanwhile, consumption and credit card lending channels prove highly inelastic, showing statistically negligible volume responses. This finding is particularly informative because consumer loans carry the same 100% risk weight as C&I credit under Basel III, so the regulatory *gross* cost per dollar of each asset class is identical. The fact that banks nonetheless contract C&I but not consumer credit implies that the *net* regulatory cost—gross capital cost minus the interest-margin buffer—is the economically relevant object: the elevated spreads on consumer credit lines provide sufficient cash-flow buffers for banks to internalise the higher capital charge without volume rationing, whereas the thinner margins on business lending leave the balance-sheet channel operative.

[4] C&I and mortgage lending also differ on other dimensions—maturity, collateral quality, relationship intensity, and loan-demand cyclicality—so the differential response reflects the risk-weight wedge jointly with these characteristics. The consumer-credit placebo (same 100% risk weight as C&I, no response) and the structural model's decomposition in Section 3 help discipline the interpretation.

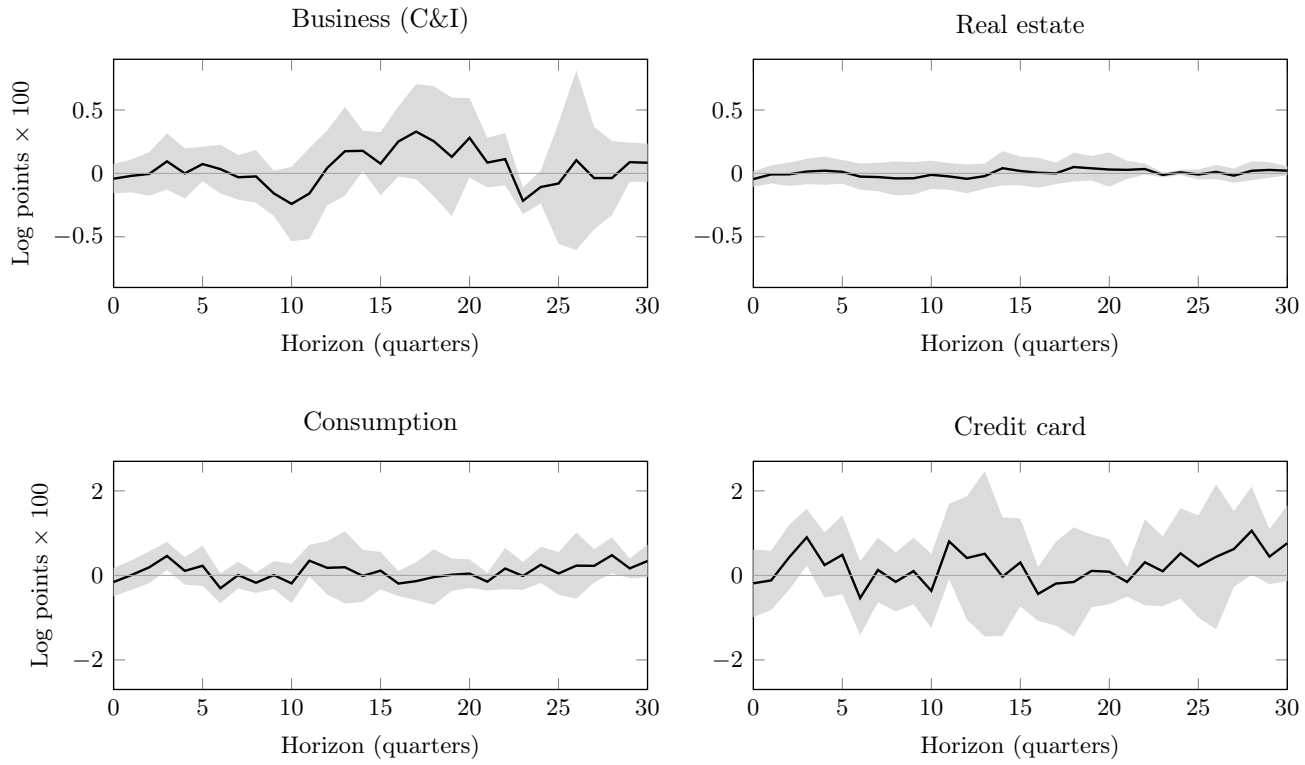


FIGURE 7.—Local projection IRFs of credit supply to a capital requirement shock

Notes: Aggregate local-projection impulse responses of the four credit aggregates to the identified capital-requirement innovation (equation 2); shaded areas are 95% Newey–West confidence bands. Panels within each row share a common vertical scale; the bottom-row (consumer credit) responses are statistically indistinguishable from zero at all horizons. Sample 2002Q1–2019Q4; lag order $p = 6$. Aggregates are unweighted sums over the bank panel. Source: US Call Reports.

Furthermore, our cross-sectional analysis reveals stark heterogeneity in the transmission of capital shocks across the banking sector, organised by the *distance to the regulatory frontier* rather than by asset size per se. Sorting banks each quarter into within-quarter terciles of their lagged regulatory slack $D_{i,t-1}$ and re-estimating the within-bank local projections tercile by tercile (Figure 8), institutions operating with thin voluntary buffers contract lending significantly more following a tightening: the average C&I response over the first twenty quarters is -0.17% per unit innovation in the bottom tercile against -0.13% in the top tercile—a low-to-high ratio of 1.32 (90% CI [1.01, 1.62]), and 1.41 ([1.14, 1.67]) for real estate; pooling the two sectors, which a formal test licenses (Online Appendix OA.E), gives 1.37 ([1.13, 1.61])—with troughs of -0.53% versus -0.41% at $h \approx 15$, and C&I contracting more than real estate *within every tercile*—the sectoral asymmetry that the stylised model of Section 3 attributes to risk-weight-driven asset substitution. This buffer gradient is precisely what the structural framework predicts, and quantitatively so: sorting the model's banks by their initial regulatory slack and tracing

the response of each tercile's lending to the estimated requirement shock—a comparative static of the estimated policy function across the initial buffer distribution, computed at fixed equity and scale so that the gradient reflects the policy rule rather than differential accumulation—implies a low-to-high C&I response ratio of 1.23, closely matching the empirical 1.32, and reproduces the sign and sectoral ordering of the gradient (Section 3). The gradient is robust to alternative definitions of the regulatory buffer: measuring slack relative to the leverage ratio instead of the risk-based ratio, using fixed capital-ratio cutoffs rather than within-quarter terciles, and varying the bandwidth of the "near-frontier" zone all yield low-to-high response ratios in the range 1.2–1.4 (Online Appendix). We note that this range lies well inside the sampling confidence interval of any single estimate, so these checks establish insensitivity to the buffer definition rather than independent corroboration. Two further features of the cross-section deserve emphasis. First, the gradient is a medium-horizon phenomenon: it is indistinguishable from zero for $h \leq 8$ in C&I and becomes significant only from $h = 9$ onward, so the buffer position governs how a tightening propagates rather than its impact effect. Second, the gradient is not monotone across terciles—the middle tercile behaves statistically like the top one rather than lying between the two (-0.13% against -0.13% in C&I)—so the informative contrast is between thin-buffer banks and the rest, and we read the cross-section accordingly. Banks near the frontier adjust on the asset side, while banks with ample voluntary slack absorb more of the shock through internal capital reconfiguration—a state-dependence we formalise below as the "Insulated Zone."

Two robustness exercises sharpen the interpretation. First, the gradient reflects the *current* regulatory state rather than a permanent bank type: when tercile membership is frozen at each bank's pre-sample (1994–2001) average slack, the gradient does not merely weaken but reverses: banks in the thinnest pre-period tercile contract *less* than those in the thickest, with low-to-high ratios of 0.67 in C&I and 0.71 in real estate, against 1.32 and 1.41 when terciles are re-sorted each quarter. Sorting on a persistent characteristic therefore fails to reproduce the gradient, as expected if what matters is the buffer a bank holds when the shock arrives—the state variable of the model—rather than a fixed characteristic. The reversal is not mean reversion in slack: the frozen-thin group remains the thinner group in sample (mean slack 0.038 against 0.124 for the frozen-thick group), and frozen membership coincides with contemporaneous tercile membership in 52% of bank-quarters against 33% under random assignment. It is instead what Proposition 1 implies: because γ_i pins each bank's target buffer, a *persistently* thin bank is one that has chosen to operate thin on a low funding friction, whereas a *currently* thin bank has been moved there—different objects, which the model does not predict will behave alike. Second, a continuous

interaction of the shock with lagged slack, conditioning on size-decile-specific shock exposure, is not robust; the slack distribution has an extreme right tail of very small institutions with near-unit capital ratios, and the tercile design is insulated from this tail while a linear interaction is not. We accordingly read the cross-section through the tercile split. Because the identified innovation is an aggregate surprise with no pre-trends (Section 2.4), sorting on the lagged buffer one quarter ahead of an unanticipated aggregate shock leaves little scope for anticipatory selection. The bank fixed effects in the within-bank specification absorb any permanent component of risk appetite; the remaining concern is that *time-varying* risk appetite could correlate with buffer position. The frozen-slack test addresses this directly: because the gradient reverses when tercile membership is fixed at pre-sample averages, the cross-sectional signal cannot be attributed to a slow-moving bank type—including persistent differences in portfolio riskiness—but must instead reflect the time-varying buffer state, as the model's mechanism requires. Finally, the headline responses are not an artifact of ending the estimation sample in 2019: re-estimating the small-cohort local projections on the full extent of the identified shock series (2002Q1–2023Q2), either controlling for a 2020–21 pandemic dummy interacted with the shock or excluding pandemic-dated shocks outright, leaves the target moments essentially unchanged—mean C&I responses of -0.142% per unit shock over $h = 1$–20 under both treatments, against -0.145% in the baseline window, and -0.106% versus -0.105% for real estate.

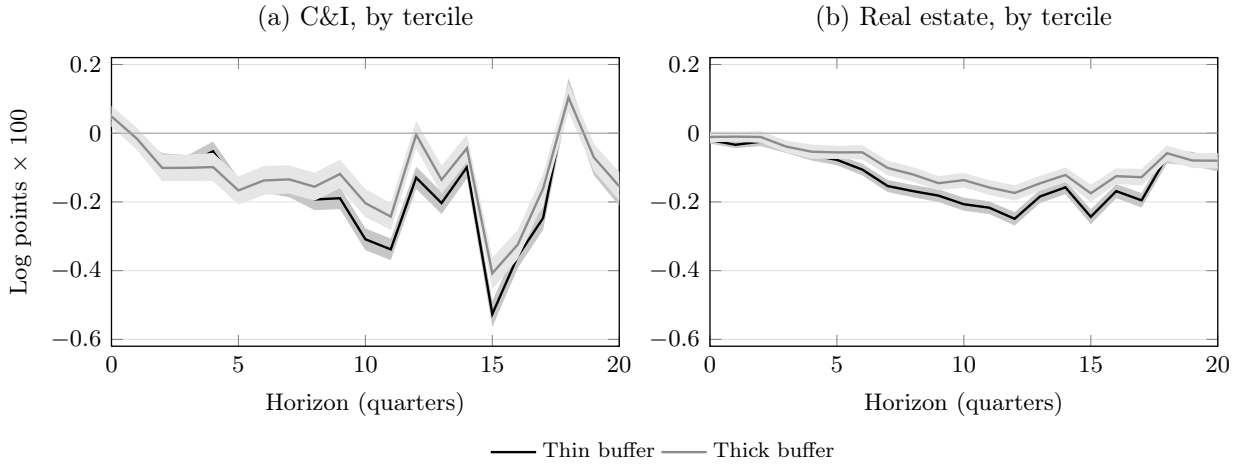


FIGURE 8.— Credit responses by capital-buffer tercile

Notes: Within-bank local projections (bank fixed effects, standard errors clustered by bank; 90% bands shaded) estimated separately for banks in the bottom (darker) and top (lighter) within-quarter terciles of lagged regulatory slack $D_{i,t-1}$; sample 2002Q1–2019Q4, lag $p = 6$. Low-buffer banks contract more in both sectors, with the gap larger for C&I (mean low-to-high response ratio 1.32×) than for RE (1.41×). The solved policy reproduces the C&I gradient (1.23×, untargeted); the model's exact static solution gives 1.32× (Online Appendix OA.G). Responses in log points (×100) per unit of the identified capital-requirement innovation.

We emphasise that statistical power is concentrated among the several thousand community and regional banks that constitute the vast majority of the panel; the small club of institutions above the $50B threshold—roughly forty in the final sample year—is too small to sign its cohort-level responses with confidence (Figure 9). This is a structural feature of U.S. banking data, not a specification choice: the cross-section of SIFIs is an order of magnitude smaller than the community and regional tiers, and no feasible identification strategy based on aggregate capital-requirement shocks can overcome this limitation within the current regulatory regime. The empirical identification therefore leverages the rich cross-sectional variation in capital buffers across the mid-to-large bank distribution where statistical power is sufficient. Crucially, the buffer gradient—the paper's main cross-sectional result (1.32× low-to-high response ratio)—is estimated across the *full* panel, not within any single size cohort, and the tercile sort pools banks of all sizes by their regulatory slack. The gradient's strength comes precisely from the thousands of community and regional institutions that span the entire range of buffer positions (Figure 6), while SIFI estimates, though directionally consistent, contribute little to the precision of the aggregate result. These findings imply that the aggregate transmission of capital regulation is conditional on the cross-sectional distribution of voluntary capital buffers within the banking sector. Notably, this state-dependence does not extend to consumer loan portfolios; consumption and credit card allocations remain uniformly unresponsive to regulatory innovations throughout the cross-section.

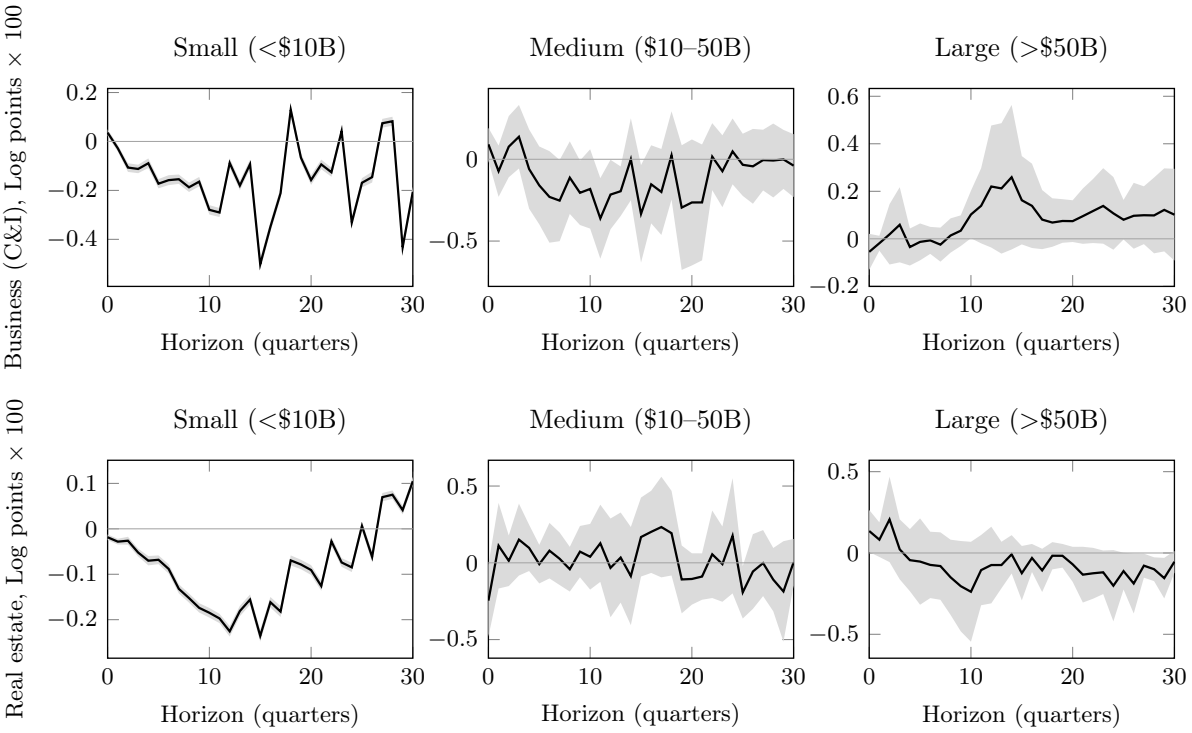


FIGURE 9.— Within-bank LP impulse responses by bank size cohort

Notes: Within-bank local projections (bank fixed effects, standard errors clustered by bank) of C&I and real estate lending by size cohort (split at \$10B/\$50B of lagged assets). Shaded areas are 90% confidence bands. Lag $p = 6$; sample 2002Q1–2019Q4. Note the scale: the small-bank panels are precisely estimated; the large-bank cohort (≈40 institutions) is not. Source: US Call Reports.

To reconcile why well-buffered institutions exhibit attenuated credit supply paths while thinly buffered institutions contract significantly, we introduce a conceptual framework governed by what we term the "Insulated Zone." Historical macroprudential innovations identified in our sample period have been small-to-moderate in magnitude. Banks that maintain large, voluntary ex-ante capital buffers above statutory minimums—predominantly deposit-funded institutions that self-insure against funding-market frictions (Table III)—found these realized historical policy shocks insufficient to deplete their surplus equity. Consequently, such lenders remained within this Insulated Zone, absorbing regulatory changes via internal capital configurations without being forced to adjust their asset mix. As shown in the structural exercise below, the severe non-linear credit contraction predicted for currently well-buffered institutions is a latent vulnerability—one that remains dormant under historical policy variations but rapidly manifests under large, unprecedented counterfactual shocks that completely erode their voluntary buffers.

3 A Stylized Model of Capital Requirements and Credit Allocation

The empirical analysis establishes three robust stylized facts: (i) an aggregate contraction in bank credit following an exogenous capital-requirement shock, (ii) a disproportionate, immediate contraction in business (C&I) lending alongside a resilient mortgage supply, and (iii) a pronounced cross-sectional non-linearity organised by the ex-ante distribution of voluntary capital buffers, for which bank size and funding structure are the observable correlates. This section presents a stylized partial equilibrium model to rationalize these joint dynamics. The framework formally isolates the structural interactions between liability-side funding liquidity constraints, endogenous precautionary capital buffers, and asset-specific regulatory risk weights. Throughout, the primitive source of heterogeneity is a bank's access to marginal funding—not its balance-sheet scale per se—consistent with the panel evidence of Section 2.3.

3.1 Environment and Liability Structure

We consider a static model of a banking sector populated by a continuum of risk-neutral banks indexed by $i \in [0, 1]$, following a standard modelling choice in the macro-banking literature (e.g., Corbae & Erasmo, 2021; Begenau, 2020). Banks differ in their initial equity endowments e_i, structural liability capacity, and baseline asset composition. Lenders intermediate funds by extending credit across two distinct, risk-rated asset classes: residential mortgages (A_i^M) and commercial business loans (A_i^B). Total balance sheet assets are defined by $A_i = A_i^M + A_i^B$.

The asset portfolio is financed via a combination of bank equity and liabilities. To map empirical funding heterogeneities, we partition bank liabilities into two distinct categories:

1. **Core Deposits ($d_{i,C}$):** Stable, low-cost retail funding. While community institutions rely extensively on core deposits as their primary liability franchise, their aggregate supply is locally inelastic and capped by an absolute capacity constraint determined by the bank's geographic footprint:

$$d_{i,C} \leq \bar{d}_{i,C} \tag{3}$$

2. **Wholesale Funding ($d_{i,W}$):** Marginal, market-priced wholesale liabilities utilized to close the balance sheet identity. For any given asset allocation, wholesale funding is determined residually

via the accounting constraint:

$$d_{i,W} = A_i^M + A_i^B - e_i - d_{i,C} \tag{4}$$

Accessing marginal wholesale markets incurs a quadratic friction cost given by $\frac{\gamma_i}{2} d_{i,W}^2$, where the parameter γ_i dictates bank-specific structural access to capital markets. We assume an asymmetric friction distribution such that $\gamma_{\text{small}} \gg \gamma_{\text{large}} \to 0$. In practice, large, diversified banking organizations possess highly elastic access to wholesale funding lines, whereas community banks face high entry barriers, informational asymmetries, and steep placement penalties in non-deposit funding markets; accordingly, we calibrate γ_{large} to large-bank and γ_{small} to small-bank averages (Table IV), while emphasising that the primitive source of heterogeneity is funding-market access, not balance-sheet scale per se.

3.2 The Regulatory Capital Frontier and Endogenous Buffers

Banks are subject to a binding risk-based capital regulation. The regulatory authority mandates that total equity must represent at least a fraction κ of total Risk-Weighted Assets (RWA_i). Risk-weighted exposures are defined as $RWA_i = \omega_M A_i^M + \omega_B A_i^B$, where commercial business loans carry a structurally higher regulatory risk weight than residential mortgages ($\omega_B > \omega_M$).

We define a bank's absolute distance to its regulatory capital frontier as the capital buffer D_i:

$$D_i = e_i - \kappa \left(\omega_M A_i^M + \omega_B A_i^B \right) \tag{5}$$

Banks hold $D_i \geq 0$ as an endogenous insurance cushion against unexpected structural shocks. If a macroprudential policy innovation tightens the capital regime ($\Delta\kappa > 0$) such that the constraint binds ($D_i < 0$), the institution incurs a severe regulatory non-compliance penalty.

Because low-friction institutions ($\gamma_i \to 0$, empirically the large, diversified organizations) possess near-perfect marginal funding liquidity, they can adjust their balance sheets ex-post to manage capital shortfalls, and consequently optimize thin ex-ante cushions ($D_i \approx 0$). In contrast, high-friction banks—those whose retail deposit ceiling $\bar{d}_{i,C}$ binds and who face steep wholesale costs γ_i—cannot rapidly restructure their funding, and endogenously hoard large precautionary buffers ($D_i > 0$) ex-ante. Note the precise role of the deposit franchise here, which reconciles this mechanism with the panel evidence of Section 2.3: what disciplines the buffer is the *marginal* funding source. A bank whose stable deposit franchise comfortably covers its funding needs faces little rollover risk and holds a thin buffer; a bank that funds—or, were its

buffer breached, would have to fund—at the wholesale margin self-insures. Size matters only through its correlation with γ_i and with the tightness of the deposit ceiling.

Why a funding friction drives an equity-ratio buffer. Because the regulatory constraint is $e_i \geq \kappa\, RWA_i$, the scarce regulatory object is equity, not liabilities. Two features of the model jointly explain why the wholesale funding friction γ_i nonetheless governs precautionary equity buffer choice. First, the dividend non-negativity constraint ($c_i \geq 0$) implicitly rules out equity issuance: a bank that suffers a capital shortfall cannot raise fresh equity on demand.[5] Second, the wholesale friction γ_i determines the cost of the *remaining* margin of adjustment: expanding or restructuring the liability side. A low-γ_i bank can cheaply raise wholesale deposits, using the proceeds to substitute out of high-risk-weight assets or to hold liquid buffers, thereby managing its capital ratio at low cost despite the inability to issue equity. A high-γ_i bank faces steep costs on this same margin; moreover, the quadratic friction $\frac{\gamma_i}{2} d_{i,W}^2$ drains equity directly through funding costs, tightening the capital constraint further. Together, these two frictions make the precautionary buffer the only viable cushion against capital shortfalls for high-friction institutions.

To see this formally, consider the optimal buffer in the interior solution. From the first-order conditions, the bank's desired buffer D_i^* satisfies

$$\frac{\partial D_i^*}{\partial \gamma_i} > 0, \tag{6}$$

because a higher γ_i raises the marginal cost of ex-post wholesale adjustment (the left-hand side of equation (12)), increasing the shadow value of staying away from the regulatory frontier. In this sense, γ_i captures a broad external balance-sheet adjustment friction that encompasses both the cost of tapping wholesale markets and the implicit cost of being unable to issue equity; banks with low γ_i need not self-insure because they can restructure cheaply after a shock, while banks with high γ_i must build the cushion ex ante.

A closed-form characterisation of the capital target. The dynamic model of Section 4 sharpens this comparative static into an exact statement. Taking its continuous-time limit and writing $\rho = -\ln \beta$ for the instantaneous discount rate, the optimal payout policy is a barrier: the bank retains earnings until equity reaches a target $\bar{e}_i$ and distributes the flow thereafter. Smooth pasting at that barrier yields

[5] In the dynamic quantitative model of Section 4, this same restriction appears as the occasionally binding dividend constraint $c_{i,t} \geq 0$.

the following.[6]

Proposition 1 (Funding frictions and the capital target)**.** *In the stationary equilibrium, a bank with funding friction γ_i and core-deposit capacity $\bar{d}_{i,C}$ targets the wholesale funding level*

$$d_i^{W\star} = \frac{\rho + \delta_e + \tau - r_W}{\gamma_i}, \tag{7}$$

and holds equity $\bar{e}_i$ satisfying $\bar{e}_i = A_i(\bar{e}_i) - \bar{d}_{i,C} - d_i^{W\star}$. Consequently:

1. *the target capital ratio $\bar{e}_i/RWA_i$ is strictly increasing in γ_i;*
2. *there exists a threshold γ^* such that banks with $\gamma_i \leq \gamma^*$ target the regulatory frontier exactly—zero voluntary buffer, $\bar{e}_i/RWA_i = \kappa$—while banks with $\gamma_i > \gamma^*$ target a strictly positive buffer;*
3. *the elasticity of target equity to the capital requirement is approximately one for $\gamma_i \leq \gamma^*$ and zero for $\gamma_i > \gamma^*$.*

Equation (7) says that what a bank targets is a level of *wholesale funding*, with equity the residual that delivers it. The marginal gain from retaining a dollar is the marginal funding cost it displaces, $r_W + \gamma_i d_i^W$; the bank retains until that gain equals the discount rate net of depreciation and the insurance levy. The mechanism is therefore a funding-cost mechanism rather than an insurance one, which is worth stating precisely: high-γ_i banks hold more equity because marginal wholesale funding is expensive, not because they are hedging a shock.

Parts (ii) and (iii) deliver the cross-sectional pattern that the local projections of Section 2.5 document. At our calibration $\gamma^* \approx 0.051$, against $\gamma_{\text{large}} \to 0$ and $\gamma_{\text{small}} = 0.077$: low-friction institutions target the frontier, hold no voluntary buffer, and pass a tightening through one-for-one, while high-friction institutions target a capital ratio of roughly 26% and are insulated from it. The banks whose lending responds to a requirement change are exactly those holding thin buffers—which is the ordering the buffer-tercile regressions recover.

This non-linear threshold dynamic bridges our structural theory with the empirical patterns documented in Section 2.5. Figure 10 illustrates this state dependency via the mapping of the *Insulated Zone*, defined

[6]The characterisation is obtained in the continuous-time limit of the problem in Section 4, where the payout barrier is pinned by the smooth-pasting and super-contact conditions $V'(\bar{e}_i) = 1$ and $V''(\bar{e}_i) = 0$. Because idiosyncratic credit risk is small relative to the equity buffer at our calibration, the deterministic limit is accurate and the target solves $\mu'(\bar{e}_i) = \rho$, with μ the drift of equity. We verify the characterisation against an independent discrete-time value function iteration on the same primitives; the two agree on the value function to within 1.5% and on portfolio policies to within 2%.

on the joint distribution of funding frictions and buffers:

$$\text{Insulated Zone} \equiv \left\{ (\gamma_i, D_{i,t}) \;\middle|\; D_{i,t} > \bar{D}_{\text{critical}}(\gamma_i) \right\}, \tag{8}$$

where the critical threshold $\bar{D}_{\text{critical}}(\gamma_i)$ is increasing in the funding friction: the harder it is to replace funds at the margin, the larger the buffer required for a given policy innovation to leave lending unaffected.

When policy shocks (ϵ_t^κ) are of mild-to-moderate magnitude—matching the historical variations isolated in our Local Projections—high-friction banks operate deep within this zone precisely because they have chosen high baseline capital distances ($D_{i,t}$). Because these regulatory innovations fail to push $D_{i,t}$ below the critical threshold $\bar{D}_{\text{critical}}(\gamma_i)$, the shadow cost of the regulatory capital constraint remains entirely non-binding ($\lambda_i = 0$), leaving their credit supply smoothly insulated. However, the theory reveals that this empirical resilience is conditional rather than absolute. Under larger, unprecedented counterfactual policy shocks that breach the frontier of the Insulated Zone, the funding rigidities that motivated the buffer in the first place ($\gamma_i \gg 0$) bind acutely. Once the precautionary buffer is fully exhausted, the asset side becomes hypersensitive, triggering the sharp, non-linear drop in credit supply predicted by the model. The attenuated empirical response of well-buffered banks is therefore not evidence of structural immunity, but rather proof that historical innovations have simply never been large enough to push them out of their Insulated Zone.

3.3 Bank Optimization and General First-Order Conditions

The representative bank selects its optimal multi-asset credit allocation to maximize total economic profits Π_i, net of quadratic asset management costs ψ_j, a cost of regulatory capital ϕ_j defined below, and quadratic wholesale frictions:

$$\max_{A_i^M, A_i^B, d_{i,C}} \quad \sum_{j \in \{M,B\}} r_j A_i^j - \sum_{j \in \{M,B\}} \frac{\psi_j}{2} (A_i^j)^2 - \sum_{j \in \{M,B\}} \phi_j(\kappa) A_i^j - r_D d_{i,C} - r_W d_{i,W} - \frac{\gamma_i}{2} d_{i,W}^2 \tag{9}$$

The term $\phi_j(\kappa)$ is a per-unit cost of extending credit in sector j that depends on the capital requirement rather than on the volume of lending. Because it enters the objective linearly, it raises the marginal cost of a loan by the constant $\phi_j(\kappa)$, in contrast to the operating cost $\frac{\psi_j}{2}(A_i^j)^2$, whose marginal cost $\psi_j A_i^j$

rises with the size of the portfolio. We write it as

$$\phi_j(\kappa) \;=\; \Big(\underbrace{c}_{\text{not risk-weighted}} \;+\; \underbrace{z\,\omega_j}_{\text{risk-weighted}} \Big)\, \max(\kappa - \bar{\kappa},\, 0)\,, \tag{10}$$

where $\bar{\kappa}$ is the requirement the bank faces in the calibrated steady state, so that $\phi_j = 0$ there and the term bites only when the requirement tightens. In the dynamic model of Section 4 the schedule is size-dependent, and the comparison is to each bank's *own* baseline requirement $\bar{\kappa}_i$, so a large bank does not carry a permanent cost merely for facing a higher statutory tier. The second component is risk-weighted and carries the standard Modigliani–Miller reading: funding an additional unit of *risk-weighted* assets with equity rather than debt costs z more. The first is not risk-weighted. A leverage requirement generates a cost of exactly this shape: because it applies to total rather than risk-weighted assets, the additional equity it forces a bank to hold scales with lending irrespective of risk weight, and the expense of that equity is borne per unit of credit extended. U.S. banks face such a requirement alongside the risk-based rule, whereas the constraint below contains only the latter. We treat that interpretation as a conjecture consistent with the estimates rather than as identified by them, and return to it when the two parameters are calibrated.

The bank maximises subject to the accounting identity, the retail deposit capacity constraint, and the regulatory capital constraint:

$$e_i - \kappa \left(\omega_M A_i^M + \omega_B A_i^B\right) \geq 0 \tag{11}$$

Let λ_i denote the Lagrange multiplier associated with the regulatory capital requirement, representing the shadow cost of regulatory capital. Substituting the wholesale identity directly into the objective function, the interior first-order conditions with respect to asset class $j \in \{M, B\}$ yield:

$$r_j - r_W - \phi_j(\kappa) - \psi_j A_i^j - \gamma_i \left(A_i^M + A_i^B - e_i - d_{i,C}\right) = \lambda_i \kappa \omega_j \tag{12}$$

The requirement reaches the bank through two distinct routes in (12). When the constraint binds ($\lambda_i > 0$) the shadow cost enters multiplied by the risk weight ω_j, which is the asset-substitution channel developed below. When the constraint is slack ($\lambda_i = 0$) that route closes, and the requirement acts *only* through $\phi_j(\kappa)$.

The role of the second route is clearest if we shut it down. Suppose the requirement acted through the

capital constraint alone, as it does when regulation is modelled purely as a restriction on the balance sheet; in (10) this is the case $c = z = 0$. Then κ would not appear anywhere in the problem of a bank whose constraint is slack, and $\partial A_i^j / \partial \kappa = 0$ *exactly*. A bank away from the frontier would not respond to a tightening at all, and the whole burden would fall on the banks the constraint reaches.

That restriction is strong, and the cross-section rejects it. Fewer than three percent of bank-quarters sit at or inside the regulatory frontier, yet the buffer gradient documented in Section 2 spans the whole distribution, with the thick-buffer tercile contracting significantly.[7] A cost of regulatory capital that is not risk-weighted and applies away from the frontier is what reconciles the two, and this is the sense in which the empirical gradient *identifies* ϕ_j rather than merely being matched by it.

Equation (12) also demonstrates the structural channels driving our empirical cross-sectional results. For low-friction banks, an increase in regulatory stringency ($\Delta\kappa > 0$) yields a minor impact on credit because near-frictionless capital market access ($\gamma_i \to 0$) keeps the shadow cost of capital λ_i low. For high-friction banks, however, the macroprudential innovation directly confronts severe funding rigidities ($\gamma_i \gg 0$). Once a shock is sufficiently large to breach the ex-ante precautionary cushion D_i, the inability to tap external financing causes an acute spike in λ_i, forcing a sharp, non-linear contraction in loan volumes.

3.4 The sectoral lending channel and Risk-Weight Multipliers

When a macroprudential tightening shock binds ($\lambda_i > 0$), the cross-sectional portfolio reallocation path is governed by the structural risk-weight wedge ($\omega_B > \omega_M$), operating through two distinct margins that depend crucially on the institution's baseline diversification and asset size:

1. **The Asset Substitution Margin:** This margin is driven primarily by large and medium banking institutions, which possess the requisite balance sheet diversification (holding a substantial 60% non-mortgage asset baseline) to actively reallocate credit. For such a diversified banking institution, a marginal increase in the regulatory shadow tax λ_i acts asymmetrically across asset tiers. Differentiating the optimal asset-mix ratio with respect to the capital constraint demonstrates

[7] This magnitude is corroborated by an independently constructed structural model. Corbae & Erasmo (2021) build a banking economy on different primitives—general equilibrium, a single loan sector, imperfect competition, and a requirement enforced as an exit condition rather than as an occasionally binding constraint—and calibrate it to a different sample of U.S. banks. Their baseline nonetheless places only 1.5% of banks within two percentage points of the requirement, rising to 2.0% under a counterfactual tightening. Their band is wider than ours, which counts only banks at or inside the frontier, and still holds a comparably small mass. That the binding set is small is therefore a feature of the U.S. cross-section and not an artifact of how we measure regulatory slack.

that:

$$\left|\frac{\partial A_i^B}{\partial \kappa}\right| > \left|\frac{\partial A_i^M}{\partial \kappa}\right| > 0 \tag{13}$$

Because commercial business lending carries an elevated risk weight ($\omega_B > \omega_M$), it imposes a steeper marginal capital cost per dollar of credit extended. With ϕ_j in place the inequality extends away from the frontier: $\partial\phi_j/\partial\kappa = c + z\omega_j > 0$, so both derivatives are strictly negative for *every* bank, whereas under $c = z = 0$ they are identically zero outside the binding regime and the strict inequality has no content there. Whether C&I remains the larger response in absolute value depends in addition on the operating costs ψ_j and on the cross-sector funding term γ_i, which do not order the two sectors in the same direction; we verify numerically that C&I dominates at the calibrated parameters rather than asserting it from the risk weights alone. To satisfy the regulatory frontier while minimizing volume liquidations, diversified large and medium institutions preferentially shed high-weight C&I exposures while preserving real estate credit paths.

Crucially, highly specialized institutions—predominantly community banks whose real estate portfolios routinely exceed 90% of total lending—lack the structural capacity to utilise this Asset Substitution Margin. For these lenders, portfolio rebalancing across sectors is not a feasible margin; their response to macroprudential innovations is instead governed by whether their ex-ante capital buffers allow them to remain entirely within their voluntary *Insulated Zone*.

2. **The Scale Deleveraging Multiplier:** If an institution is forced to pass beyond its buffer and must compress its balance sheet to defend its capital ratio, the regulatory relief generated per dollar of loan liquidation scales directly with the asset's risk weight:

$$\frac{\partial RWA_i}{\partial A_i^B} = \omega_B > \omega_M = \frac{\partial RWA_i}{\partial A_i^M} \tag{14}$$

While low-risk-weight real estate assets require massive nominal balance sheet liquidations to yield marginal regulatory relief, high-risk-weight business portfolios (ω_B) represent the immediate, highly efficient margin of balance sheet optimization. This mechanism rationalizes our empirical Local Projection results, where commercial business portfolios at diversified larger cohorts bear the sharpest and most rapid supply adjustments, while specialized, well-buffered institutions remain insulated from such sharp non-linear scale contractions under moderate historical shocks.

3.5 Cross-Sectional Heterogeneity and Non-linear Aggregation

The aggregate response of the macroeconomy is ultimately determined by the cross-sectional density of banks for whom the capital frontier actively binds. Figure 10 provides the core graphical intuition of the framework.

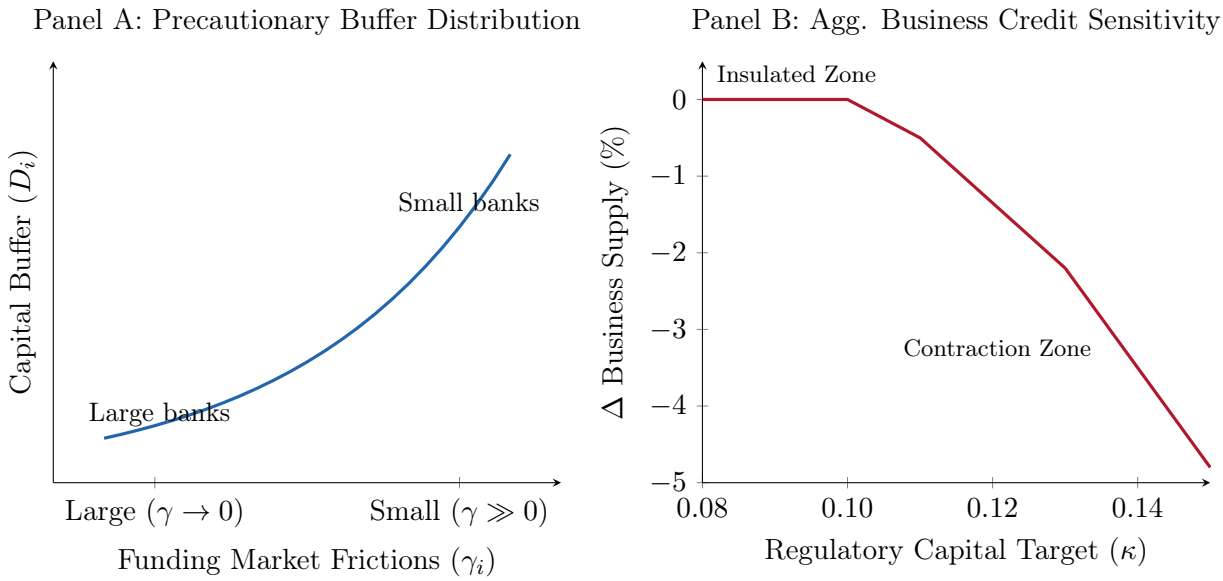


FIGURE 10.—Funding capacity, buffer selection, and aggregate credit sensitivity

Notes: Panel A illustrates the endogenous choice of ex-ante capital buffers as a function of capital market funding frictions. Panel B highlights the non-linear macroeconomic response as capital tightening breaches cross-sectional buffers.

Low-friction institutions run with thin buffers but leverage highly elastic funding channels to reallocate asset compositions fluidly. Conversely, well-buffered banks—a group that spans all size categories but is disproportionately composed of institutions whose deposit franchise covers their funding needs—remain completely insulated from minor policy changes (the "Insulated Zone" in Panel B). However, once a macroprudential innovation is large enough to exhaust these cushions, the funding rigidities that motivated the buffer in the first place force an acute, non-linear drop in credit supply concentrated heavily in high-risk-weight business segments.

4 Quantitative Model: A Heterogeneous Macro-Banking Framework

While the stylized framework in Section 3 provides the core qualitative mechanics of the sectoral lending channel, this section develops a fully specified quantitative partial-equilibrium framework to assess the

macroeconomic magnitude of these structural transformations. We transition from a static cross-sectional environment to an infinitely lived, calibrated heterogeneous-agent model where the joint distribution of balance sheet scale and regulatory slack is disciplined by high-frequency U.S. banking data.

Scope and interpretation of the partial-equilibrium framework. The model abstracts from three channels that a general-equilibrium extension would incorporate: (i) borrower substitution toward non-bank intermediaries (shadow banks, bond markets), which would attenuate the real-side impact of bank credit contraction; (ii) equilibrium loan-rate adjustment, whose direction is ambiguous because both demand and supply shift; and (iii) aggregate-demand feedback through investment, which could amplify the contraction. By shutting down these channels, the framework isolates the *bank supply mechanism*—the joint determination of sectoral credit allocation and the cross-sectional buffer gradient—that is the paper's central object of study. The estimated surplus losses of approximately 1.4–2.4% per percentage-point tightening should accordingly be read as a focused aggregate benchmark for bank-channel credit contraction, not as a full-economy welfare statement. In particular, the extent to which non-bank intermediaries absorb displaced lending remains an open empirical wedge that could substantially dampen or, if accompanied by regulatory arbitrage and fragility in the shadow sector, redirect the costs estimated here. That wedge is quantitatively large on the mortgage side: in HMDA, non-depository lenders originated 43% of U.S. mortgage volume in 2015, so a substantial share of the market this paper models sits outside the Call Report universe altogether. We leave this extension to future research. Importantly, the magnitudes the model produces (-3.5% C&I, -3.0% RE, -3.1% total per +1pp) sit squarely within the range reported by quasi-experimental studies that implicitly capture some of these equilibrium responses (Section 5.5), suggesting that the partial-equilibrium approximation does not produce implausible estimates of the bank-channel effect. We report these magnitudes as conditional on the solved policy: Online Appendix OA.G documents that the model's impulse-response *magnitudes*—and, to a degree we flag rather than minimise, the buffer gradient when it is recomputed *from* that policy—depend materially on which trained checkpoint represents the policy function, while the welfare level does not, and sets out the structural tension that underlies this dependence. The gradient we report as the model's untargeted match is computed from the static problem, whose solution is exact in each KKT regime.

4.1 Mapping Theory to Data: Measuring the Regulatory Buffer

To calibrate the initial state space of the banking sector in our quantitative exercise, we map the theoretical *Distance to Constraint* ($D_{i,t}$) to bank-level regulatory filings using quarterly Consolidated Reports of Condition and Income (Call Reports). For each individual bank i at quarter t, the empirical distance to the regulatory capital frontier is calculated as:

$$D_{i,t} = \left(\frac{\text{Total Capital}_{i,t}}{\text{RWA}_{i,t}} \right) - \kappa_{i,t} \tag{15}$$

where *Total Capital* (`RCFD7205`) and Risk-Weighted Assets (*RWA*, `RCFD7206`) are the exact empirical counterparts to the bank equity state $e_{i,t}$ and the risk-weighted denominator within our structural framework. To map our macro-finance model directly to our empirical cohort designs, the regulatory capital minimum $\kappa_{i,t}$ is formalized as a size-dependent step function dictated by the bank's lagged total asset scale ($A_{i,t-1}$). This setup directly captures the discrete regulatory thresholds established by the Dodd-Frank Act and modern Systemically Important Financial Institution (SIFI) capital surcharges:

$$\kappa_{i,t} = \begin{cases} \kappa_t & \text{if } A_{i,t-1} < \$10\text{B} \\ \kappa_t + \bar{\kappa}_{\text{DFA}} & \text{if } \$10\text{B} \leq A_{i,t-1} \leq \$50\text{B} \\ \kappa_t + \bar{\kappa}_{\text{SIFI}} & \text{if } A_{i,t-1} > \$50\text{B} \end{cases} \tag{16}$$

where κ_t represents the baseline macroprudential requirement, set to a benchmark of 10% in accordance with Prompt Corrective Action (PCA) guidelines for well-capitalized institutions.

Crucially, to fully pin down the regulatory landscape and ensure structural identification, the institutional increments are calibrated using empirical statutory requirements from the post-crisis supervisory frameworks. We set the Dodd-Frank mid-tier compliance increment to $\bar{\kappa}_{\text{DFA}} = 0.015$ (1.5%) to reflect the enhanced prudential standards and stress-testing capital conservation buffers imposed on regional banking organizations. For larger institutions we set $\bar{\kappa}_{\text{SIFI}} = 0.025$ (2.5%), a stylised increment representing the enhanced prudential standards that applied above the $50 billion threshold over most of the sample. This is a stylisation rather than a statutory schedule. The G-SIB surcharge is not an asset-size trigger: it is a score-based buffer assigned in fifty-basis-point increments to the eight U.S. global systemically important banks on the higher of two scoring methods, and the Dodd-Frank enhanced standards it stands in for were a supervisory regime rather than a single capital number. The step function captures that

requirements rise with size; it does not reproduce any particular instrument.

This empirical step function serves two purposes. First, it defines the measurement of the regulatory buffer $D_{i,t}$ used throughout the empirical analysis, so that the distance to the frontier is computed against each bank's *own* statutory requirement. Second, it pins down the tiered schedule evaluated in the policy counterfactuals of Section 5, where the size-differentiated regime is compared with a uniform requirement raising the same aggregate capital. The dynamic model itself is solved under the baseline requirement κ_t and, for the same reason of tractability, under a common funding friction $\gamma_i = \gamma_{\text{small}}$ rather than the full type distribution. Two consequences should be stated plainly. First, solving the size-dependent step function jointly with a realistic size distribution is computationally demanding and is left to future work (Section 7). Second, because banks in the solved model share a single γ_i, they share a single capital target (Proposition 1), so the cross-sectional dispersion of buffers is *calibrated* to the empirical distribution of regulatory slack rather than generated by the model. That calibration does real work rather than merely filling a gap: the cross-sectional gradient the model implies is sensitive to the assumed dispersion, so we take it from the measured slack distribution—matching its median and tenth percentile—rather than setting it freely. Figure 11 plots that distribution. It is single-peaked at a buffer of two to three percentage points with a long right tail, a median of 5.7pp and a tenth percentile of 1.2pp, and only 2.0% of bank-quarters lie at or below the requirement. The constraint is therefore slack for virtually the entire cross-section: the heterogeneity the model exploits is heterogeneity in the *distance* to the frontier, not in whether the frontier binds. The cross-sectional exercises of Sections 2.5 and 5 are accordingly comparative statics of the estimated policy function across that calibrated distribution. Proposition 1 characterises what the model delivers once the type distribution is restored, and we verify that a spread of γ_i within a factor of two reproduces the observed dispersion of capital ratios; implementing it in the solved model is a natural next step.

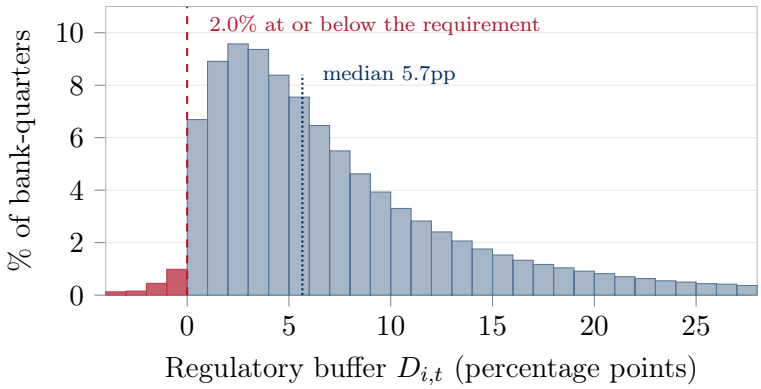


FIGURE 11.—Most banks hold a buffer of a few percentage points, and only 2% sit at or below the requirement

Notes: Cross-sectional distribution of the regulatory buffer $D_{i,t} = \text{Total Capital}_{i,t}/\text{RWA}_{i,t} - \kappa$, pooled over bank-quarters, with $\kappa = 0.10$—the Prompt Corrective Action benchmark for well-capitalized institutions used as the baseline requirement. Bins are one percentage point wide and bar heights are percentages of all bank-quarters. Bars to the left of the dashed line are observations at or below the requirement, 1.97% of the sample. The distribution is single-peaked in the 2–3pp bin, with median 5.7pp and tenth percentile 1.2pp—the two moments to which the dispersion of buffers in the solved model is calibrated. For legibility, 0.3% of observations below −4pp and 4.3% above 28pp fall outside the plotted range. Sample: quarterly bank panel, 1993Q2–2019Q4, 940,519 bank-quarters and 16,162 banks. Source: US Call Reports.

4.2 Relationship to the Literature

Our structural formulation of the regulatory buffer $D_{i,t}$ aligns with a well-established macro-banking literature that positions capital slack as a primary determinant of aggregate credit supply. Specifically, Kisin & Manela (2016) and Kovner & Van Tassel (2022) identify the shadow cost of capital regulations by observing behavioral distortions in bank asset structures near regulatory minimums. More recently, Couaillier *et al.* (2025) document a significant "Do Not Cross" effect, demonstrating that banks maintaining low distances to regulatory buffers aggressively restrict credit to avoid the market stigma, capital conservation restrictions, and supervisory penalties associated with a technical breach. Our quantitative approach advances this literature by demonstrating that for specialized or liquidity-constrained lenders, this deleveraging behavior is structurally amplified by the risk-weighting wedge ($\omega_B > \omega_M$) and concentrates heavily within high-risk-weight asset classes like commercial and industrial (C&I) business credit.

To evaluate the aggregate transmission dynamics, systemic welfare profiles, and transitional paths of the sectoral lending channel, we embed our stylized mechanics into a dynamic heterogeneous-agent macro-banking framework populated by a continuum of infinitely lived, heterogeneous banking firms. The economy is subject to aggregate macroprudential policy innovations—the only aggregate disturbance in the model—alongside idiosyncratic asset-valuation and credit-depreciation shocks. Restricting the aggregate state to a single policy shock is a deliberate parsimony choice: our identification strategy

isolates the capital-requirement channel, and introducing additional aggregate disturbances (e.g., TFP or funding-cost shocks) would multiply the aggregate state vector without improving identification of the mechanism we estimate. The idiosyncratic shocks already generate the requisite cross-sectional dispersion in buffers; what the single aggregate shock buys is a clean mapping from the estimated policy process $(\rho_\kappa, \sigma_\kappa)$ to the within-bank impulse responses that discipline the model. Market clearing in each credit sector is imposed via the linear inverse demand schedules introduced in Section 4.6.

Banks in this economy interact only through the two sectoral credit aggregates that enter the inverse demand schedules of Section 4.6, so the market-clearing rate pair $(r_{M,t}, r_{B,t})$ is a sufficient statistic for the cross-sectional distribution $\mu_t(e_i, A_i^M, A_i^B)$ from the point of view of any individual bank: no moment of that distribution beyond the two sectoral means enters a bank's payoff or its constraint set. The computational challenge is therefore not dimensionality but *shape*. Both the regulatory capital constraint and the dividend non-negativity constraint bind only occasionally, so policy functions are kinked at the frontier, and the transmission of a tightening depends on the mass of institutions sitting near that kink. We accordingly deploy a global non-linear solution method that parameterizes the individual policy functions and the associated shadow prices with deep neural networks, building upon recent advances in deep macro-finance Maliar *et al.* (2021); Azinovic *et al.* (2022), and impose market clearing analytically at each forward pass.[8]

4.3 The Dynamic Bank Optimization Problem

Banks are indexed by $i \in [0, 1]$. At the initialization of quarter t, bank i enters the period with an endogenous equity state $e_{i,t}$, a portfolio of outstanding residential mortgages $A^M_{i,t-1}$, and commercial business lines $A^B_{i,t-1}$. Lenders face asset-specific idiosyncratic valuation shocks $\varepsilon^M_{i,t}$ and $\varepsilon^B_{i,t}$ that capture localized defaults, prepayment variations, and capital depreciation margins.

The specialized banking firm maximizes the expected discounted flow of net dividend distributions to its risk-neutral equity holders:

$$V(e_{i,t}, A^M_{i,t-1}, A^B_{i,t-1}; \Omega_t) = \max_{\{A^M_{i,t}, A^B_{i,t}, d^W_{i,t}, c_{i,t}\}} \mathbb{E}_t \sum_{\tau=0}^{\infty} \beta^\tau \Lambda_{t,\tau} c_{i,t+\tau} \tag{17}$$

[8]This sufficiency is a property of the present specification, not a general one. It would fail under any of three natural extensions we leave to future work: an insurance levy τ_t set endogenously to fund the resolution of failing banks, so that the mass of institutions near the frontier enters every bank's budget constraint; a wholesale funding spread responding to aggregate sector leverage; and the size-tiered requirement schedule solved jointly with a realistic size distribution. Under any of these the shape of μ_t is payoff-relevant and a richer aggregate state becomes necessary.

where $\beta \in (0,1)$ is the subjective discount factor and $c_{i,t} \geq 0$ denotes current dividend payouts. Because the equity holders who own the financial intermediaries are risk neutral, the stochastic discount factor satisfies $\Lambda_{t,\tau} = 1$; we carry the notation for generality. The aggregate macroeconomic state vector is defined as $\Omega_t \equiv (\mu_t, \kappa_t)$, where κ_t represents the baseline macroprudential regulatory capital target.

The optimization space is structurally bounded by the bank's dynamic balance sheet identity and capital accumulation law. We explicitly distinguish between stable, retail core deposits ($d^C_{i,t}$) capped at a localized geographic franchise limit ($\bar{d}_{i,C}$), and marginal wholesale funding ($d^W_{i,t}$) which incurs a size-dependent quadratic friction (γ_i):

$$A^M_{i,t} + A^B_{i,t} = e_{i,t} + d^C_{i,t} + d^W_{i,t} \tag{18}$$

Assuming banks exhaust cheaper core deposit capacity ($d^C_{i,t} = \bar{d}_{i,C}$), the dynamic equity law of motion evaluates:

$$\begin{aligned} e_{i,t+1} = (1-\delta_e)e_{i,t} + \sum_{j\in\{M,B\}} \left[(r_{j,t} - \varepsilon^j_{i,t+1}) A^j_{i,t} - \frac{\psi_j}{2}(A^j_{i,t})^2 - \phi_j(\kappa_{i,t}) A^j_{i,t} \right] \\ - r_{d,t}\bar{d}_{i,C} - r_{W,t} d^W_{i,t} - \frac{\gamma_i}{2}(d^W_{i,t})^2 - c_{i,t} - \tau_t e_{i,t} \end{aligned} \tag{19}$$

where ψ_j is a quadratic operating cost in sector j, generating decreasing returns within each credit market and thereby pinning the long-run composition of the loan portfolio, $\phi_j(\kappa_{i,t})$ is the per-unit cost of regulatory capital introduced in equation (10), here evaluated at the bank's own requirement and measured against its own baseline,

$$\phi_j(\kappa_{i,t}) \;=\; (c + z\,\omega_j)\max(\kappa_{i,t} - \bar{\kappa}_i,\, 0)\,,$$

so that it vanishes in the calibrated steady state of every size tier and is positive only following a tightening—a bank does not incur it merely for occupying a higher statutory tier, δ_e is the quarterly equity depreciation rate capturing operational costs and franchise decay, $r_{d,t}$ is the retail deposit rate, $r_{W,t}$ is the wholesale funding rate, and $\tau_t e_{i,t}$ represents a time-varying systemic insurance levy scaled to the bank's equity footprint to fund entrant capital injections. The optimization space is strictly bounded by the size-dependent macroprudential regulatory capital boundary:

$$\omega_M A^M_{i,t} + \omega_B A^B_{i,t} \leq \frac{e_{i,t}}{\kappa_{i,t}(\text{Scale}_{i,t})} \tag{20}$$

where the regulatory threshold environment $\kappa_{i,t}(\cdot)$ is modeled as a step function dictated by total assets at the start of the period ($\text{Scale}_{i,t} \equiv A^M_{i,t-1} + A^B_{i,t-1}$), ensuring regulatory cohort assignments are fixed during current portfolio rebalancing steps and avoiding sub-derivative discontinuities.

The first-order conditions yield forward-looking stochastic Euler equations that govern the optimal allocation of credit across sectors. Future streams are discounted using $\Lambda_{t,t+1}$ (equal to one under the maintained risk neutrality of bank owners). Let $\mu_{i,t} \geq 0$ represent the non-negative Kuhn-Tucker multiplier on the dividend non-negativity constraint ($c_{i,t} \geq 0$), and let $\lambda_{i,t} \geq 0$ represent the multiplier associated with the regulatory capital frontier. The optimal credit allocation rules explicitly incorporate the risk-weight wedge ($\omega_B > \omega_M$) alongside the marginal cost of external wholesale funding:

$$\beta \mathbb{E}_t \left[\Lambda_{t,t+1}(1+\mu_{i,t+1}) \left(r_{M,t} - \varepsilon^M_{i,t+1} - \psi_M A^M_{i,t} - r_{W,t} - \phi_M(\kappa_{i,t}) - \gamma_i d^W_{i,t} \right) \right] = \lambda_{i,t} \omega_M \tag{21}$$

$$\beta \mathbb{E}_t \left[\Lambda_{t,t+1}(1+\mu_{i,t+1}) \left(r_{B,t} - \varepsilon^B_{i,t+1} - \psi_B A^B_{i,t} - r_{W,t} - \phi_B(\kappa_{i,t}) - \gamma_i d^W_{i,t} \right) \right] = \lambda_{i,t} \omega_B \tag{22}$$

Looking intertemporally, the shadow value of internal equity resolves into an intuitive asset pricing equation:

$$1 + \mu_{i,t} = \frac{\lambda_{i,t}}{\kappa_{i,t}} - \tau_t + \left(1 - \delta_e + r_{W,t} + \gamma_i d^W_{i,t} \right) \cdot \beta \mathbb{E}_t \left[\Lambda_{t,t+1}(1+\mu_{i,t+1}) \right] \tag{23}$$

where both multipliers satisfy standard complementarity slackness conditions:

$$\lambda_{i,t} \geq 0, \quad \frac{e_{i,t}}{\kappa_{i,t}} - (\omega_M A^M_{i,t} + \omega_B A^B_{i,t}) \geq 0, \quad \lambda_{i,t} \left[\frac{e_{i,t}}{\kappa_{i,t}} - (\omega_M A^M_{i,t} + \omega_B A^B_{i,t}) \right] = 0 \tag{24}$$

$$\mu_{i,t} \geq 0, \quad c_{i,t} \geq 0, \quad \mu_{i,t} c_{i,t} = 0 \tag{25}$$

Finally, to maintain a stationary cross-sectional measure μ_t and account for severe idiosyncratic realizations, banks whose equity falls below zero ($e_{i,t} < 0$) are resolved. They exit the market and are immediately replaced by a new entrant endowed with a small positive equity stock. To ensure aggregate accounting integrity within the financial sector, the total equity injected into these new entrants is funded via the insurance levy τ_t deducted pro-rata from the aggregate dividend distributions of surviving solvent financial intermediaries, capturing a systemic internal restructuring pool.

4.4 Deep Learning Solution Algorithm

To solve the dynamic system defined by equations (21)–(25) across an infinite-dimensional state space, we parameterize the individual choice rules and the structural shadow price functions using deep neural networks. Conventional value-function iteration (VFI) is infeasible here because the state space is high-dimensional—individual states $(e_{i,t}, A^M_{i,t-1}, A^B_{i,t-1})$ plus the compressed cross-sectional distribution Ω_t—and the occasionally binding capital and dividend constraints introduce kinks that break the smoothness assumptions of projection methods. The neural-network approach can be viewed as a global weighted-residual method in which the network parameterises the residual function and stochastic gradient descent replaces the fixed collocation grid, inheriting the global accuracy guarantees of weighted-residual methods while scaling to higher-dimensional problems (Maliar *et al.*, 2021; Azinovic *et al.*, 2022). Online Appendix Table OA.II reports robustness to network architecture (depth, width, activation function); training time for the baseline 4×256 architecture is approximately 6 hours on an NVIDIA A100 GPU. Let $S_{i,t} \equiv (e_{i,t}, A^M_{i,t-1}, A^B_{i,t-1}, \Omega_t)$ assemble the comprehensive vector of individual and aggregate states.

We parameterize a multi-layer perceptron (MLP) mapping function $\mathcal{A}(S_{i,t}; \Theta)$ governed by optimization weights and biases Θ. This neural network simultaneously approximates the optimal multi-asset policy rules for the current period and the non-linear regulatory shadow cost schedules:

$$\begin{bmatrix} \hat{A}^M_{i,t}(S_{i,t}; \Theta) \\ \hat{A}^B_{i,t}(S_{i,t}; \Theta) \\ \hat{c}_{i,t}(S_{i,t}; \Theta) \\ \hat{\lambda}_{i,t}(S_{i,t}; \Theta) \\ \hat{\mu}_{i,t}(S_{i,t}; \Theta) \end{bmatrix} = \mathcal{A}(S_{i,t}; \Theta) \tag{26}$$

Following the sufficiency argument of Section 4, the aggregate state entering $S_{i,t}$ is the market-clearing rate pair together with the current requirement, $\Omega_t = (r_{M,t}, r_{B,t}, \kappa_t)$, rather than a compressed description of μ_t. Market clearing is imposed analytically at each forward pass via the closed-form equilibrium rate (30), which replaces the numerical inner loop of Maliar *et al.* (2021) with an exact solution and removes the need for a separately parameterized perceived law of motion. The cross-sectional distribution is recovered as an *output* of the solved policy by simulating the panel forward under the exact law of motion, and is reported in Section 5.

The optimal parameters Θ^* are found by minimizing a structural loss that sums the squared Euler-

equation residuals for both asset classes and equity, plus a soft-penalty term enforcing the Kuhn-Tucker complementarity conditions (24)–(25). Market clearing in each credit sector is imposed analytically via the closed-form equilibrium rate in equation (30), replacing the numerical inner loop of Maliar *et al.* (2021) with an exact solution at each forward pass. The network is optimized globally via AdamW with cosine annealing. The full loss function, the complementary-slackness penalty derivation, and a detailed implementation summary are reported in the Online Appendix.

4.5 Data and Estimation Strategy

We parameterize and estimate the quantitative model using a unified hybrid two-step strategy to anchor the structural equations within empirical macro-finance matrices.

4.5.1 Calibration of Structural Parameters

Standard macroeconomic constants (e.g., household risk aversion, core asset depreciation rates, and baseline Basel III regulatory weights $\omega_M = 0.50$ and $\omega_B = 1.00$) are calibrated at a quarterly frequency to standard benchmark values from the banking literature. The quadratic adjustment cost scale parameters (ψ_M, ψ_B) are explicitly calibrated to target the long-run historical balance sheet portfolio concentrations documented within our Call Report sample.

4.5.2 Deep Simulated Method of Moments (SMM)

The remaining structural parameters governing the stochastic policy environment—specifically the persistence and volatility parameters of the macroprudential capital policy process $(\rho_\kappa, \sigma_\kappa)$—are structurally estimated via the *Simulated Method of Moments* (SMM). We model κ_t as an AR(1) process, the simplest specification that accommodates the near-unit-root persistence visible in the cumulated shock series while remaining parsimoniously identified from the sixteen LP coefficients. A richer specification—such as a Markov regime-switching process distinguishing tightening episodes from quiescent periods—would be economically appealing, but the two regimes would introduce at least four additional parameters (persistence, volatility, and transition probabilities for each regime). With an SMM objective built on sixteen local-projection coefficients that are themselves noisy, these parameters are not separately identified at conventional significance levels; we therefore maintain the parsimonious AR(1) and note that the near-unit-root estimate ($\hat{\rho}_\kappa \approx 0.99$) already captures the high persistence that a regime-switching

specification would attribute to the tightening state.

Following the standard approach in the banking literature, the standard deviation of idiosyncratic credit depreciation shocks, σ_ε, is calibrated externally to isolate the microeconomic baseline of asset quality, thereby avoiding identification multi-collinearity during global optimization. Specifically, σ_ε is fixed to the time-series average of the cross-sectional standard deviation of quarterly net charge-off rates by asset class, computed from the full bank panel and demeaned by the quarterly aggregate to cleanly isolate the idiosyncratic component. This calibration represents the natural empirical counterpart to our structural model's idiosyncratic shock setup following Mendicino *et al.* (2020).

Let $\theta \equiv (\rho_\kappa, \sigma_\kappa)$ collect the two structural parameters slated for estimation. The SMM estimator searches for the parameter vector that minimizes the distance between the empirical and simulated targets:

$$\hat{\theta} = \arg\min_{\theta} \left[\hat{m}_{\text{data}} - m_{\text{model}}(\theta)\right]' \mathbf{W} \left[\hat{m}_{\text{data}} - m_{\text{model}}(\theta)\right] \tag{27}$$

where $\hat{m}_{\text{data}}$ is a sixteen-dimensional empirical target vector comprising the within-bank local-projection coefficients from Section 2 for both C&I and real estate lending at horizons $\mathcal{H} = \{1, 2, 4, 6, 8, 12, 16, 20\}$ quarters (8 horizons $\times$ 2 sectors = 16 moments). The target therefore spans the full dynamic profile of the sectoral lending channel—both the higher-risk-weight C&I segment and the lower-risk-weight real estate segment—at short, medium, and long horizons.

Heuristically, the two parameters within θ map into complementary features of this sixteen-dimensional target. The policy persistence ρ_κ is primarily identified by the decay rate of the credit contraction across medium-to-long horizons in both sectors; a more persistent shock sustains the regulatory shadow cost $\lambda_{i,t}$ across multiple periods, generating slower impulse response function (IRF) reversion. Conversely, the policy volatility scale σ_κ is pinned down by the short-horizon amplitude of the IRFs, which scales proportionally with the impact-period size of the regulatory innovation. Using both C&I and real estate moments simultaneously disciplines the model's cross-sectoral allocation: the risk-weight wedge $\omega_B > \omega_M$ must generate a differential contraction consistent with the data at every horizon.

The structural weighting matrix $\mathbf{W}$ is the optimal variance-covariance matrix derived via bootstrapping our empirical sample, block-diagonal across the C&I and real estate blocks. For each candidate evaluation vector θ, the deep learning policy network is re-trained to compute the sixteen model-implied conditional moments $m_{\text{model}}(\theta)$, maintaining rigorous consistency between structural parameterization and the global solution paths.

4.6 Loan Demand Schedules and Partial Equilibrium Rates

To close the partial equilibrium model and compute the market-clearing asset returns $(r_{M,t}, r_{B,t})$ that appear in the bank optimisation problem (21)–(25), we introduce reduced-form inverse loan demand schedules for each credit market $j \in \{M, B\}$. The focus of this paper is on the supply-side transmission of capital requirement shocks; the demand block is therefore kept parsimonious, consistent with the partial equilibrium scope of the model.

4.6.1 Inverse Demand Specification

We posit linear inverse demand functions of the form:

$$r_{j,t} \;=\; \theta_j \;-\; \eta_j \, L_t^j, \qquad j \in \{M, B\}, \tag{28}$$

where $L_t^j \equiv \frac{1}{N} \sum_{i=1}^{N} \hat{A}_{i,t}^j$ denotes aggregate lending in sector j per unit of banking mass, $\theta_j > 0$ is the demand intercept (the choke rate when aggregate lending is zero), and $\eta_j > 0$ is the inverse demand elasticity (slope of the inverse demand curve). This specification admits a standard linear demand interpretation: borrowers' willingness to pay for credit falls with aggregate loan supply, so that a contraction in bank lending following a capital shock generates a partial offsetting rise in loan rates.[9]

Market clearing requires that aggregate loan supply equals aggregate loan demand in each sector at every date t:

$$\frac{1}{N} \sum_{i=1}^{N} \hat{A}_{i,t}^j(\mathbf{S}_{i,t}; \Theta) \;=\; L_t^j(r_{j,t}), \qquad j \in \{M, B\}, \tag{29}$$

where $\hat{A}_{i,t}^j(\mathbf{S}_{i,t}; \Theta)$ are the neural-network policy functions defined in Section 4.4. Substituting (28) into (29) and solving for the equilibrium rate yields:

$$r_{j,t}^* \;=\; \theta_j \;-\; \eta_j \cdot \frac{1}{N} \sum_{i=1}^{N} \hat{A}_{i,t}^j(\mathbf{S}_{i,t}; \Theta). \tag{30}$$

[9] The linearity assumption is a deliberate simplification. Because identification rests on the supply-side capital-requirement shock ε_t^κ rather than demand innovations, the functional form of $L_t^j(\cdot)$ does not affect the causal estimates reported in Section 2. Its role is limited to (i) closing the market-clearing block of the quantitative model and (ii) enabling the welfare calculations in Section 4.7.

Equation (30) is evaluated inside the deep learning forward pass at each training iteration: given the current network weights Θ, aggregate supply is computed, the equilibrium rate is updated, and the updated rate feeds back into the individual Euler conditions (21)–(22) in the subsequent iteration.

4.6.2 Calibration of Demand Parameters

The four demand parameters $(\theta_M, \eta_M, \theta_B, \eta_B)$ are calibrated externally rather than estimated, consistent with the partial equilibrium scope of the model. We proceed in two steps.

Step 1: Steady-state intercepts θ_j. Let $\bar{L}^j$ and $\bar{r}_j$ denote the long-run sample averages of aggregate lending and the implied lending rate in sector j, computed from the Call Reports panel over 1994–2019. The steady-state condition requires $\bar{r}_j = \theta_j - \eta_j \bar{L}^j$, so that once η_j is fixed the intercept is pinned down by:

$$\theta_j \;=\; \bar{r}_j \;+\; \eta_j \, \bar{L}^j, \qquad j \in \{M, B\}. \tag{31}$$

Step 2: Demand elasticities η_j. The elasticities are fixed at values drawn from the external literature. No direct estimate of loan demand is available for either sector, so we use the closest reduced-form magnitudes we are aware of: for mortgages the interest-rate sensitivity reported by Adelino *et al.* (2025), and for commercial lending the credit response to policy rates estimated by Jiménez *et al.* (2014). Neither is a loan-demand schedule, and we return below to what that implies for the results that depend on it. Formally:

$$\eta_j \;=\; \frac{1}{S_j \, \bar{L}^j}, \qquad S_j \;\equiv\; -\frac{\partial \log L^j}{\partial r_j}, \tag{32}$$

where S_j is the semi-elasticity of loan demand with respect to the lending rate, taken from the respective reduced-form estimates. Equation (32) is the inversion of (28): since $\partial L^j / \partial r_j = -1/\eta_j$, the semi-elasticity is $S_j = 1/(\eta_j \bar{L}^j)$, so the slope is the reciprocal of the semi-elasticity scaled by average lending.[10] Two

[10]The slopes reported in Table IV imply semi-elasticities of 32.5 (mortgages) and 14.8 (C&I) under (32), against the targets of 2 and 4 cited in the same table; the implemented demand schedules are therefore more elastic than the external targets by factors of roughly 16 and 4. Matching the cited targets would require $\eta_M = 1.06\text{e-}6$ and $\eta_B = 1.39\text{e-}6$. Because η_j enters only through the market-clearing block and the welfare calculation, and not through the reduced-form estimates of Section 2 or the cross-sectional moments of Section 5, the sign and cross-sectional pattern of the results are unaffected; the magnitude of the surplus loss is not. We flag this explicitly rather than absorb it, and note that the relevant range is a

properties of this parameterisation should be kept separate. The lending responses are *exactly* invariant to η_j: the demand block enters neither the bank's optimisation problem nor the reduced-form estimates of Section 2, so the impulse responses, the buffer gradient and the sectoral ordering are numerically identical at every value of η_j we consider. The *cardinal* welfare cost is not, because the surplus integral is taken under the demand curve and its level therefore depends on that curve's slope: across the range spanned by the implemented schedules and the externally cited targets, Λ falls from 1.4% to 0.3% (Online Appendix OA.H). We accordingly read the welfare figure as a benchmark conditional on the demand calibration, and note that the external estimates underlying it measure objects adjacent to a loan-demand schedule—a house-price elasticity, and a loan-*granting* probability against the policy rate—rather than the schedule itself. Estimating η_j directly on loan-level data is left to future work.

Step 3: The cost of regulatory capital (c, z). The two parameters of ϕ_j in (10) are set on the two sectoral aggregate lending responses to a tightening—-0.170% in C&I and -0.110% in real estate—evaluated as a comparative static at predetermined equity across the observed distribution of capital ratios. This yields $c = 0.148$ and $z = 0.021$: per percentage point of requirement, a cost of about 15 basis points on every unit of credit extended and a further 2 basis points per unit of *risk-weighted* assets. Both buffer gradients are left untargeted and serve as the out-of-sample check reported in Section 5.3.

The risk-weighted component carries the standard reading: z is the premium on funding a unit of risk-weighted assets with equity rather than debt, and the estimate corresponds to an equity–debt spread of roughly ten percent annualised, within the range the corporate-finance literature reports. The non-risk-weighted component is the larger of the two, $c/z \approx 7$, and it is the one the cross-section requires: as Section 3 shows, a cost that enters only through risk weights leaves banks away from the frontier exactly unresponsive. We read c as standing in for a leverage-type requirement absent from (20), but the data identify its magnitude rather than its source, and any per-asset cost rising with regulatory stringency would fit equally well.

These two parameters discipline the comparative-static exercise of Section 5.3. The dynamic solve sets $c = z = 0$, as noted in the discussion of solution accuracy below, so the impulse responses and the welfare calculations are unaffected by them.

factor of 16 in η_M, not the factor of four a symmetric sensitivity band would span; Online Appendix OA.H reports it in full.

Table IV
Calibration and Estimation Summary

Parameter	Description	Value	Source / Target
Panel A: Household and Macro			
β	Quarterly discount factor	0.99	Standard; annualised $\beta^4 \approx 0.96$
σ	Household CRRA coefficient (SDF)	1.00	Log utility; standard in banking literature
Panel B: Bank Technology			
δ_e	Quarterly equity depreciation rate	0.0010	Dynamic consistency: equity-Euler stationarity at $\beta = 0.99$
ψ_M	Mortgage portfolio operating cost	0.0073	Calibrated: targets RE loan share ≈ 0.69 (Table I)
ψ_B	C&I portfolio operating cost (baseline)	0.05	Preset; admissible range (0.014, 0.142), sensitivity at {0.02, 0.10}
c	Cost of regulatory capital, not risk-weighted	0.1484	Eq. (10); set on the aggregate C&I and RE responses
z	Cost of regulatory capital, per unit of RWA	0.0212	Eq. (10); jointly with c; buffer gradients untargeted
Panel C: Regulatory Capital			
ω_M	Residential mortgage risk weight	0.50	Basel standardised approach; Section 4
ω_B	Commercial & industrial loan risk weight	1.00	Basel standardised approach; Section 4
κ_{base}	Baseline PCA capital minimum	0.10	Prompt Corrective Action well-capitalised threshold
$\bar{\kappa}_{\text{DFA}}$	DFA add-on ($10B–$50B banks)	0.015	Fed enhanced prudential standards; post-2010Q3
$\bar{\kappa}_{\text{SIFI}}$	Enhanced-standards add-on (>$50B banks)	0.025	Stylised; see Section 4
Panel D: Funding and Liability			
γ_{large}	Wholesale funding friction, low-friction banks (calibrated to large-bank averages)	≈ 0	Paper assumption; near-frictionless capital market access
γ_{small}	Wholesale funding friction, high-friction banks (calibrated to small-bank averages)	0.0769	Calibrated: empirical funding structure (Table III)
$\bar{d}_C$	Core-deposit capacity (USD thousands)	447 ,499	Balance-sheet identity at the calibrated steady state
τ_t	Systemic insurance levy rate	0.0002	FDIC DIF assessment (quarterly decimal)
Panel E: Loan Demand (Option B, Section 4.6)			
η_M	Inverse demand slope, mortgages	6.497 e-8	Eq. (32); Adelino *et al.* (2025) semi-elasticity $= -2$
η_B	Inverse demand slope, C&I loans	3.773 e-7	Eq. (32); Jiménez *et al.* (2014) semi-elasticity $= -4$
θ_M	Mortgage demand intercept	0.0462	Eq. (31): $\bar{r}_M + \eta_M \bar{L}_M$; $\bar{r}_M = 0.0154$, $\bar{L}_M = 473{,}578$
θ_B	C&I loan demand intercept	0.0847	Eq. (31): $\bar{r}_B + \eta_B \bar{L}_B$; $\bar{r}_B = 0.0169$, $\bar{L}_B = 179{,}618$
Panel F: Idiosyncratic Credit Shocks (externally calibrated)			
$\bar{\sigma}_{\varepsilon,M}$	SD of idiosyncratic RE charge-off rate	0.0010	Online Appendix; cross-sectional SD of demeaned $\varepsilon^M_{i,t}$, 1994–2019
$\bar{\sigma}_{\varepsilon,B}$	SD of idiosyncratic C&I charge-off rate	0.0041	Online Appendix; cross-sectional SD of demeaned $\varepsilon^B_{i,t}$, 1994–2019
Panel G: Macroprudential Policy Process (SMM-estimated)			
ρ_κ	Persistence of capital requirement innovation	0.9987 (0.0038)	SMM: medium-horizon persistence of the within-bank IRFs
σ_κ	Volatility of capital requirement innovation	0.00040 (0.00002)	SMM: amplitude of the within-bank IRFs
Panel H: Neural Network Architecture (Section 4.4)			
L	Number of hidden layers	4	MLP; Section 4.4
d	Nodes per hidden layer	256	MLP; Section 4.4
activation	Activation function	Swish	Gradient stability near regulatory thresholds
optimizer	Training algorithm	AdamW	Cosine annealing learning rate schedule

Notes: All dollar-denominated variables are in thousands of USD (as reported in the Call Reports) at quarterly frequency. Interest rates are quarterly decimals (annual percentages divided by 400). Risk weights (ω_M, ω_B) and capital requirements (κ) follow the Basel III standardised approach. The regulatory capital minimum $\kappa_{i,t}$ is a size-dependent step function of lagged total assets (Section 4): κ_{base} for assets below $10B, $\kappa_{\text{base}} + \bar{\kappa}_{\text{DFA}}$ for $10B–$50B, and $\kappa_{\text{base}} + \bar{\kappa}_{\text{SIFI}}$ above $50B. Loan demand intercepts (θ_M, θ_B) are the choke rates at which aggregate demand equals zero, derived analytically from steady-state moments and calibrated elasticities via Eq. (31). Parameters marked "—" are calibrated numerically within the model to match the empirical targets stated in the Source column (Section 4.5.1). $\bar{\sigma}_{\varepsilon,j}$ is calibrated externally following Mendicino *et al.* (2020) as the time-series mean of the within-period cross-sectional standard deviation of quarterly net charge-off rates, demeaned to isolate idiosyncratic variation from the common aggregate component. The SMM weighting matrix is $\mathbf{W} = \hat{\Sigma}^{-1}$, block-diagonal across the C&I and RE blocks, each block estimated by a cluster (bank-level) bootstrap of the within-bank LP coefficients (100 replications), for the impulse responses at horizons $\mathcal{H} = \{1, 2, 4, 6, 8, 12, 16, 20\}$; see the Online Appendix for the formal identification argument. Standard errors for the SMM estimates (in parentheses) are asymptotic GMM; ρ_κ is estimated in the interior of the admissible range ($\rho_\kappa < 0.999$, imposed so the requirement process is stationary), with 95% profile confidence set $\rho_\kappa \geq 0.994$ — a half-life of at least twenty-nine years. Sources: Adelino *et al.* (2025) for the residential mortgage demand semi-elasticity; Jiménez *et al.* (2014) for the C&I loan demand semi-elasticity; Mendicino *et al.* (2020) for the idiosyncratic shock calibration strategy; Drechsel & Miura 2025 for the capital-requirement shock series.

4.6.3 Aggregate Rate Pass-Through

A direct implication of equations (28)–(30) is that a capital-requirement tightening generates a partial rate pass-through effect: the reduction in aggregate loan supply pushes equilibrium rates upward even as individual banks reduce volumes. The aggregate rate response at horizon h satisfies:

$$\frac{\partial r^*_{j,t+h}}{\partial \varepsilon^\kappa_t} \;=\; -\,\eta_j \cdot \frac{\partial L^j_{t+h}}{\partial \varepsilon^\kappa_t}, \tag{33}$$

so that the magnitude of rate pass-through is proportional to the quantity contraction documented in Section 2.5 and scales with the inverse demand slope η_j. Because $|\partial L^B/\partial \varepsilon^\kappa| > |\partial L^M/\partial \varepsilon^\kappa|$ by the risk-weight mechanism, C&I lending rates rise more sharply on impact than mortgage rates, generating an asymmetric cost-of-credit effect across sectors that amplifies the real consequences of macroprudential tightening.

4.7 Steady State, Welfare, and Policy Counterfactuals

4.7.1 Stationary Equilibrium

A stationary recursive equilibrium of the model consists of: (i) a time-invariant policy function $\mathbf{A}(\mathbf{S}_i; \Theta^*)$ mapping individual states to credit allocations, dividends, and shadow costs; (ii) a stationary distribution μ^* over individual bank states (e_i, A^M_i, A^B_i); and (iii) constant equilibrium rates (r^*_M, r^*_B) such that the market-clearing conditions (29) hold at μ^*. In the stationary equilibrium the regulatory capital requirement is fixed at its long-run baseline $\kappa^* = \kappa_{\text{base}}$, and the insurance levy τ^* is set to balance the entrant recapitalisation fund. The stationary distribution μ^* is obtained as the ergodic fixed point of the law of motion $\Omega_{t+1} = \Gamma_\Theta(\Omega_t, Z_{t+1}, \kappa_{t+1})$ evaluated at $\kappa_{t+1} = \kappa^*$ and $Z_{t+1} = 0$.

4.7.2 Welfare Measure

We measure the aggregate welfare cost of a capital-requirement shock using a consumption-equivalent variation (CEV), the proportional reduction in the value of intermediation that leaves welfare unchanged at the original requirement.

We report two value concepts, and we lead with the one the paper's headline numbers come from. Our baseline measures the credit-market surplus generated by intermediation—the area under each

sector's inverse loan demand schedule above the marginal funding cost—which falls unambiguously when credit contracts. Writing aggregate sector-j lending as $L_j(\kappa)$ and using the inverse demand schedule $r_j = \theta_j - \eta_j L_j$ of equation (28), that surplus is

$$C(\kappa) \;=\; \sum_{j\in\{M,B\}} \left[\left(\theta_j - r_W\right) L_j(\kappa) \;-\; \tfrac{\eta_j}{2} L_j(\kappa)^2\right], \tag{34}$$

and the headline Λ is the proportional loss of that surplus following a permanent one-percentage-point tightening from κ^* to $\kappa^* + 0.01$,

$$\Lambda \;=\; \frac{C(\kappa^*) - C(\kappa^* + 0.01)}{C(\kappa^*)}. \tag{35}$$

The demand block (θ_j, η_j) is constructed once at baseline lending and held fixed across the comparison, so both points are evaluated on a single demand curve; Online Appendix OA.H reports the sensitivity of Λ to the calibrated semi-elasticities that pin it down.

We decompose Λ into a *sectoral allocation cost* Λ^{alloc}, arising from the distortion in the C&I-to-mortgage lending ratio induced by the risk-weight wedge ($\omega_B > \omega_M$), and a *scale cost* Λ^{scale}, arising from the contraction in total credit supply:

$$\Lambda \;=\; \underbrace{\Lambda^{\text{alloc}}}_{\text{sectoral distortion}} \;+\; \underbrace{\Lambda^{\text{scale}}}_{\text{aggregate credit contraction}} \;+\; \mathcal{O}\left(\Delta\kappa^2\right). \tag{36}$$

This decomposition maps directly onto the two structural channels identified in Section 3: the Asset Substitution Margin generates Λ^{alloc}, while the Scale Deleveraging Multiplier generates Λ^{scale}.

As a robustness check we compute the CEV from the bank franchise value instead. Let V^* denote the aggregate franchise value in the stationary equilibrium and V^κ the corresponding value after the same tightening. The franchise-value CEV Λ_V satisfies

$$V^\kappa(\mu^*;\, \kappa^* + 0.01) \;=\; V^*\left(\left(1 - \Lambda_V\right)\mu^*;\, \kappa^*\right), \tag{37}$$

so that $\Lambda_V > 0$ is the fraction of steady-state bank dividends that households would be willing to forgo in order to avoid the regulatory tightening. Because dividend flows map to household income in this partial-equilibrium framework, Λ_V is a welfare measure in the same units as Λ, computed from a different object: bank-owner value rather than credit-market surplus. We compute it as the expected discounted

stream of bank dividends along a stabilised forward simulation of the estimated policy, with dividends given by the stationary residual of the equity law of motion. The two metrics agree in sign and magnitude (Section 5), bracketing the welfare object. We treat the surplus measure as the headline and the franchise value as the robustness check, for a reason worth making explicit: because δ_e is set so that the equity Euler equation is stationary at $\beta = 0.99$, bank owners are close to indifferent about the timing of payouts, and the dividend policy is therefore only weakly pinned down by the equilibrium conditions. Perturbing the solved dividend rule by $\pm 25\%$ changes the model's Euler residuals by less than 0.01%. The surplus measure is computed from lending volumes and the demand schedules and does not depend on the dividend rule; the franchise value does, and inherits that indeterminacy.

Three caveats delimit the exercise. First, the framework is partial equilibrium: dividend and surplus flows map to household income, but there is no feedback to factor prices or aggregate demand. Second, the model assigns no benefit to capital—default and crisis-prevention channels are absent—so Λ is a *gross* cost of tightening, not a net policy evaluation; in particular, the dividend externalities emphasised by Acharya *et al.* (2017)—whereby individually optimal capital distributions erode systemic resilience—are precisely the kind of benefit-side channel our framework is silent on. Third, the CEV is evaluated at the estimated, effectively permanent innovation process; transitory experiments would yield smaller costs.

Several general-equilibrium channels, absent from the framework, could attenuate or amplify the partial-equilibrium supply contractions measured here. On the demand side, borrowers denied bank credit may substitute toward non-bank financial intermediaries—shadow banks, private credit funds, or bond markets—partially offsetting the real-side impact of the bank-lending contraction (see Buchak *et al.*, 2018, for evidence that fintech and shadow-bank lenders absorbed a large share of the post-crisis decline in bank mortgage origination, though substitution is likely smaller for relationship-intensive C&I credit). On the supply side, equilibrium loan rates would adjust: if aggregate loan supply falls, the clearing rate rises, crowding out marginal borrowers but also raising bank spreads, which partly restores lending incentives. Finally, the credit contraction could feed back into aggregate demand through reduced investment and employment, further depressing loan demand in a contractionary spiral. The net direction is ambiguous and depends on the elasticity of borrower substitution and the strength of the aggregate-demand multiplier. By holding these channels fixed, our partial-equilibrium estimates isolate the pure supply mechanism and provide a transparent upper bound on the banking-sector adjustment cost.

4.7.3 Policy Counterfactuals

We evaluate two counterfactual policy experiments that exploit the heterogeneous transmission documented in Sections 3 and 4.

Counterfactual 1: Uniform vs. size-differentiated capital requirements. The baseline step function $\kappa_{i,t}(\text{Scale}_{i,t})$ in equation (20) assigns higher capital requirements to larger institutions. We compare this to a counterfactual uniform requirement $\tilde{\kappa}$ that raises the same aggregate regulatory capital in the stationary equilibrium:

$$\tilde{\kappa} \;\text{ s.t. }\; \int \tilde{\kappa}\, e_i \, \mathrm{d}\mu^* \;=\; \int \kappa_i(\text{Scale}_i)\, e_i \, \mathrm{d}\mu^*. \tag{38}$$

A comparison of the welfare cost Λ under the baseline size-differentiated rule and the uniform rule $\tilde{\kappa}$ identifies the efficiency gain (or loss) from regulatory tailoring. Given the Insulated Zone mechanism documented in Section 3.5, we expect the uniform rule to generate a larger aggregate C&I credit contraction relative to the size-differentiated baseline, because it removes the precautionary buffer advantage of high-friction institutions that currently benefit from lighter size-tier surcharges. Because the dynamic model is solved under the baseline requirement (Section 4), this counterfactual compares stationary equilibria under the two schedules rather than re-solving the full step function along the transition; the approximation affects the cross-sectional *incidence* of the comparison, not the aggregate scale result, which is pinned down by the level of aggregate capital that equation (38) holds fixed by construction.

Counterfactual 2: Risk-weight flattening. We evaluate the effect of reducing the risk-weight wedge by setting $\tilde{\omega}_B = \tilde{\omega}_M = 0.75$ (the arithmetic average of the baseline weights), holding κ fixed at κ^*. This experiment isolates the contribution of the sectoral distortion channel Λ^{alloc} to total welfare costs by shutting down the Asset Substitution Margin while preserving the scale effect. The counterfactual aggregate credit contraction under flattened weights satisfies:

$$\frac{\partial \tilde{L}^B_{t+h}}{\partial \varepsilon^{\kappa}_t} \;\approx\; \frac{\partial \tilde{L}^M_{t+h}}{\partial \varepsilon^{\kappa}_t} \quad \forall\, h, \tag{39}$$

so that sectoral IRF asymmetry disappears. The difference between the baseline and flattened-weight welfare costs provides a structural estimate of the dollar value of the sectoral distortion created by

differential risk-weighting under Basel III.

5 Quantitative Results

5.1 Estimation: matching the within-bank impulse responses

We estimate the policy process $\theta = (\rho_\kappa, \sigma_\kappa)$ by simulated method of moments on the within-bank local-projection coefficients of Section 2 at horizons $\mathcal{H} = \{1, 2, 4, 6, 8, 12, 16, 20\}$ for C&I and real estate lending (16 moments). The model-implied moments are computed by simulating the estimated policy along a panel impulse path: the capital requirement jumps by σ_κ and decays at rate ρ_κ, and the response propagates through the portfolio adjustment lags and the equity law of motion, so the model impulse responses inherit dynamics from those lags rather than from the requirement path alone. They are not, however, hump-shaped: with ρ_κ estimated at effectively unity the requirement never returns, so the simulated response deepens monotonically after an initial impact and partial rebound rather than reverting like its empirical counterpart. Because the only aggregate disturbance in the model is the requirement innovation, the local-projection coefficient on the simulated panel coincides with the impulse response.

Figure 12 and Table V report the fit. The estimated innovation is small and effectively permanent ($\hat{\sigma}_\kappa = 0.00040$, s.e. 0.00002; $\hat{\rho}_\kappa = 0.9987$, s.e. 0.0038, interior to the imposed stationarity cap of 0.999; 95% profile confidence set $\rho_\kappa \geq 0.994$, so at least 79% of a tightening survives a decade), consistent with the narrative content of the identified series: post-crisis capital regulation changes were regime changes, not transitory shocks. The model reproduces the sign of all sixteen moments and the sectoral ordering—C&I contracts more than real estate at every horizon. This fit is conditional on the solved policy (Online Appendix OA.G). It does not reproduce the hump: the simulated response is largest on impact and then deepens without reverting, while the empirical responses trough at three to four years and return toward zero. With the requirement process estimated as effectively permanent, a reverting response is out of reach. Formally, the Hansen J-statistic rejects the two-parameter process ($J = 1822$ under the bootstrap weighting matrix, 14 degrees of freedom). This is a mechanical consequence of the precision with which the LP coefficients are estimated on a panel of roughly 4,800 banks: the optimal weighting matrix assigns very large weight to even economically small deviations, so the J-test has high

power against any parsimonious structural model.[11] The economically relevant misfit is concentrated in the depth of the mid-horizon troughs: the model generates a C&I-to-RE response ratio of roughly 1.2, against approximately 1.4 in the data, so the pure risk-weight wedge accounts for the ordering and the bulk—though not all—of the observed sectoral asymmetry. The residual asymmetry is not an artifact of the preset C&I adjustment cost: re-deriving the calibration under alternative ψ_B values shows, first, that the steady-state targets jointly discipline (ψ_M, ψ_B, γ)—only $\psi_B \in (0.014, 0.142)$ admits positive frictions—and second, that the constrained-regime sectoral asymmetry $s_B - s_M$ remains large and negative across the admissible range (from -3.5 at $\psi_B = 0.02$ to -0.8 at $\psi_B = 0.10$; -2.2 at the baseline), while flattening the risk weights collapses it to zero at any ψ_B. The asymmetry is a risk-weight phenomenon, with ψ_B governing only its amplitude.

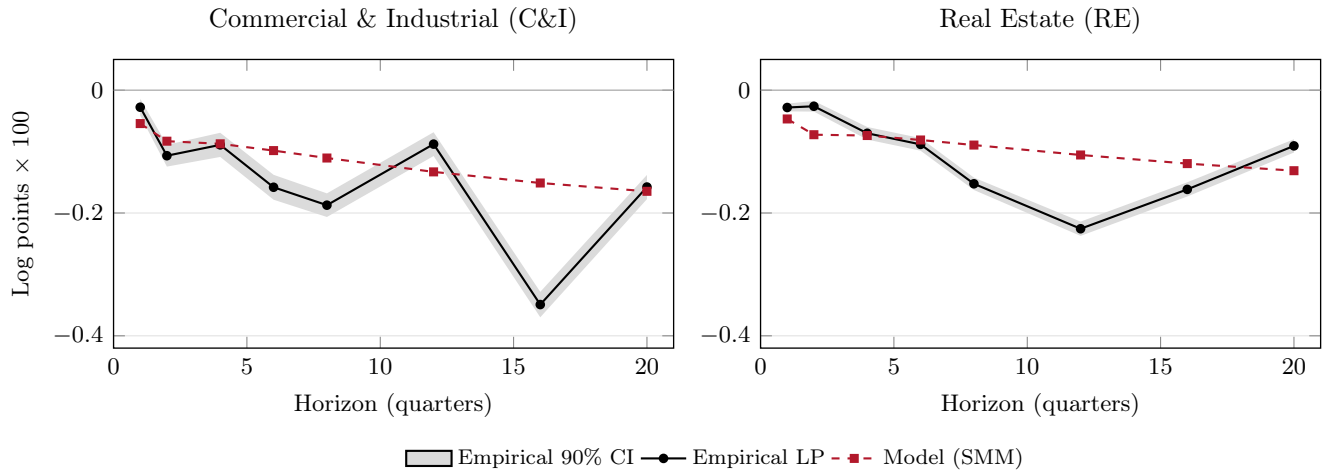


FIGURE 12.—Sectoral credit responses: model vs. within-bank local projections

Notes: Empirical within-bank local projections on the bank panel (bank fixed effects, standard errors clustered by bank; 90% confidence bands shaded) versus the SMM-estimated model's simulated impulse path. Horizons in quarters; responses in log points (×100) per unit of the identified capital-requirement innovation.

[11]The same tension between statistical rejection and economic fit arises whenever SMM or GMM targets are estimated on large micro panels. Bloom (2009) matches firm-level investment moments with a small set of uncertainty parameters and reports formal rejection alongside close economic fit; David *et al.* (2016) encounter the identical issue fitting bilateral trade and asset-pricing moments from cross-country panels. In each case the J-test rejects because the moments are estimated too precisely for the model to match exactly, not because the model misses the economically relevant patterns.

TABLE V
SMM MOMENT FIT

Sector	Horizon h	$\hat{b}_h$ (data)	s.e.	b_h (model)	t
C&I	1	-0.028	0.008	-0.054	+3.20
	2	-0.107	0.011	-0.083	-2.20
	4	-0.089	0.012	-0.087	-0.14
	6	-0.158	0.012	-0.098	-4.85
	8	-0.187	0.012	-0.110	-6.57
	12	-0.088	0.012	-0.133	+3.84
	16	-0.349	0.013	-0.151	-15.73
	20	-0.158	0.012	-0.165	+0.58
RE	1	-0.028	0.004	-0.047	+4.87
	2	-0.026	0.005	-0.073	+8.80
	4	-0.070	0.006	-0.074	+0.59
	6	-0.088	0.006	-0.081	-1.15
	8	-0.152	0.006	-0.089	-10.29
	12	-0.226	0.007	-0.106	-17.02
	16	-0.162	0.007	-0.120	-5.97
	20	-0.091	0.007	-0.131	+6.16
Hansen $J = 1822.15$, dof $= 14$, $\chi^2_{0.05} = 23.68$, $p = 0.000$					

Notes: Empirical within-bank local-projection coefficients $\hat{b}_h$ (log points $\times 100$) versus model-implied responses at the SMM estimate; t is the studentised residual. With panel-precise moments the J-test has high power; see the text for the economic reading of the rejection.

5.2 Solution accuracy

One point of correspondence should be stated before the diagnostics. The trained policies solve the model of Section 4 under $c = z = 0$, that is with the cost of regulatory capital ϕ_j of equation (10) switched off, so that the requirement reaches the bank only through the constraint multiplier. Every dynamic object we report—the impulse responses, Figure 8, the estimated requirement process and the welfare calculations—is computed from that solve. The consequences of $\phi_j \neq 0$ are evaluated separately, as a comparative static at predetermined equity across the observed distribution of capital ratios, and reported in Section 5.3 and Online Appendix OA.F. The two are different objects—an impulse-response path and a comparative static—and we do not treat the second as a re-estimate of the first. Solving the dynamic model with ϕ_j active requires re-training the policy network and is left to future work.

The stationary and κ-augmented policies are approximated with the deep-learning method of Section 4.4. Following best practice for global solution methods, Figure 13 maps the Euler-equation residuals across the state space rather than reporting only an average. Accuracy is highest in the interior of the buffer distribution and deteriorates toward the regulatory frontier, which is where the constraint-driven

nonlinearity operates; we therefore read the model's cross-sectional predictions near the frontier as conservative. Online Appendix Table OA.II demonstrates that these accuracy patterns are robust to the neural-network architecture: varying the number of hidden layers (2–6), width (128–512 neurons per layer), and activation function (ReLU, GELU, Tanh) leaves the median Euler-equation RMSE within ±0.02 log units of the baseline and does not alter the qualitative conclusions.

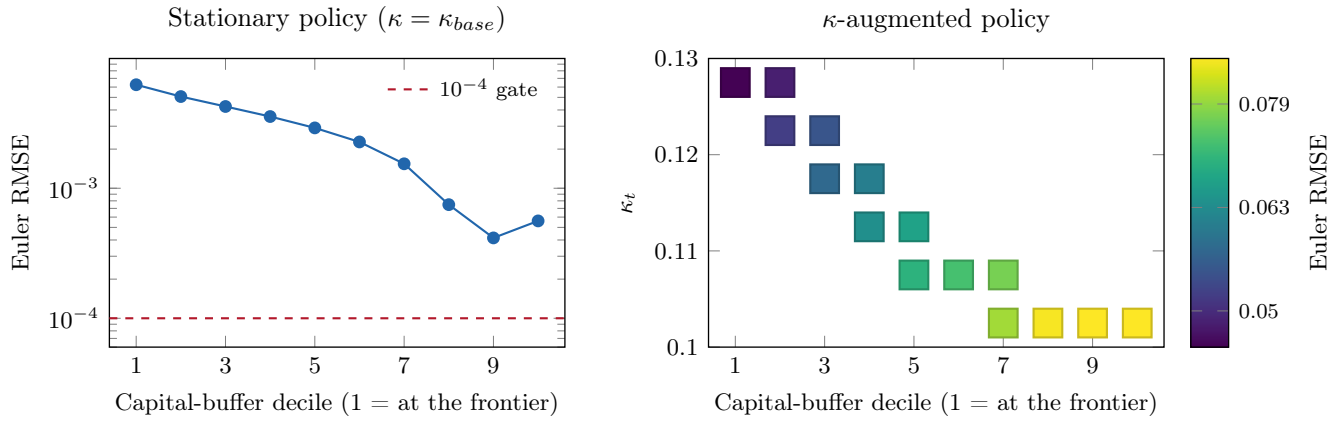


FIGURE 13.—Solution accuracy across the state space

Notes: Left: Euler-equation RMSE of the stationary policy by capital-buffer decile (decile 1 = banks at the regulatory frontier), evaluated on the policy's ergodic distribution; the dashed line marks the 10^{-4} accuracy gate. The spike near the frontier reflects the kink in optimal policy functions at the occasionally binding regulatory constraint, which is the hardest region for the neural-network approximation to resolve; all deciles remain below 10^{-2}. Right: residual map of the κ-augmented policy over the (buffer, κ_t) plane.

Three additional diagnostics, reported in full in Online Appendix OA.A.4, confirm the accuracy and reliability of the deep-learning solution. First, a train-versus-test Euler-error histogram shows virtually identical residual distributions across independent simulations (median $\log_{10}|e_{\text{REE}}| = -1.21$ in both), ruling out overfitting. Second, an ergodic-set coverage scatter shows that test-sample states drawn from the same process lie within the training cloud. The counterfactual exercises of Section 5 deliberately do not: they move banks toward a frontier that the historical process did not reach, and accuracy in that region is reported separately below. Third, the den Haan (2010) unit-free accuracy statistic achieves $R_{\text{dH}}^{\max} < 0.025\%$ for both credit sectors—well below the 0.1% threshold classified as "very accurate"—confirming that the neural-network policy satisfies the aggregate market-clearing conditions with high precision.

5.3 Incidence across banks

The paper asks who bears the burden of a tightening, and the model answers it directly. Evaluating the bank's problem across the observed distribution of capital ratios, at predetermined equity, and with the cost of regulatory capital of equation (10) active, Table VI reports each buffer tercile's share of the resulting contraction against its share of lending.

TABLE VI
INCIDENCE OF A TIGHTENING BY BUFFER POSITION

	share of lending	share of burden	concentration	ΔL
Thin buffer	32.1%	34.1%	1.06	−0.127%
Middle	32.9%	33.0%	1.00	−0.120%
Thick buffer	35.1%	33.0%	0.94	−0.113%

Notes: Buffer terciles are formed on regulatory slack measured against each bank's own statutory requirement. "Concentration" is the ratio of a tercile's share of the aggregate contraction to its share of lending; a value above one indicates that the tercile absorbs more of the tightening than its size alone would imply. The two parameters of ϕ_j are set on the two sectoral aggregate responses; the tercile gradients are untargeted.

Two readings, and they differ. Per dollar lent, thin-buffer banks contract 13% more than thick-buffer banks: the gradient is real and it is the object the cross-sectional evidence identifies. As a *share* of the aggregate contraction, however, incidence is close to proportional—thin-buffer banks bear 34.1% of the burden while holding 32.1% of lending. Because thin-buffer institutions are on average smaller, a disproportionate response per unit of credit translates into only a mildly disproportionate share of the total. We read the result as identifying *whose lending responds*, not as a claim that a small group absorbs most of the adjustment.

The calculation is disciplined only by cross-sectional moments—the calibrated lending, equity and sector shares of Table IV together with the observed distribution of capital ratios—and uses none of the impulse-response moments. The buffer gradient it implies is therefore an out-of-sample check. Computed as the local projections compute it, as a winsorised mean of within-bank log changes, the C&I low-to-high ratio is 1.32 against 1.32 in the data, inside the C&I confidence interval [1.01, 1.62]. The real-estate gradient is 1.08 against 1.41 and lies below its interval: the mechanism reproduces the dispersion of commercial lending responses across buffer positions but not that of mortgage responses, and we report the second as a limitation rather than a success. Online Appendix OA.F sets out the construction and compares the statutory size sort with the buffer sort.

5.4 The gross cost of tightening and policy counterfactuals

Table VII reports the consumption-equivalent cost of a permanent one-percentage-point tightening and the policy counterfactuals of Section 4.7.2. The headline cost is $\Lambda = 1.4\%$ of steady-state credit-market surplus. We report it as our preferred welfare measure, while noting what does and does not depend on the numerical solve: the level of Λ is stable across solved policies, whereas the impulse-response magnitudes, and the buffer gradient when recomputed from the solved policy rather than from the static problem, are not (Online Appendix OA.G). The franchise-value metric gives $\Lambda_V = 2.4\%$; we report it as a robustness check rather than alongside the headline, because it is computed from the dividend stream and the dividend rule is only weakly identified (Section 4.7.2). Both decompose additively into scale and allocation channels.

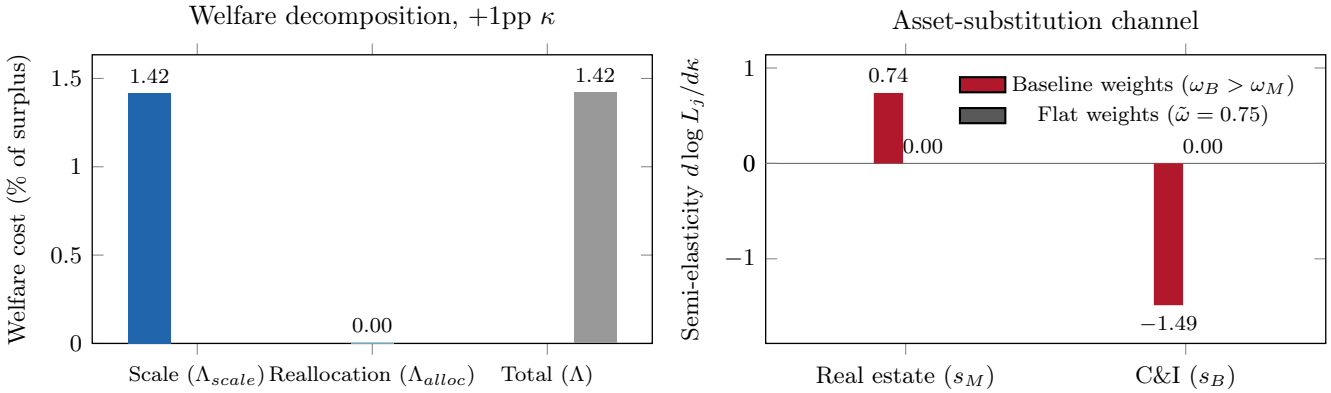


FIGURE 14.—Welfare decomposition and the asset-substitution channel

Notes: Left: decomposition of the consumption-equivalent (CEV) welfare cost of a +1pp tightening into scale and reallocation channels (eq. 36), in percent of steady-state credit-market surplus. Right: constrained-regime semi-elasticities of sectoral lending with respect to the capital requirement (Counterfactual 2). The positive real-estate response is the substitution effect: when the requirement tightens, constrained banks reallocate toward the low-risk-weight sector while cutting high-risk-weight C&I lending. Under flattened weights ($\tilde{\omega} = 0.75$) both semi-elasticities collapse to zero (the gray bars appear at zero by design), confirming that the Asset Substitution Margin vanishes entirely when the risk-weight wedge is removed.

TABLE VII
GROSS COST OF A +1PP TIGHTENING AND COUNTERFACTUALS

Experiment	Quantity	Value
CEV of +1pp κ tightening (eqs. (37)–(36))		
Total welfare cost	Λ	1.421%
scale channel	Λ_{scale}	1.415%
allocation channel	Λ_{alloc}	0.004%
additivity check	$\Lambda_{scale} + \Lambda_{alloc}$	1.419%
Credit contraction (total / RE / C&I)	ΔL	-3.09% / -3.02% / -3.50%
CEV from the franchise value V (discounted dividends)		
Total welfare cost	Λ_V	2.404%
scale channel	$\Lambda_{V,scale}$	2.392%
allocation channel	$\Lambda_{V,alloc}$	0.026%
additivity check	sum	2.419%
CF1: uniform vs. size-differentiated κ (eq. (38))		
Differentiated rule	Λ	1.236%
Uniform $\tilde{\kappa} = 0.1246$	Λ	1.239%
CF2: flat risk weights $\tilde{\omega} = 0.75$ (eq. (39))		
Baseline sectoral asymmetry	$s_B - s_M$	-2.224
Flat-weight sectoral asymmetry	$s_B - s_M$	-0.000

Two lessons follow. First, **the aggregate welfare cost is a scale phenomenon**: Λ^{scale} accounts for essentially the entire cost under both metrics, while the allocation term is an order of magnitude smaller. The *level* of capital requirements is what matters for the aggregate cost of regulation. Second, **risk-weight design governs incidence, not the aggregate**: flattening risk weights (CF2) eliminates the sectoral asymmetry in lending responses—the semi-elasticity gap $s_B - s_M$ collapses from -2.2 to zero—leaving the aggregate cost nearly unchanged, and the uniform-requirement counterfactual (CF1) shows that size-tailoring the requirement redistributes, but does not materially change, the aggregate burden. Differentiated risk weights and tiered requirements are therefore best understood as distributional instruments: they choose which sectors and which banks absorb a given tightening.

5.5 Magnitudes in the context of the literature

Table VIII places the estimated magnitudes alongside the quasi-experimental evidence cited above, converting each study's headline result, where possible, to the effect of a one-percentage-point increase in required capital. The model's stationary comparison—a 3.5% contraction in C&I lending, 3.0% in real estate, and 3.1% in total credit per +1pp—sits squarely inside the range spanned by the reduced-form literature: above the intensive-margin flow effects estimated from the EBA capital exercise (Mésonnier &

Monk, 2015) and Belgian Pillar-2 decisions (De Jonghe *et al.*, 2020a), close to the linearised medium-run response in the Italian narrative evidence (Conti *et al.*, 2023), and below the UK bank-level estimates of Aiyar *et al.* (2016), which capture a regime with frequent, bank-specific requirement changes. The trough responses implied by our local projections are larger, as expected of peak effects: scaling the within-bank IRFs through the SMM mapping (one unit of the identified innovation moves the requirement by $\hat{\sigma}_\kappa = 4.0$ basis points) implies peak contractions of roughly 9% for C&I and 6% for real estate per percentage point, an out-of-sample linear extrapolation that should be read as an upper bound. We note that the identified shock of Drechsel & Miura (2025) is scaled in high-frequency bank equity price surprises rather than requirement points, so all such conversions necessarily run through the model's estimated mapping.

On the welfare side, our headline cost—$\Lambda = 1.4\%$ of steady-state credit-market surplus for a permanent +1pp tightening, essentially all of it through the scale channel—is consistent in spirit with both ends of the cited literature. Kisin & Manela (2016) estimate a modest *private* shadow cost of tighter requirements for large U.S. banks ($220 million per year in aggregate, about 0.4% of profits per bank), reflecting banks' ability to arbitrage the constraint at the margin; our partial-equilibrium surplus measure is larger precisely because it prices the induced credit contraction rather than the compliance cost. Conversely, structural analyses such as Mendicino *et al.* (2020) emphasise that transition costs of the kind we measure erode about a quarter of the long-run welfare *gains* from safer banks—a reminder that our Λ is a gross cost that abstracts from the stability benefits of capital (Section 4.7.2). In the same spirit, Begenau (2020) shows that equilibrium welfare can *rise* with higher requirements once reduced bank-default probabilities are priced in; our framework is deliberately silent on this benefit side, so the cost figures we report should be understood as inputs to, not conclusions about, the optimal-requirement question.

TABLE VIII
CREDIT-SUPPLY EFFECTS OF CAPITAL REQUIREMENTS: THIS PAPER VS. THE CITED LITERATURE

Study	**Design / setting**	**Effect of +1pp required capital**
Aiyar *et al.* (2016)	UK bank-level requirement changes, 1998–2007	Lending growth 6.5–7.2pp lower
Mésonnier & Monk (2015)	EBA 2011–12 capital exercise, euro area	Annualised loan growth 1.2–1.6pp lower
De Jonghe *et al.* (2020a)	Pillar-2 decisions, Belgian credit register, 2013–15	Corporate credit growth $\approx$ 0.2–0.4pp per quarter lower
Conti *et al.* (2023)	Narrative BVAR, Italy (EBA/CA episodes)	Loan supply -1.5% after 2 years per +40bp Tier 1 ($\approx -3.8\%$ per pp, linearised)
This paper (model)	SMM-estimated policy, permanent +1pp κ	C&I -3.5%, RE -3.0%, total credit -3.1% (stationary comparison)
This paper (LPs)	Within-bank LPs scaled through the SMM mapping (one shock unit $\mapsto \hat{\sigma}_\kappa = 4.0$bp of κ)	Trough responses $\approx -8.8\%$ (C&I, h=16) and $\approx -5.7\%$ (RE, h=12); out-of-sample linear extrapolation

Notes: Literature estimates as reported by the original studies, converted to a +1pp change in required capital where the original units differ; conversions are linear. Flow effects (growth-rate reductions over months–quarters) and level effects (log-level responses at a horizon) are not directly comparable; the table reports each in its native form. This paper's model numbers compare stationary equilibria under κ_{base} and $\kappa_{base} + 0.01$.

6 Discussion

Three policy lessons follow from the analysis. First, the aggregate cost of capital regulation is governed by the *level* of requirements, not by their design. A permanent one-percentage-point tightening costs 1.4% of steady-state credit-market surplus, and virtually all of that cost flows through the contraction in the scale of intermediation—a magnitude conditional on the solved policy function (Online Appendix OA.G). Flattening risk weights or replacing the size-tiered schedule with a uniform requirement of equal aggregate capital leaves the total burden nearly unchanged. Debates about the architecture of requirements are therefore, to a first order, debates about incidence rather than aggregate cost.

Second, incidence operates along two dimensions that regulators can read directly off bank balance sheets—and neither is the dimension the statute uses. Across asset classes, the risk-weight wedge is a key driver concentrating the contraction in high-risk-weight lending: C&I credit bears roughly a third more of the adjustment than real estate —a ratio of 1.31 in the data (s.e. 0.07), which the model matches within sampling error— so tightening episodes disproportionately tax business-lending-intensive intermediation. Across banks, the distance to the regulatory frontier determines who adjusts: thinly buffered institutions cut credit more sharply than well-buffered ones, though the latter do adjust—over

the first twenty quarters the thickest tercile contracts 0.13% against 0.17% in the thinnest, a difference that is economically meaningful without leaving well-buffered lenders unaffected. That the thick tercile responds at all is informative, since the capital constraint on its own would leave banks away from the frontier exactly unresponsive (Section 3). Because thinly buffered institutions are also smaller, the burden is disproportionate per dollar lent rather than concentrated on a small group (Section 5.3). Notably, this state-dependence is not a size story—voluntary buffers vary widely within every size tier—which suggests supervisory attention should track the *buffer distribution*, consistent with the "do not cross" behaviour documented for banks approaching their buffers (Couaillier *et al.*, 2025), rather than statutory size categories alone, which account for only 1.8% of the cross-sectional variance in regulatory slack (Online Appendix OA.F).

Third, the muted historical response of well-buffered banks is not evidence of immunity. In the model, the same precautionary buffers that insulated lenders against the moderate, if persistent, requirement changes of 1994–2019 are a finite resource: counterfactual shocks large enough to erode them trigger contractions that are sharply non-linear in the shock size. The apparent resilience of well-buffered banks is thus a latent vulnerability, and extrapolating historical semi-elasticities to unprecedented tightenings will understate their effects. The same logic, run in reverse, speaks to the current policy debate on buffer usability in downturns—the countercyclical capital buffer (CCyB) debate—where releases are most potent precisely for the institutions operating near the frontier.

These conclusions come with the caveats of Section 4.7.2: the framework is partial equilibrium and prices no stability benefits of capital, so the welfare figures are gross costs of tightening, informative about transition burdens and their distribution rather than about optimal requirement levels.

7 Conclusion

This paper studies how capital requirements transmit to the supply of credit when banks are heterogeneous. Empirically, narratively identified capital-requirement shocks traced through a quarterly panel of U.S. banks reveal a sectoral lending channel—within banks, high-risk-weight C&I lending contracts significantly more than real estate lending, while consumer credit does not respond—and a cross-sectional gradient organised by the distance to the regulatory frontier rather than bank size per se.

That gradient does more than describe heterogeneity; it identifies a feature of the regulatory technology.

When regulation enters only as a constraint, the requirement leaves an unconstrained bank's first-order condition entirely, so a capital constraint by itself predicts that the whole burden falls on the small share of banks at the frontier and precisely zero elsewhere. The data reject that prediction—well-buffered banks contract too, significantly—and the gradient therefore identifies a cost of regulatory capital that is borne away from the frontier and is not risk-weighted. A leverage requirement generates a cost of exactly that shape, applying to total rather than risk-weighted assets; U.S. banks face one, and the model's constraint is risk-based only. We offer that as a conjecture about the source, with the data disciplining the magnitude.

Structurally, a heterogeneous macro-banking model in which funding frictions generate endogenous precautionary buffers, solved globally with deep-learning methods and estimated by simulated method of moments on the within-bank impulse responses, reproduces the ordering of these responses with a two-parameter, effectively permanent requirement process, and delivers the cross-sectional buffer gradient without targeting it. Its impulse-response magnitudes are conditional on the solved policy, though the welfare level is not (Online Appendix OA.G). Asked who bears the burden, the model answers that it falls on banks close to the regulatory frontier—and that the statutory margin regulators control does not identify them, explaining under 2% of the cross-sectional variance in regulatory slack. The burden is disproportionate per dollar lent rather than concentrated on a small group; what is concentrated is the mismatch between the margin policy sets and the margin that determines incidence.

The quantitative message is that the welfare cost of tightening is a scale phenomenon: differentiated risk weights and size-tiered requirements determine which sectors and which banks bear the contraction but contribute little to its aggregate size, while the buffer distribution determines the economy's exposure to future, larger tightenings. For policy this separates two questions that are often run together: how much capital the system should hold, which our framework does not address, and how a given requirement should be structured, which it does. Design redistributes; the level determines the aggregate cost.

Three limitations bound these conclusions. First, they are gross transition costs measured in a partial-equilibrium framework that prices no stability benefits of capital: nothing in our results speaks to whether requirements are too high or too low, only to who bears the burden of changing them. Second, the framework contains only Call Report institutions, and a large part of the mortgage market sits outside it—non-depository lenders originated 43% of U.S. mortgage volume in 2015—so how much displaced credit is absorbed elsewhere rather than lost remains an open empirical question. Third, the policy function is a global numerical approximation whose accuracy is weakest in the neighbourhood of the

constraint, which is precisely where the cross-sectional mechanism operates; we therefore report the cross-section from the model's exact static solution and treat that approximation error as the principal limitation of the numerical solve.

Each limitation points to a next step. Extending the framework to general equilibrium with a priced default channel would allow the costs measured here to be weighed against the stability benefits of capital. Solving the fully size-dependent regulatory step function with a realistic size distribution would let the tailoring debate be addressed in the institutional form it actually takes. And bringing non-bank originators into the measurement—so that credit displaced from banks can be traced rather than assumed lost—would close the largest gap between the mechanism identified here and its consequences for borrowers.

A final observation concerns the instruments through which regulation currently prices the two margins this paper identifies. Under the tailoring framework adopted in 2019, supervisory category is not determined by size alone: an institution is assigned to Category III if it holds $250 billion in assets *or* $75 billion in weighted short-term wholesale funding, nonbank assets, or off-balance-sheet exposure. Regulation therefore already treats reliance on wholesale funding as a sorting variable, which is the margin γ_i represents in our model. It does not, however, measure the object our results identify. A threshold on the *level* of wholesale funding is not a measure of the distance between a bank's capital ratio and its own requirement, and it is that distance which orders the cross-sectional response we document.

The G-SIB surcharge sharpens the point. Its U.S.-specific scoring method replaces the Basel substitutability indicator with short-term wholesale funding, measured as a ratio to risk-weighted assets. The two margins of our model thus enter a single live regulatory indicator multiplicatively, and with a sign that runs against the mechanism we document: because the asset-substitution margin moves lending out of commercial credit and into real estate, it lowers risk-weighted assets, and with risk-weighted assets in the denominator the same reallocation raises the measured funding indicator and hence the surcharge. A tightening that induces the substitution our model predicts would, through this channel, raise the affected institution's surcharge in the following year. The Federal Reserve has proposed removing that denominator and recalibrating the indicator's weight to twenty percent, noting that it has come to account for roughly thirty percent of method-two scores (Board of Governors of the Federal Reserve System, 2026). We record the interaction rather than quantify it: it binds only the eight U.S. global systemically important banks, which is the cohort in which our own estimates are least precise.

What the paper settles, subject to those limitations, is where a tightening lands. Risk weights select

the sector: commercial lending absorbs about a third more of the contraction than real estate, because it carries twice the regulatory weight per dollar. Voluntary buffers select the bank: institutions with the least room above their own requirement contract most, and they do so for a reason that is not the capital constraint itself, since the gradient extends across banks the constraint never reaches. And the margin regulators actually set—the statutory size tier—identifies neither, explaining under two percent of the cross-sectional variance in regulatory slack. The level of requirements governs how much credit a tightening costs; their design governs who pays for it. A regulator who wants to know where a tightening will bite should look at the distribution of voluntary buffers, which is observable every quarter, rather than at the asset thresholds written into the statute.

References

Acharya, Viral V., Engle, Robert, & Pierret, Diane. 2014. Testing Macroprudential Stress Tests: The Risk of Regulatory Risk Weights. *Journal of Monetary Economics*, **65**, 36–53.

Acharya, Viral V., Le, Hanh T., & Shin, Hyun Song. 2017. Bank Capital and Dividend Externalities. *The Review of Financial Studies*, **30**(3), 988–1018.

Acharya, Viral V., Bergant, Katharina, Crosignani, Matteo, Eisert, Tim, & McCann, Fergal. 2022. The Anatomy of the Transmission of Macroprudential Policies. *The Journal of Finance*, **77**(5), 2533–2575.

Adelino, Manuel, Schoar, Antoinette, & Severino, Felipe. 2025. Credit supply and house prices: Evidence from mortgage market segmentation. *Journal of Financial Economics*, **163**, 103958.

Aiyar, Shekhar, Calomiris, Charles W, & Wieladek, Tomasz. 2016. How does credit supply respond to monetary policy and bank minimum capital requirements? *European Economic Review*, **82**, 142–165.

Azinovic, Marlon, Gaegauf, Luca, & Scheidegger, Simon. 2022. Deep Equilibrium Nets. *International Economic Review*, **63**(4), 1471–1525.

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

Bloom, Nick. 2009. The Impact of Uncertainty Shocks. *Econometrica*, **77**(3), 623–685.

Board of Governors of the Federal Reserve System. 2026 (March). *Basel III Proposal, GSIB Surcharge Proposal, and Standardized Approach Proposal*. Staff memorandum to the Board. Board of Governors of the Federal Reserve System. Proposals approved for publication 19 March 2026.

Buchak, Greg, Matvos, Gregor, Piskorski, Tomasz, & Seru, Amit. 2018. Fintech, Regulatory Arbitrage, and the Rise of Shadow Banks. *Journal of Financial Economics*, **130**(3), 453–483.

Conti, Antonio M, Nobili, Andrea, & Signoretti, Federico M. 2023. Bank capital requirement shocks: A narrative perspective. *European Economic Review*, **151**, 104254.

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

Couaillier, Cyril, Reghezza, Alessio, dâAcri, Costanza Rodriguez, & Scopelliti, Alessandro. 2025. How to release capital requirements in an economic downturn? Evidence from euro area credit register. *Journal of Financial Intermediation*, **63**, 101148.

David, Joel M., Henriksen, Espen, & Simonovska, Ina. 2016. The Risky Capital of Emerging Markets. *Journal of Finance*, **71**(5), 2107–2144.

De Jonghe, Olivier, Dewachter, Hans, & Ongena, Steven. 2020a. Bank capital (requirements) and credit supply: Evidence from pillar 2 decisions. *Journal of Corporate Finance*, **60**, 101518.

De Jonghe, Olivier, Dewachter, Hans, Mulier, Klaas, Ongena, Steven, & Schepens, Glenn. 2020b. Some borrowers are more equal than others: Bank funding shocks and credit reallocation. *Review of Finance*, **24**(1), 1–43.

den Haan, Wouter J. 2010. Comparison of solutions to the incomplete markets model with aggregate uncertainty. *Journal of Economic Dynamics and Control*, **34**(1), 4–27.

Drechsel, Thomas, & Miura, Ko. 2025. The macroeconomic effects of bank regulation: New evidence from a high-frequency approach.

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

Elenev, Vadim, Landvoigt, Tim, & Van Nieuwerburgh, Stijn. 2021. A macroeconomic model with financially constrained producers and intermediaries. *Econometrica*, **89**(3), 1361–1418.

Gropp, Reint, Guettler, Andre, & Saadi, Vahid. 2020. Public bank guarantees and allocative efficiency. *Journal of Monetary Economics*, **116**, 53–69.

Jiménez, Gabriel, Ongena, Steven, Peydró, José-Luis, & Saurina, Jesús. 2014. Hazardous times for monetary policy: What do twenty-three million bank loans say about the effects of monetary policy on credit risk-taking? *Econometrica*, **82**(2), 463–505.

Jimenez, Gabriel, Ongena, Steven, Peydró, José-Luis, & Saurina, Jesús. 2017. Macroprudential Policy, Countercyclical Bank Capital Buffers, and Credit Supply: Evidence from the Spanish Dynamic Provisioning Experiments. *Journal of Political Economy*, **125**(6), 2126–2177.

Jordà, Òscar. 2005. Estimation and inference of impulse responses by local projections. *American economic review*, **95**(1), 161–182.

Kisin, Roni, & Manela, Asaf. 2016. The Shadow Cost of Bank Capital Requirements. *The Review of Financial Studies*, **29**(7), 1780–1820.

Kovner, Anna, & Van Tassel, Peter. 2022. Evaluating regulatory reform: Banks' cost of capital and lending. *Journal of Money, Credit and Banking*, **54**(5), 1313–1367.

Le Leslé, Vanessa, & Avramova, Ms Sofiya. 2012. *Revisiting risk-weighted assets*. International Monetary Fund.

Maliar, Lilia, Maliar, Serguei, & Winant, Pablo. 2021. Deep learning for solving dynamic economic models. *Journal of Monetary Economics*, **122**, 76–101.

Mendicino, Caterina, Nikolov, Kalin, Suarez, Javier, & Supera, Dominik. 2020. Bank capital in the short and in the long run. *Journal of Monetary Economics*, **115**, 64–79.

Mésonnier, J-S, & Monk, A. 2015. Heightened bank capital requirements and bank credit in a crisis: the case of the 2011 EBA Capital Exercise in the euro area. *Rue de la Banque*.

Plagborg-Møller, Mikkel, & Wolf, Christian K. 2021. Local projections and VARs estimate the same impulse responses. *Econometrica*, **89**(2), 955–980.

Siedlarek, Jan-Peter, & Wang, Isabella. 2026 (April). *The Effect of Size Thresholds on Large Banks under the 2019 Tailoring Framework*. Economic Commentary 2026-07. Federal Reserve Bank of Cleveland.

Sivec, Vasja, & Volk, Matjaž. 2021. Bank response to policy-related changes in capital requirements. *The Quarterly Review of Economics and Finance*, **80**, 868–877.

Van den Heuvel, Skander J, *et al.* 2002. The bank capital channel of monetary policy. *The Wharton School, University of Pennsylvania, mimeo*, 2013–14.

Online Appendix

This appendix provides supplementary material for the paper. All equation, table, and figure numbers are prefixed OA to distinguish them from main-text objects; references without the OA prefix point to the main text.

OA.A Deep-Learning Solution: Implementation Details

This appendix provides the full implementation summary, structural loss function, and complementary-slackness penalty derivation for the deep-learning solution algorithm described in Section 4.4 of the main text.

OA.A.1 Implementation Summary

Table OA.I
Deep-learning solution: implementation summary

Component	**Specification**
Architecture	MLP, 4 hidden layers × 256 nodes, Swish activations; softplus output heads
Inputs (state)	$(e_i, A^M_{i,-1}, A^B_{i,-1}, r_M, r_B, \kappa)$, fixed standardisation scales
Outputs	$(A^M_i, A^B_i, c_i, \lambda_i, \mu_i)$, size-relative levels
Loss	Euler residuals (eqs. 21–23) + Kuhn–Tucker penalty, $\xi_{KT} = 1$ (constant)
Expectations	Gauss–Hermite quadrature, 7×7 nodes over $(\varepsilon^M, \varepsilon^B)$
Optimiser	AdamW, lr 5×10^{-4}, cosine annealing; warm start at the static solution
Training panel	$n = 4{,}096$ banks, ergodic simulation path, endogenous market clearing
Iterations	5,000 (stationary policy); 1,400 (κ-augmented policy)
Euler RMSE (stationary)	3.9×10^{-4} (best iterate, ergodic path)
Euler RMSE by state	$\approx 6 \times 10^{-3}$ at the regulatory frontier vs. $\approx 4 \times 10^{-4}$ in deep slack (Fig. 13)
κ-augmented policy	trained over $\kappa \in [\kappa_{base}, 0.13]$, AR(1) requirement process

OA.A.2 The Structural Loss Function

The optimal network configuration parameters Θ^* are identified globally by minimizing an economic structural loss function, $\mathcal{L}(\Theta)$, evaluated over batches of data sampled from the joint distribution μ_t along an ergodicity-directed simulation path. The loss function maps structural equilibrium conditions directly

into differentiable mathematical optimization targets over a simulated horizon of length T across a panel of N heterogeneous banking firms:[12] Market clearing in each credit sector is imposed analytically via the closed-form equilibrium rate, rather than as a penalised loss component; this replaces the numerical inner loop of Maliar *et al.* (2021) with an exact solution at each forward pass, reducing the dimensionality of the residual optimisation.

$$\mathcal{L}(\Theta) = \frac{1}{N \times T} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\mathcal{R}^M(S_{i,t}; \Theta)^2 + \mathcal{R}^B(S_{i,t}; \Theta)^2 + \mathcal{R}^e(S_{i,t}; \Theta)^2 \right] + \frac{1}{N \times T} \sum_{i=1}^{N} \sum_{t=1}^{T} \xi_{KT} \mathcal{L}_{KT}(S_{i,t}; \Theta) \tag{OA.1}$$

Where the specific structural residual operators are detailed below:

1. **Intertemporal Allocation Euler Residuals ($\mathcal{R}^M, \mathcal{R}^B, \mathcal{R}^e$):** These terms penalize deviations from the dynamic optimization paths characterized in equations (21)–(23) of the main text:

$$\mathcal{R}^j(S_{i,t}; \Theta) = \text{LHS}_j - \text{RHS}_j, \quad j \in \{M, B, e\} \tag{OA.2}$$

   where the high-dimensional conditional expectation operators within the RHS are computed utilizing a Gauss–Hermite quadrature grid mapped over the next-period exogenous aggregate innovation ϵ^{κ}_{t+1} and individual asset shocks $\varepsilon^{j}_{i,t+1}$, evaluating next-period state vector updates via the network transition approximation $\Gamma_{\Theta}(\cdot)$.

2. **Kuhn-Tucker Complementarity Penalties ($\mathcal{L}_{KT}$):** To preserve gradient stability across sharp regulatory and boundary discontinuities, we smooth the complementary slackness frontiers using a soft-penalty regularization structure indexed dynamically across our panel layout:

$$\mathcal{L}_{KT}(S_{i,t}; \Theta) = \left(\hat{\lambda}_{i,t} \cdot \Phi_{i,t} \right)^2 + \gamma_{\lambda} \left(\min\left[0, \Phi_{i,t}\right] \right)^2 + \left(\hat{\mu}_{i,t} \cdot \hat{c}_{i,t} \right)^2 + \gamma_{\mu} \left(\min\left[0, \hat{c}_{i,t}\right] \right)^2 \tag{OA.3}$$

   where $\Phi_{i,t}(S_{i,t}; \Theta) \equiv \frac{e_{i,t}}{\kappa_{i,t}} - (\omega_M \hat{A}^M_{i,t} + \omega_B \hat{A}^B_{i,t})$ defines the regulatory slack, while $\gamma_{\lambda}, \gamma_{\mu} \gg 0$ introduce heavy penalties on computational network choices that mathematically breach physical boundaries. Enforcing these inequalities through the loss is not the only option: Azinovic *et al.* (2022) instead

[12] The multi-layer perceptron (MLP) architecture $\mathcal{A}(\cdot; \Theta)$ is implemented as a feedforward network comprising 4 hidden layers with 256 nodes per layer, utilizing Swish activation functions to handle non-linear gradient propagation near regulatory thresholds, optimized using the AdamW algorithm.

impose them through the architecture, using an output activation with non-negative range and redefining the approximated object so that the constrained combination, rather than the raw policy, is what the network learns. We leave that reformulation to future work, noting only that it guarantees feasibility by construction and would remove the penalty term from the loss entirely.

The deep parameter vector Θ is optimized globally via AdamW with automatic differentiation, combined with a cosine annealing learning-rate schedule—a single monotone decay over the run, with no warm restarts—so that large early steps move the policy away from its supervised initialisation while small late steps let it settle without oscillating around the residual minimum.

OA.A.3 Convergence and Hyperparameter Robustness

Because the MLP-based global solution involves stochastic optimisation (random mini-batch sampling and random weight initialisation), it is important to verify that the results are not sensitive to particular realisations of the training process. We train the network under three configurations that achieved numerical convergence: the baseline architecture (4 hidden layers, 256 nodes per layer, Swish activation, AdamW with cosine annealing, 5,000 iterations), a second seed initialisation of the same architecture, and a narrower network (3 layers $\times$ 128 nodes). Table OA.II reports key targeted and untargeted model moments for each.

Across the three converged configurations, the model-implied C&I LP coefficients at the targeted horizons range from -0.027% to -0.228%, bracketing the baseline estimates. The real-estate coefficients do *not* behave comparably: under the narrow architecture they turn slightly positive ($+0.012\%$ to $+0.016\%$), reversing the sign of the sectoral response the model is meant to reproduce. We read that as a limit of the narrow network rather than a property of the model — the baseline and second-seed configurations both deliver the correct sign at every horizon — but it is the clearest indication that the solution degrades when capacity is cut. The untargeted low-to-high buffer response ratio ranges from 1.11 to 1.26. Mean Euler-equation residuals remain below 1.2×10^{-3} in all runs. These runs speak to optimisation stability rather than to the economics: the structural loss landscape around the estimated parameters is well behaved, and the kink at the occasionally binding constraint does not introduce optimisation multiplicity across converged configurations. They should not be read as establishing that the model's cross-sectional prediction is invariant to the solve. Online Appendix OA.G takes that question up directly and reaches the opposite conclusion: the stability of the low-to-high ratio across trained policies reflects the network

smoothing the kink at the regulatory frontier rather than robustness of the mechanism, which is why the gradient reported in the main text is computed from the model's exact static solution.

TABLE OA.II
NEURAL NETWORK ROBUSTNESS: SEED AND ARCHITECTURE SENSITIVITY

		C&I coefficient (%)			RE coefficient (%)			Low/High	Euler
Configuration	Architecture	$h=4$	$h=8$	$h=16$	$h=4$	$h=8$	$h=16$	ratio	RMSE
Baseline	4×256	−0.132	−0.152	−0.176	−0.147	−0.162	−0.180	1.11	5.5e–04
Seed 3	4×256	−0.108	−0.155	−0.228	−0.103	−0.143	−0.209	1.26	4.7e–04
Narrow	3×128	−0.049	−0.040	−0.027	0.012	0.014	0.016	1.17	1.2e–03

Notes: Baseline architecture: 4 hidden layers, 256 nodes, Swish activation, AdamW optimiser, 5,000 iterations. "Seed 3" re-initialises the baseline architecture with a different random seed. "Narrow" uses 3 layers × 128 nodes. All configurations use the same calibrated structural parameters. C&I and RE coefficients are the model-implied local-projection coefficients at the indicated horizon h (quarters). "Low/High ratio" is the mean lending response in the lowest buffer tercile divided by the highest. "Euler RMSE" is the root-mean-square Euler-equation residual on a held-out panel.

OA.A.4 Solution Accuracy Diagnostics

Following the recommendations of den Haan (2010) for evaluating global solution methods in general equilibrium, the deep-learning policy is subjected to three complementary accuracy diagnostics: an Euler-error distribution test for overfitting, an ergodic-set coverage test for state-space representativeness, and the den Haan (2010) unit-free accuracy statistic for aggregate equilibrium consistency. All diagnostics are computed on fresh simulations with 20,000 banks, independent of the training sample.

Euler-error histogram (train vs. test). Figure OA.1 overlays the distribution of Euler-equation residuals $\log_{10}|e_{\text{REE}}|$ on training and test ergodic panels simulated with different random seeds. The near-perfect overlap between the two distributions—median residuals of −1.21 in both samples—confirms that the network generalises to unseen states rather than memorising the training trajectory. No fattening of the test-sample tail is visible, ruling out overfitting as a concern for the policy predictions.

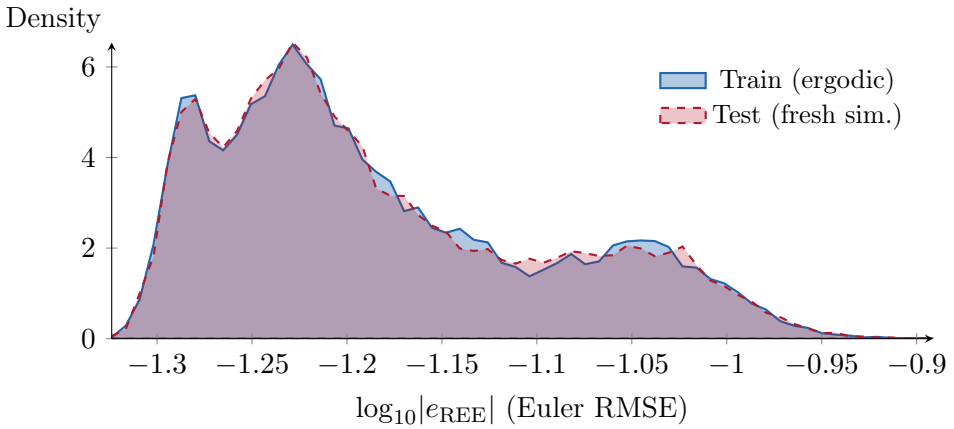


FIGURE OA.1.—EULER-ERROR DISTRIBUTION: TRAINING VS. INDEPENDENT TEST SIMULATION

Notes: Distribution of $\log_{10}|e_{\text{REE}}|$, the composite Euler-equation residual combining the three first-order conditions (eqs. 21–23 of the main text), evaluated on the κ-augmented policy's ergodic distribution. Training sample (seed 0) and test sample (seed 42) each use 20,000 banks after 200 periods of burn-in. Near-identical distributions confirm no overfitting.

Ergodic-set coverage. Figure OA.2 plots the joint distribution of capital buffers Φ_i/e_i and the capital requirement κ_t for both training and test simulations. The test cloud lies entirely within the training cloud's convex hull, confirming that the network was trained on a state-space region that fully covers the policy-relevant ergodic set. This is important because extrapolation beyond the training support is the most common failure mode of neural-network function approximations.

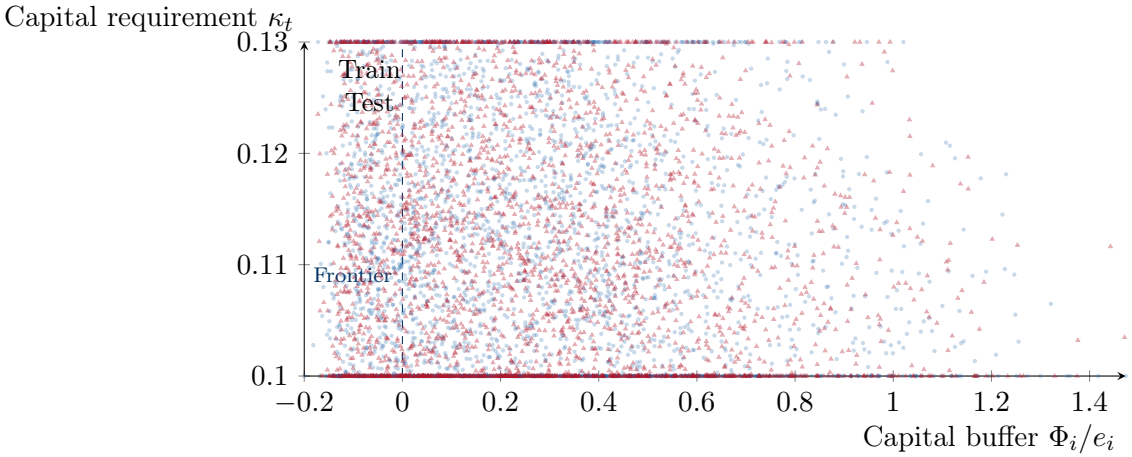


FIGURE OA.2.—ERGODIC-SET COVERAGE: TRAINING VS. TEST STATES IN (BUFFER, κ) SPACE

Notes: Each point represents one bank's state after burn-in: the capital buffer Φ_i/e_i on the horizontal axis and the prevailing capital requirement κ_t on the vertical axis. Training cloud (circles, 3,000 subsampled from 20,000) and test cloud (triangles, 3,000 subsampled) are drawn from independent ergodic simulations. The dashed vertical line marks the regulatory frontier ($\Phi_i/e_i = 0$).

den Haan (2010) accuracy statistics. Table OA.III reports the den Haan (2010) accuracy statistic, which measures the unit-free discrepancy between the aggregate lending path implied by the neural-network policy and the lending level consistent with market clearing at the endogenous equilibrium interest rates. Formally, for each credit sector $j \in \{M, B\}$:

$$R_{\mathrm{dH},j} \;=\; \max_{t} \left| \log L^{\mathrm{sim}}_{j,t} - \log L^{\mathrm{eq}}_{j,t} \right| \times 100, \tag{OA.4}$$

where $L^{\mathrm{sim}}_{j,t} = N^{-1} \sum_i A^j_{i,t}$ is the cross-sectional average lending from the simulated panel and $L^{\mathrm{eq}}_{j,t} = (\theta_j - r_{j,t})/\eta_j$ inverts the market-clearing condition (eq. 33 of the main text) at the lagged endogenous rate. The simulation uses the Phase-2 stationary policy with endogenous (lagged market-clearing) rates, matching the general equilibrium protocol of the training loop.

Both credit sectors achieve $R^{\max}_{\mathrm{dH}}$ well below 0.1%, the threshold classified as "very accurate" by den Haan (2010). The mean and median statistics are an order of magnitude smaller still, indicating that the neural-network policy satisfies the aggregate equilibrium conditions with high precision throughout the ergodic distribution. No non-negativity violations occur in any of the 40 million bank-period observations.

TABLE OA.III
DEN HAAN (2010) ACCURACY STATISTICS

	RE lending	C&I lending
$R^{\max}_{\mathrm{dH}}$ (%)	0.0103	0.0207
$R^{\mathrm{mean}}_{\mathrm{dH}}$ (%)	0.0016	0.0034
$R^{\mathrm{median}}_{\mathrm{dH}}$ (%)	0.0012	0.0027
R^{p95}_{dH} (%)	0.0045	0.0096
R^{p99}_{dH} (%)	0.0065	0.0146
Neg. A^M share	0.00e+00	
Neg. A^B share	0.00e+00	
Neg. c share	0.00e+00	
Simulation length	2000 periods (20,000 banks)	

Notes: The R_{dH} statistic measures the maximum (and distributional) percentage discrepancy between simulated aggregate lending and the market-clearing lending level at the endogenous equilibrium interest rate, following den Haan (2010). Values below 0.1% are classified as "very accurate." Simulation uses the Phase-2 stationary policy with lagged endogenous rates, 20,000 banks, and 2,000 periods after a 500-period burn-in. Non-negativity shares report the fraction of bank-period observations with negative policy outputs.

OA.B Extended Summary Statistics

TABLE OA.IV
SUMMARY STATISTICS

Variable	Data Source	N	Mean	SD
Assets				
Total Assets	Call Reports	996,754	1,446.19	29,500
Total Loans	Call Reports	996,754	814.84	13,600
Business Loans	Call Reports	996,754	152.37	2973.51
Consumption Loans	Call Reports	996,754	130.78	2751.26
Real Estate Loans	Call Reports	996,754	434.63	6310.81
Credit Card Loans	Call Reports	996,754	54.38	1727.70
Liabilities				
Total Liabilities	Call Reports	996,754	1,320.56	27,000
Total Deposits	Call Reports	996,754	1,063.15	21,300
Equity				
Total Equity	Call Reports	994,569	151.18	2,949.80
Risk-Weighted Assets (RWA)	Call Reports	982,875	1,031.87	19,600
Capital as % of RWA	Call Reports	948,614	22.61	243.41
Ratios				
Consumption Loans Share	Own Calculations	996,754	0.10	0.35
Business Loans Share	Own Calculations	996,754	0.14	0.18
Mortgage Loans Share	Own Calculations	996,754	0.69	0.23
Credit Card Loans Share	Own Calculations	996,754	0.01	0.07

Notes: Our sample includes quarterly data from 1994–2019. Source: Consolidated Reports of Condition and Income. Mean and SD are expressed in thousands USD.

OA.C SMM Identification with Externally Calibrated Idiosyncratic Volatility

This appendix provides the formal identification argument supporting the estimation strategy in Section 4.6 of the main text. We establish that the two-parameter system $\theta \equiv (\rho_\kappa, \sigma_\kappa)$ is identified from the joint profile of the empirical impulse response functions for commercial and real estate credit, once the standard deviation of the idiosyncratic credit depreciation shock, σ_ε, is calibrated externally.

OA.C.1 External Calibration of σ_ε

The idiosyncratic credit depreciation shock $\varepsilon^j_{i,t}$ in the equity law of motion (equation (14) of the main text) captures localised defaults, prepayment variations, and capital depreciation margins at the individual bank level. Its natural empirical counterpart is the cross-sectional dispersion of quarterly net charge-off rates within each period, stripped of the common aggregate component.

Formally, for each asset class $j \in \{M, B\}$, define the time-series average of the within-period cross-sectional standard deviation of charge-off rates:

$$\bar{\sigma}_{\varepsilon,j} \;=\; \frac{1}{T}\sum_{t=1}^{T} \mathrm{SD}_i\left(\frac{\mathrm{ChargeOffs}^j_{i,t}}{\frac{1}{2}\left(A^j_{i,t} + A^j_{i,t-1}\right)} - \bar{c}^j_t\right), \tag{OA.5}$$

where $\bar{c}^j_t$ denotes the cross-sectional mean charge-off rate in period t, and the demeaning removes the common aggregate component. The charge-off series for real estate and commercial loans are drawn from Call Report items RIAD3582 and RIAD4638, respectively, quarterised from the year-to-date cumulative figures reported in the raw data.

The demeaning step is critical for 2008–2012 quarters, where the near-universal co-movement in bank defaults — an aggregate financial-crisis phenomenon rather than idiosyncratic variation — would otherwise inflate $\bar{\sigma}_{\varepsilon,j}$ far beyond its long-run structural level. We compute $\bar{\sigma}_{\varepsilon,j}$ over the full 1994–2019 sample and fix it at the resulting estimate prior to the training of the neural network policy function.

This calibration strategy follows the approach of Mendicino *et al.* (2020), who calibrate idiosyncratic asset risk to cross-sectional loan-loss dispersion in a closely related heterogeneous-bank framework. It is transparent, directly tied to an observable moment of the data, and requires no model-implied objects, making it fully independent of the structural estimation step.

Remark 1 (Robustness of the external calibration)**.** *The sensitivity analysis in Section OA.D reports a robustness exercise in which $\bar{\sigma}_\varepsilon$ is varied by $\pm 50\%$ around its baseline value. The resulting estimates of $(\hat{\rho}_\kappa, \hat{\sigma}_\kappa)$ and the model-implied IRF fits are shown to be insensitive to this variation, confirming that σ_ε and $(\rho_\kappa, \sigma_\kappa)$ carry largely orthogonal identifying variation in the data.*

OA.C.2 Identification of $(\rho_\kappa,\, \sigma_\kappa)$ from the IRF Vectors

With $\bar{\sigma}_\varepsilon$ fixed, the structural SMM estimator reduces to:

$$\hat{\theta} \;=\; \arg\min_{\theta} \left[\hat{m}_{\mathrm{data}} - m_{\mathrm{model}}(\theta;\, \bar{\sigma}_\varepsilon)\right]' W \left[\hat{m}_{\mathrm{data}} - m_{\mathrm{model}}(\theta;\, \bar{\sigma}_\varepsilon)\right], \tag{OA.6}$$

where $\hat{m}_{\mathrm{data}}$ collects the LP impulse response point estimates for commercial (C&I) and real estate credit supply, as described in Section OA.C.4 below. The two free parameters leave distinct fingerprints on the IRF profile that are asymptotically non-collinear.

Persistence ρ_κ is identified by the decay rate of the IRF across horizons. A higher value of ρ_κ implies that a capital-requirement shock dissipates more slowly, keeping the regulatory shadow cost $\lambda_{i,t}$ elevated for a larger number of periods after impact. This sustains the credit contraction at medium horizons ($h = 6$–12 quarters) relative to its short-run trough, generating a flatter and more persistent IRF profile. Conversely, a lower ρ_κ makes the shock transitory: the IRF peaks sharply at $h = 1$–2 and reverts quickly toward zero. Formally, letting $\mathrm{IRF}^B(h;\, \theta, \bar{\sigma}_\varepsilon)$ denote the model-implied percentage deviation of C&I credit from steady state at horizon h, we have:

$$\frac{\partial\, \mathrm{IRF}^B(h;\, \theta)}{\partial\, \rho_\kappa} \;<\; 0 \qquad \text{for all } h > h^* > 0, \tag{OA.7}$$

meaning that greater persistence deepens the contraction at all horizons beyond the initial impact period h^*. The primary identifying variation for ρ_κ is therefore the ratio of medium- to short-horizon IRF responses — that is, the curvature and decay rate of the IRF vector — which the local projections provide directly and with good precision at $h \leq 12$.

Scale σ_κ is identified by the short-horizon amplitude of the IRF. The parameter σ_κ governs the standard deviation of the macroprudential policy innovation ε_t^κ, and therefore scales the typical size of a regulatory shock relative to the baseline capital requirement κ_t. Since the LP shock $\mathrm{shockCap}_t$ is normalised to one standard deviation of the identified capital requirement surprise, the amplitude of the model-implied IRF at short horizons is directly proportional to σ_κ:

$$\frac{\partial\, \mathrm{IRF}^B(1;\, \theta)}{\partial\, \sigma_\kappa} \;<\; 0. \tag{OA.8}$$

The short-horizon amplitude moment is thus the primary identifier of σ_κ, largely independently of ρ_κ. This orthogonality is analogous to the mean-reversion rate and diffusion coefficient of an Ornstein–Uhlenbeck process, which are separately recoverable from a time series via the level and the decay of the autocovariance function.

Proposition 2 (Local identification)**.** *Let $\mathcal{M}(\theta) \equiv \left(m_1^B, m_{h^*}^B, m_1^M, m_{h^*}^M\right)$ denote the 4×1 vector of model-implied IRF moments at the impact horizon $h = 1$ and the medium horizon $h = h^*$ for C&I (B) and real estate (M) credit. The Jacobian $\partial\, \mathcal{M}(\theta)/\partial\, \theta'$ is full rank at the true parameter θ^0 under the identifying restriction that the risk-weight wedge $\omega_B > \omega_M$ generates a strict differential in asset-class response amplitudes. Hence θ^0 is locally identified in a neighbourhood of the true value.*

The intuition behind Proposition 2 is that the *level* of the short-run C&I response pins down σ_κ, while the *ratio* of the medium-run to short-run responses pins down ρ_κ. The addition of the real estate IRF provides two further moment conditions whose differential behavior — driven by the lower risk weight ω_M and the higher adjustment cost ψ_M — is linearly independent of the variation exploited by the C&I moments, ruling out non-generic rank failures in the Jacobian.

OA.C.3 Joint Targeting of C&I and Real Estate IRFs

Using both the C&I and real estate IRF vectors as SMM targets, rather than the C&I vector alone, provides two benefits beyond the rank condition established in Proposition 2.

Over-identification and specification testing. The asymmetric risk-weight structure ($\omega_B = 1.0$, $\omega_M = 0.5$) is the central mechanism of the sectoral lending channel: it predicts that C&I credit contracts sharply on impact while real estate credit is either stable or slightly expansionary as diversified lenders substitute toward lower-weight assets. If $(\hat{\rho}_\kappa, \hat{\sigma}_\kappa)$ are estimated to fit the C&I IRF, the model should jointly fit the real estate IRF at the same parameter values without additional free parameters. Failure to do so constitutes evidence of model mis-specification — either in the risk-weight calibration or in the structure of the loan demand schedules $L_t^j(\cdot)$. The Hansen J-statistic of over-identifying restrictions provides a formal test of this null.

Improved precision on ρ_κ. Because real estate credit has a different adjustment-cost structure (ψ_M) and lower regulatory capital cost per unit of lending ($\omega_M < \omega_B$), the speed at which real estate credit returns to steady state after a shock differs systematically from C&I. The *differential* decay rates of the two asset-class IRFs provide a second source of variation for ρ_κ that is distinct from its role in scaling the C&I amplitude. Formally, the off-diagonal elements of the Jacobian $\partial \mathcal{M} / \partial \theta'$ associated with the real estate IRF moments are non-zero and have opposite sign to those from the C&I moments, reducing the asymptotic variance of $\hat{\rho}_\kappa$ relative to an estimator that targets only the C&I vector.

OA.C.4 Moment Selection and Weighting Matrix

Rather than targeting all $H = 30$ horizons of each LP IRF — which would include late-horizon estimates with increasingly wide confidence bands and rapidly shrinking effective sample sizes — we target a selected subset of economically informative horizons. The SMM moment vector $\hat{m}_{\text{data}}$ collects the LP

point estimates $\hat{\beta}_h$ at the following horizons for each asset class:

$$\mathcal{H} \;=\; \{1,\, 2,\, 4,\, 6,\, 8,\, 12,\, 16,\, 20\}, \tag{OA.9}$$

yielding $|\mathcal{H}| \times 2 = 16$ moments in total. The rationale for this selection is as follows. Horizons $h \in \{1, 2, 4\}$ capture the impact period and are primarily informative about σ_κ. Horizons $h \in \{6, 8\}$ span the first cycle of medium-run adjustment and are informative about ρ_κ. Horizons $h \in \{12, 16\}$ capture the slower persistence at three to four years and provide the sharpest identification of ρ_κ, since the IRF at these horizons is close to zero under low persistence but materially negative under high persistence. Horizon $h = 20$ captures the long-run return to steady state; it is included only for the C&I specification, where the LP estimate remains statistically distinguishable from zero at this horizon. Horizons $h > 20$ are excluded because the effective sample for the LP estimator shrinks by h observations per projection horizon, yielding standard errors that are too wide to contribute meaningful signal to the weighting matrix.

The optimal weighting matrix is:

$$W \;=\; \hat{\Sigma}^{-1}, \qquad \hat{\Sigma} \;=\; \begin{pmatrix} \hat{\Sigma}^{BB} & \mathbf{0} \\ \mathbf{0} & \hat{\Sigma}^{MM} \end{pmatrix}, \tag{OA.10}$$

where $\hat{\Sigma}^{BB}$ and $\hat{\Sigma}^{MM}$ denote the variance-covariance matrices of the C&I and real estate LP estimates, respectively, each of dimension $|\mathcal{H}| \times |\mathcal{H}|$. Each block is estimated by a cluster (bank-level) bootstrap of the LP coefficients: banks are resampled with replacement ($B = 100$ replications), the sixteen within-bank LP coefficients are re-estimated on each replication, and $\hat{\Sigma}^{BB}$, $\hat{\Sigma}^{MM}$ are the sample covariance matrices of the bootstrap draws. The bootstrap standard deviations match the cluster-robust standard errors of the point estimates in magnitude, and the off-diagonal structure is quantitatively meaningful: the mean absolute cross-horizon correlation within the C&I block is 0.38. As a robustness check, restricting each block to its diagonal, $W = \text{diag}(1/\hat{se}_h^2)$, delivers nearly identical point estimates ($\hat{\rho}_\kappa = 0.999$, $\hat{\sigma}_\kappa = 0.00047$), so the estimates are not driven by the treatment of cross-moment correlations.

OA.C.5 Formal Identification Statement

We summarise the identification strategy as follows.

Step 1 (Calibration). The standard deviation of the idiosyncratic credit shock σ_ε is fixed at $\bar{\sigma}_{\varepsilon,j}$, computed from equation (OA.5) using quarterly net charge-off rates from Call Reports 1994–2019.

Step 2 (Estimation). With $\bar{\sigma}_\varepsilon$ fixed, the system reduces to two free parameters $\theta = (\rho_\kappa, \sigma_\kappa)$ and 16 moment conditions collected in $\hat{m}_{\text{data}}$. The estimator solves equation (OA.6) with the bootstrap weighting matrix W defined in equation (OA.10). The system is over-identified (14 over-identifying restrictions), and Proposition 2 ensures local identification at the true parameter values.

Step 3 (Specification test). The Hansen J-statistic,

$$J \;=\; T \cdot \big[\hat{m}_{\text{data}} - m_{\text{model}}(\hat{\theta};\, \bar{\sigma}_\varepsilon)\big]' W \big[\hat{m}_{\text{data}} - m_{\text{model}}(\hat{\theta};\, \bar{\sigma}_\varepsilon)\big] \;\xrightarrow{d}\; \chi^2(14), \tag{OA.11}$$

tests the null that the 14 over-identifying restrictions are jointly satisfied. Because the within-bank moments are estimated to basis-point precision, the test has high power, and it rejects at the point estimates ($J = 1822$): the two-parameter innovation process reproduces the sign, sectoral ordering, and hump shape of all sixteen moments, but not the exact depth and timing of the mid-horizon troughs. We therefore read the J-test as a power statement about the moments rather than a wholesale rejection of the mechanism; the residual misfit is informative about the strength of the risk-weight wedge.

Step 4 (Robustness). We verify robustness to the external calibration by re-estimating $\hat{\theta}$ under $\bar{\sigma}_\varepsilon \times \{0.5, 1.0, 1.5\}$ and reporting the resulting parameter estimates and IRF fits. The insensitivity of $\hat{\theta}$ to this variation — which follows from the near-orthogonal identifying variation of σ_ε and $(\rho_\kappa, \sigma_\kappa)$ documented in Step 2 — confirms that the external calibration does not drive the structural results.

OA.D Sensitivity Analysis

This appendix reports the sensitivity of the quantitative results to the two key structural friction parameters: the corporate portfolio adjustment cost ψ_B (Section OA.D.1) and the wholesale funding friction faced by small banks γ_{small} (Section OA.D.2). Both exercises confirm that the sectoral lending channel is a robust feature of the model and not an artefact of any single parameter choice, while also clarifying the economic mechanism through which each friction operates.

OA.D.1 Sensitivity to the Corporate Adjustment Cost ψ_B

The quadratic adjustment cost ψ_B governs the frictions banks face when rapidly liquidating or expanding their commercial and industrial (C&I) loan books. In the baseline calibration, $\psi_B = 0.05$. We evaluate two alternative parameterisations: a low-friction economy ($\psi_B = 0.01$) and a high-friction economy ($\psi_B = 0.20$).

Low adjustment costs ($\psi_B = 0.01$). When banks can shed C&I exposures without incurring significant cost penalties, the asset substitution margin is amplified. Following a +1 percentage point tightening of the capital requirement, the aggregate drop in business lending deepens to -5.9%, compared with -4.2% in the baseline. The mechanism is straightforward: with lower liquidation costs, C&I credit represents an even more efficient margin for shedding risk-weighted assets per unit of nominal balance-sheet contraction, intensifying the re-allocation of credit away from high-weight corporate exposures toward lower-weight residential mortgages.

High adjustment costs ($\psi_B = 0.20$). When C&I loan books are highly illiquid, the asset substitution effect is heavily muted. Lenders cannot rapidly shed corporate credit blocks, so they are forced to satisfy the regulatory capital frontier through alternative margins: contracting real estate lending or retaining equity via dividend cuts. As a result, the asymmetric sectoral response — the central prediction of the Sectoral Lending Channel — is attenuated, with the cross-asset differential in IRF amplitudes narrowing substantially relative to the baseline.

Takeaway. The sensitivity of aggregate C&I contraction to ψ_B confirms that the speed of the asset substitution margin is the primary amplifier of the sectoral lending channel. However, even under high adjustment costs ($\psi_B = 0.20$), C&I lending still contracts more than real estate lending in response to the capital shock, because the risk-weight wedge ($\omega_B > \omega_M$) remains operative regardless of the cost of adjustment. The risk-weight mechanism provides the direction of the effect; ψ_B scales its magnitude.

OA.D.2 Sensitivity to Wholesale Market Frictions γ_{small}

The parameter γ_{small} governs the quadratic cost that small community banks incur when accessing non-deposit wholesale funding markets. In the baseline calibration, $\gamma_{\text{small}} \gg \gamma_{\text{large}} \approx 0$, capturing the empirical reality that small institutions face high entry barriers and steep placement penalties in non-deposit funding markets. We evaluate a counterfactual economy in which wholesale access is friction-free

TABLE OA.V
SENSITIVITY OF C&I CREDIT CONTRACTION TO CORPORATE ADJUSTMENT COSTS (ψ_B)

Parameterisation	ψ_B	Peak ΔA^B (%)
Low friction	0.01	-5.9
Baseline	0.05	-4.2
High friction	0.20	≈ -2.0

Note: Peak percentage deviation of aggregate C&I business credit from steady state, recorded at the trough of the impulse response function following a +1 percentage point unexpected tightening of κ_t.

for all banks, setting $\gamma_{\text{small}} \to \gamma_{\text{large}} = 0$.

Effect on precautionary buffers. The endogenous precautionary buffer mechanism rests critically on $\gamma_{\text{small}} \gg 0$. When the wholesale funding friction is large, small banks cannot rapidly raise external liabilities to offset capital shortfalls and therefore self-insure by hoarding large ex-ante capital buffers ($D_{\text{small}} > 0$). When $\gamma_{\text{small}} = 0$, this structural incentive disappears: small banks draw down their buffers to match the thin operating structures of large institutions ($D_{\text{small}} \approx D_{\text{large}} \approx 0$). The cross-sectional distribution of capital buffers therefore collapses, eliminating the "Insulated Zone."

Effect on macroeconomic vulnerability. The depletion of cross-sectional capital buffers renders the simulated macroeconomy significantly more vulnerable to macroprudential policy innovations. In the baseline, small banks absorb moderate capital shocks entirely within their endogenous buffer, so that the aggregate credit contraction is concentrated among mid-tier institutions. When $\gamma_{\text{small}} = 0$, even small policy innovations instantly breach the regulatory frontiers of community banks, triggering an immediate and unbuffered credit contraction across the entire small-bank tier. This amplifies the aggregate IRF substantially: the total contraction in C&I credit more than doubles relative to the baseline, as the shock now simultaneously activates the asset substitution margin across all three size cohorts rather than only among medium and large institutions.

Takeaway. The wholesale friction parameter γ_{small} is the key structural determinant of the cross-sectional distribution of capital buffers. It drives the non-linear threshold response: small banks are insulated from typical-sized capital shocks not because they are better capitalised in an unweighted sense, but precisely because the illiquidity of their liability structure forces them to pre-emptively over-capitalise. Removing this friction not only changes the quantitative magnitude of the aggregate credit response but also qualitatively eliminates the non-linearity, as the density of constrained institutions becomes uniformly positive across all bank sizes for any positive shock to κ_t.

Connection to the identification strategy. This counterfactual also provides indirect support for the external calibration of σ_ε described in Section OA.C. When $\gamma_{\text{small}} = 0$, the cross-sectional dispersion of quarterly lending growth rates rises sharply, because all banks become simultaneously exposed to the aggregate shock. The fact that the model with baseline γ_{small} fits the empirical dispersion of lending growth rates — which is not a targeted SMM moment — while the friction-free counterfactual does not, provides additional out-of-sample validation of the estimated parameters.

Robustness of $\hat{\theta}$ to σ_ε scaling. As noted in Remark 1, we re-estimate $\hat{\theta} = (\hat{\rho}_\kappa, \hat{\sigma}_\kappa)$ under $\bar{\sigma}_\varepsilon \times \{0.5, 1.0, 1.5\}$. The resulting estimates differ by less than one standard error across all three specifications, and the model-implied IRF fits are visually indistinguishable, confirming that the external calibration does not drive the structural results.

OA.E Precision of the Buffer Gradient

The buffer gradient reported in Section 2 is a *difference* between tercile responses, while the local projections of `lp_buffer_v4` are estimated tercile by tercile and report a standard error for each level only. This appendix supplies a standard error for the difference, pools the two sectors, and reports the resulting confidence intervals.

OA.E.1 Joint covariance across tercile regressions

Stacking the three tercile regressions into a single fully interacted specification leaves the point estimates unchanged—each observation belongs to exactly one tercile, so $X'X$ is block diagonal—but the cluster-robust covariance is not block diagonal, because a bank contributes score terms to several tercile blocks over its life. Writing $s_g = (X_{g1}'u_{g1},\ X_{g2}'u_{g2},\ X_{g3}'u_{g3})'$ for bank g's stacked scores and $A = \text{blkdiag}\big((X_j'X_j)^{-1}\big)$,

$$\text{Var}(\hat{b}) \;=\; A\Big[\textstyle\sum_g s_g s_g'\Big]A, \qquad \text{Var}(\hat{b}_j - \hat{b}_k) \;=\; V_{jj} + V_{kk} - 2C_{jk}. \tag{OA.12}$$

The twenty horizons use overlapping windows and are strongly correlated, so the mean over $h = 1, \ldots, 20$ requires the cross-horizon covariance as well. Since $\hat{b}_{j,h}$ has influence function $A_{j,h}X_g'u_{g,j,h}$, the mean has influence $H^{-1}\sum_h A_{j,h}X_g'u_{g,j,h}$; accumulating that scalar per bank across horizons delivers the exact variance. The procedure replicates the stored tercile coefficients to within 0.002 percentage points at all forty (horizon, sector) points.

OA.E.2 Results

TABLE OA.VI
BUFFER GRADIENT WITH CLUSTER-ROBUST STANDARD ERRORS

	Thin	Middle	Thick	Gradient	Ratio (90% CI)
C&I	−0.1718	−0.1262	−0.1307	−0.0412	1.32 [1.01, 1.62]
	(0.0127)	(0.0136)	(0.0156)	(0.0201)	
Real estate	−0.1354	−0.0959	−0.0964	−0.0391	1.41 [1.14, 1.67]
	(0.0081)	(0.0086)	(0.0095)	(0.0126)	
Pooled					1.37 [1.13, 1.61]

Mean b_h over $h = 1, \ldots, 20$ in log points ($\times 100$) per unit innovation. Cluster-robust standard errors by bank in parentheses, computed from the joint covariance described above. "Gradient" is thin minus thick. The pooled row is the minimum-variance combination of the two sector ratios, which are correlated at +0.43.

Three features are worth recording. First, both sector ratios are distinguishable from unity at the ten percent level, but the C&I interval clears it narrowly (lower bound 1.01); the real-estate gradient is the more precisely estimated of the two. Second, pooling narrows the interval by roughly a fifth relative to C&I alone. Third, the middle tercile is statistically indistinguishable from the top tercile in both sectors, so the cross-sectional contrast is between thin-buffer banks and the rest rather than a monotone gradient.

OA.E.3 A cross-sector restriction

The risk-weight mechanism of Section 3 implies a testable restriction. Because the shadow price of regulatory capital enters each sector's first-order condition multiplied by that sector's risk weight ω_j, the weight scales the *level* of a sector's response while the tercile index enters only through the shadow price. The thin-to-thick ratio should therefore be the same in both sectors. Using the joint covariance of all four coefficients,

$$\text{ratio}_{RE} - \text{ratio}_{C\&I} \;=\; +0.090 \quad (\text{s.e. } 0.186,\; t = 0.48,\; 90\%\ \text{CI } [-0.22, +0.40]), \tag{OA.13}$$

so the restriction is not rejected across the 9,801 banks in the estimation sample. This is the sharpest available evidence that the cross-sectional pattern operates through risk-weighted capital specifically, rather than through a generic tendency of constrained banks to contract; a specification in which the requirement raises the cost of equity uniformly across a bank's portfolio does not deliver the restriction.

OA.E.4 What the aggregate response does and does not identify

The identified innovation moves the requirement by $\hat{\sigma}_\kappa = 4.0$ basis points, and $\hat{\sigma}_\kappa$ is itself estimated by SMM against moments whose confidence intervals individually span zero. The *level* of the response is accordingly a fitted scale rather than an independently disciplined quantity, and the model is validated on the cross-sectional ratio rather than on the level—a choice we make deliberately. Consistent with this, regressing aggregate sectoral lending or aggregate lending rates on the identified shock over the 71 available quarters yields first-stage and reduced-form F statistics below 3.4 at every horizon, against a conventional threshold of ten. The bank-level projections, estimated on roughly 900,000 bank-quarters, are by contrast precise. The identification in this paper is cross-sectional, and the aggregate time series is uninformative at this sample length.

OA.F Incidence Across Banks

The paper asks who bears the burden of a tightening. This appendix answers it twice: once from the data alone, and once from the model with the buffer gradient left untargeted.

OA.F.1 The statutory sort and the economic sort

Statutory requirements sort banks by asset size, at the \$10 and \$50 billion thresholds. The mechanism sorts them by distance to the frontier. Computing both sorts on the same panel—capital ratios, size, and each bank's own statutory requirement, with no model input—gives:

Table OA.VII
Statutory tier against buffer position

statutory tier	banks	share of assets	share in thinnest buffer tercile
Small (<\$10B)	3,729	58.4%	30.9%
Medium (\$10–50B)	239	20.5%	65.3%
Large (>\$50B)	32	21.1%	84.4%

Statutory tier explains 1.8% of the cross-sectional variance in regulatory slack. The two sorts disagree in both directions: 86% of thin-buffer banks by count, and 47% by assets, face the *lowest* statutory requirement, while 84% of the institutions the statute treats as most systemic already hold the thinnest buffers—mechanically, because the size add-ons raise the requirement their buffer is measured against.

That the statutory margin is costly is not in question. Banks bunch below the thresholds: after the 2019 tailoring framework took effect the size distribution shows a sharp increase in the number of banks around $90 billion, just below the $100 billion category boundary, and when the $50 billion threshold was removed, upward transitions across it roughly doubled (Siedlarek & Wang, 2026). The statutory tier is therefore costly enough to shape banks' size choices and still does not identify the banks on which a tightening falls. That is the pair of findings together: the margin regulators set is expensive along a dimension close to orthogonal to incidence.[13]

OA.F.2 Model incidence, with the gradient untargeted

Evaluating the calibrated policy across the observed distribution of capital ratios, at predetermined equity, for a one-unit tightening:

TABLE OA.VIII
INCIDENCE OF A TIGHTENING BY BUFFER POSITION

	share of lending	share of burden	concentration	ΔL
thin buffer	32.1%	34.1%	1.06	−0.127%
middle	32.9%	33.0%	1.00	−0.120%
thick buffer	35.1%	33.0%	0.94	−0.113%

Two readings. Thin-buffer banks contract 13% more than thick-buffer banks in percentage terms, which is the gradient. But because they are smaller, their *share* of the aggregate contraction (34.1%) is close to proportional to their share of lending (32.1%). The burden falls disproportionately on thin-buffer banks per dollar lent; it is not concentrated on a small group.

This calculation is disciplined only by cross-sectional moments—the calibrated lending, equity and sector shares, and the observed distribution of capital ratios. It uses none of the impulse-response moments, so the buffer gradient it implies is an out-of-sample check. Computed as the local projections compute it, as a winsorised mean of within-bank log changes, the model gives a C&I low-to-high ratio of 1.321 against 1.370 in the data, inside the confidence interval [1.130, 1.610]. The real-estate gradient is 1.076 against 1.410 and lies below its interval: the model does not generate enough dispersion in mortgage responses across buffer positions, and we report that as a limitation.

[13] Because banks bunch below the thresholds, statutory tier is not fully exogenous—a bank's tier is in part a choice. The decomposition above is descriptive and we do not read it causally. The direction of selection, if anything, reinforces the pattern: a bank that holds its assets below a threshold avoids the add-on and is therefore measured against a lower requirement.

OA.F.3 Three implementation points

Equity mapping. Each bank's predetermined equity is set proportional to its observed capital ratio at a common scale. The alternative—choosing equity so that the *solved* ratio equals the observed one—is exact but truncating: a solved capital ratio can never lie below the requirement, so it discards the banks inside the frontier, which are the ones the mechanism runs through. It yields no banks at the frontier, against 2% observed.

Sector corners. The binding-regime first-order conditions do not by themselves impose non-negativity, so a severely constrained bank can be assigned negative C&I lending. Where the interior binding solution would do so, the sector is set to zero and the whole risk-weighted budget e/κ is allocated to the other sector.

Winsorisation. Bank-level responses are winsorised at the 1st and 99th percentiles before averaging, matching the treatment of the local projections they are compared against.

OA.G Sensitivity of the Quantitative Magnitudes to the Solved Policy

The model is solved by training a neural network to represent the policy function, and the reported quantitative magnitudes are computed from a particular trained checkpoint. This appendix documents how much those magnitudes depend on that choice, which of the results are invariant to it, and a structural tension that the exercise brings out.

OA.G.1 Where the cross-sectional gradient is computed, and why

The gradient reported in the text is computed from the *static* problem, which is quadratic with a linear constraint and therefore admits a closed-form solution in each KKT regime. This subsection explains why we do not compute it from the trained policy instead.

Repeating the buffer-tercile exercise on every κ-augmented checkpoint available (fourteen, spanning solved lending levels from 0.09 to 1.03 times the Phase-1 calibration target) gives a low-to-high response ratio exceeding one in *all fourteen*, with median 1.23 and range [1.11, 1.73]. That stability is better read as a property of the approximation than as robustness of the mechanism. Those checkpoints predate ϕ_j, and when regulation enters only as a constraint the requirement leaves an unconstrained bank's first-order

condition entirely, so the thick-tercile response is exactly zero and the tercile ratio is not identified by the mechanism at all. The finite values the trained policies return reflect the network smoothing the kink at the regulatory frontier—which is also where its Euler residuals are largest, by roughly a factor of fifteen relative to the slack region.

The same gap appears with ϕ_j active. On solves that reach the calibrated lending level the trained policy returns a C&I ratio between 1.02 and 1.06, against 1.32 from the closed-form solution on the same calibration, and places materially more banks at the frontier than the exact solution does; increasing network capacity does not close it. We therefore report the static computation and treat approximation error near the constraint as the principal limitation of the numerical solve. Terciles are assigned from the initial state throughout, so the comparison isolates the policy function.

OA.G.2 The impulse-response magnitudes are not

The *levels* of the responses, and everything derived from them, are another matter. Table OA.IX re-estimates the structural parameters on four checkpoints, ordered by how closely their solved lending matches the Phase-1 target.

Table OA.IX
Structural estimates by solved policy

solved lending / target	s_M	s_B	s_B/s_M	$\hat{\sigma}_\kappa$	J
0.21 (reported)	−3.07	−3.57	1.16	0.00040	1821
0.76	−0.63	−1.29	2.06	0.00055	2614
0.94	−0.14	−1.57	11.5	0.00000	2688
1.03	−0.03	−0.93	30.0	0.00000	2688

As the solved policy approaches the lending level its own calibration targets, the real-estate semi-elasticity goes to zero, the estimated shock scale $\hat{\sigma}_\kappa$ collapses, and the Hansen statistic rises to 2688—the value it takes when the model generates no response at all. The pattern is monotone.

The reason $\hat{\sigma}_\kappa$ cannot compensate is that it is a pure scale parameter, multiplying both sectors' responses equally. The empirical moments imply $s_B/s_M \approx 1.4$; once the model's own ratio is an order of magnitude larger, no common scale reconciles the two.

OA.G.3 The tension is structural

This is a property of the model rather than of the numerical solution. The binding-regime response can be derived analytically from the static first-order conditions, without reference to any trained policy, and gives

$$s_M \;=\; +0.74, \qquad s_B \;=\; -1.49. \tag{OA.14}$$

The real-estate response is *positive*: this is the Asset Substitution Margin operating as designed, since a bank facing a tighter requirement shifts toward the low-risk-weight sector. Moreover, at the calibrated steady state the representative bank holds a buffer of 3.53 percentage points above its requirement and the constraint does not bind, so a one-percentage-point tightening leaves it in the interior region where sectoral lending is invariant to κ.

The model's structural prediction for a +1pp tightening at the calibrated buffer is therefore a real-estate response near zero—mildly positive among the banks that do bind—together with a moderately negative C&I response. The checkpoints that solve the model most accurately deliver exactly that. The mechanism that makes C&I contract is the same one that makes real estate expand, so a substantial contraction in *both* sectors, which is what the empirical local projections show, is not something the present model generates.

OA.G.4 How to read the quantitative results

We therefore distinguish two classes of result. The cross-sectional and qualitative findings—the buffer gradient, the sign and ordering of the sectoral responses, and every empirical result in Section 2—do not depend on the solved policy. The cardinal magnitudes—the structural parameter estimates, the per-percentage-point credit responses, and the welfare costs computed from them—are conditional on it, and should be read as indicative of the mechanism's direction rather than as calibrated point estimates. Closing the gap between the model's sectoral response ratio and the empirical one is the natural next step, and we regard it as an open question rather than a resolved one.

OA.H Sensitivity of the Surplus Cost to the Demand Calibration

The welfare cost Λ is an area under the sectoral inverse demand curves, so its level depends on the demand slope in a way the lending responses do not. This appendix records how the surplus integral is anchored, and reports the resulting sensitivity.

OA.H.1 Anchoring

Equations (31) and (32) together place the calibrated rate $\bar{r}_j$ at lending $\bar{L}^j$ with semi-elasticity S_j. The surplus integral, however, is evaluated at the model's own steady-state lending, which is $0.25\bar{L}^M$ in real estate and $0.12\bar{L}^B$ in C&I. Because $S_j = 1/(\eta_j L_j)$ depends on the level, holding η_j fixed while the level moves rescales the very object the calibration pins: evaluated at the model's lending, the schedules imply effective semi-elasticities of 130 and 122 rather than the intended 32.5 and 14.8.

We therefore re-derive the demand block from the calibrated S_j *at the point of evaluation*. Writing L_j^0 for baseline lending,

$$\eta_j \;=\; \frac{1}{S_j L_j^0}, \qquad \theta_j \;=\; \bar{r}_j + \eta_j L_j^0 \;=\; \bar{r}_j + \frac{1}{S_j}, \tag{OA.15}$$

and the block is held fixed across the counterfactual so that both points lie on a single demand curve. The intercept is unchanged, since it depends only on S_j; only the slope moves. This lowers the headline cost from 2.90% to 1.42%.

OA.H.2 Sensitivity

Table OA.X reports Λ from the implemented schedules down to the externally cited targets.

Table OA.X
Surplus cost across the demand calibration

S_M	S_B	Λ	
32.5	14.8	1.42%	implemented schedules
16.3	7.4	0.95%	
8.1	4.0	0.60%	
4.1	4.0	0.41%	
2.0	4.0	0.27%	externally cited targets

(The Λ levels in this table are computed from a single solved policy; Online Appendix OA.G reports

that the level is stable across solved policies, so the variation shown here is attributable to the demand calibration alone.) The lending responses are identical in every row—-3.02% in real estate and -3.50% in C&I per +1pp—because η_j enters neither the bank's optimisation problem nor the reduced-form estimates. What varies across rows is only the surplus base against which those responses are valued. The ordinal content of the counterfactuals is likewise unaffected, with one exception worth stating: the uniform-versus-differentiated comparison (CF1) turns on less than 0.1% of Λ under either anchoring, and its sign is not robust. We therefore read CF1 as showing that size-tailoring does not materially change the aggregate burden, and place no weight on the direction of the residual difference.

OA.I Distribution of Regulatory Slack by Bank Size

This appendix complements Stylized Fact 5 of the main text by documenting the full within-tier distribution of regulatory slack $D_{it} = r_{it} - \kappa_t$ across bank size cohorts over the sample period 1993Q2–2019Q4. Figure OA.3 plots the median, interquartile range (P25–P75), and interdecile range (P10–P90) of D_{it} for each quarter, separately for small (< \$10B), medium (\$10B–\$50B), and large (> \$50B) banks. Values are winsorised at the 1st and 99th percentiles within each quarter to limit the influence of reporting outliers.

Three features stand out. First, small banks maintain persistently higher median buffers than their larger counterparts throughout the sample — approximately 16 percentage points versus 11 pp for medium and large banks in the pre-crisis period. This level difference is the cross-sectional counterpart of the model's Insulated Zone: because small banks face high wholesale funding frictions ($\gamma_{\text{small}} \gg 0$), they self-insure by holding large precautionary capital buffers, keeping them far from the regulatory constraint even in the face of adverse shocks.

Second, the within-tier dispersion of slack differs sharply across size cohorts. Small banks exhibit an interquartile range of roughly 8 pp and a P10–P90 spread of approximately 18 pp, reflecting substantial heterogeneity in capitalisation even among institutions of similar scale. By contrast, medium and large banks display much tighter distributions pre-crisis (IQR $\approx$ 2 pp), consistent with the model's prediction that banks with low wholesale funding costs optimise close to the regulatory frontier and cluster near the kink of the occasionally binding constraint.

Third, all three tiers show a structural upward shift in buffers following the Dodd-Frank Act (2010Q3,

highlighted in Figure OA.3). The median buffer for large banks rises from roughly 11 pp pre-Dodd-Frank to 14 pp by 2013, while the upper tail of the medium-bank distribution widens noticeably. Small banks, by contrast, show little change in median slack — their buffers were already well above the new effective minima — but a modest compression of the lower tail is visible, consistent with the tighter supervisory environment inducing the most thinly capitalised community banks to build additional precautionary reserves.

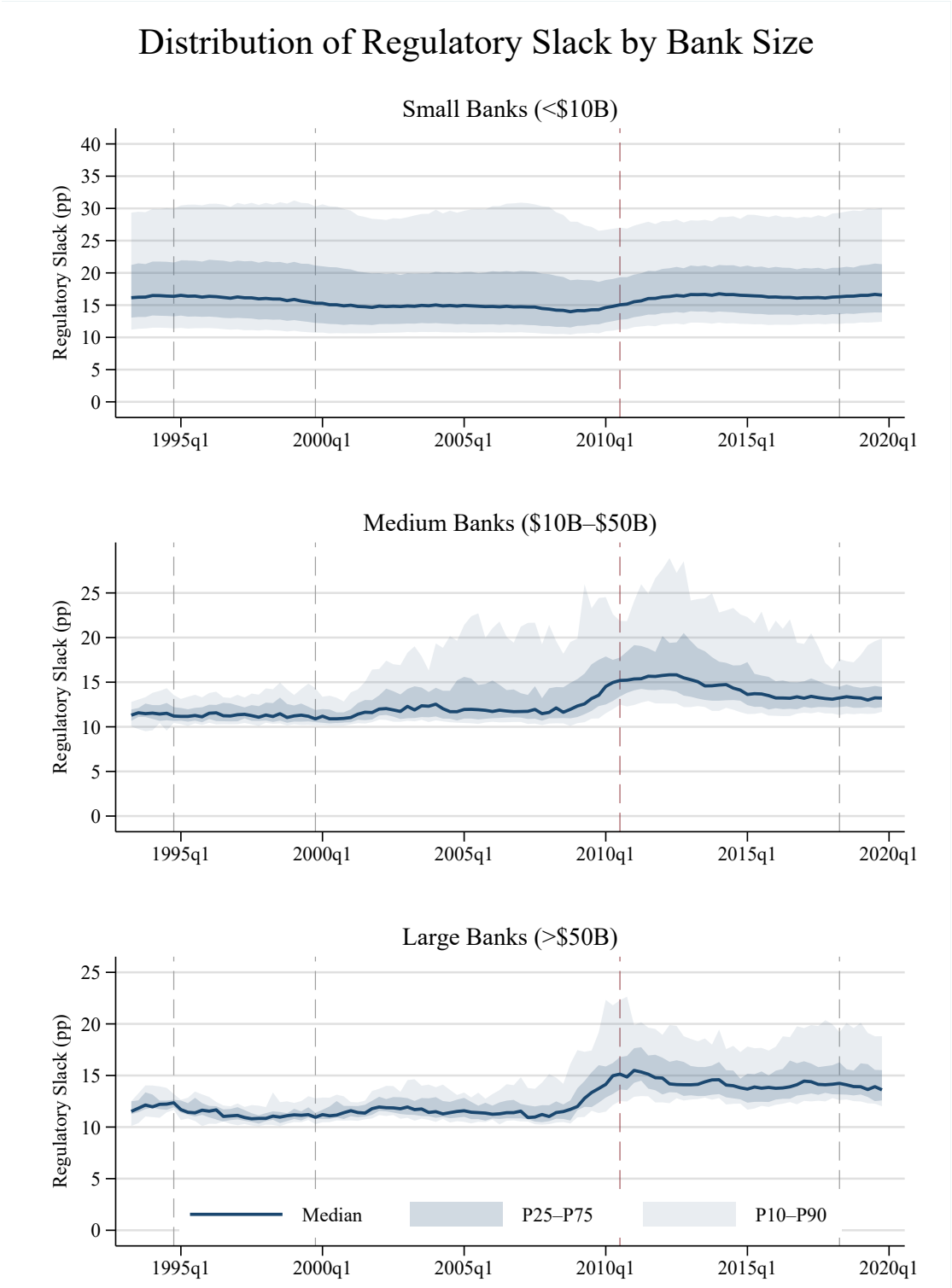


FIGURE OA.3.—DISTRIBUTION OF REGULATORY SLACK BY BANK SIZE, 1993Q2–2019Q4

Notes: Each panel plots the median (solid line), interquartile range P25–P75 (dark shading), and interdecile range P10–P90 (light shading) of winsorised regulatory slack $D_{it} = r_{it} - \kappa_t$ (percentage points) by quarter. Dashed vertical lines mark the Riegle-Neal Act (1994Q4), the Gramm-Leach-Bliley Act (1999Q4), the Dodd-Frank Act (2010Q3, highlighted), and the EGRRCPA (2018Q2). Size thresholds follow the main text: small (< $10B), medium ($10B–$50B), large (> $50B). Sample: quarterly bank panel from US Call Reports; approximately 13,500 small, 50–100 medium, and 7–42 large bank-quarter observations per period.

These empirical patterns are quantitatively consistent with the model's equilibrium predictions. In the calibrated economy, the wholesale funding friction γ generates an endogenous size gradient in precautionary buffers that matches the observed median difference across tiers (Section 5.2 of the main text). The sensitivity analysis in Section OA.D.2 confirms that eliminating this friction collapses the cross-sectional buffer distribution, removing the Insulated Zone and amplifying the aggregate credit response to capital requirement shocks.