Financial Intermediation and Aggregate Demand: A Sufficient Statistics Approach

Yu-Ting Chiang[†] Piotr Zoch[‡]

August 27, 2024

Click here for the most recent version.

Abstract

We show the financial sector's asset supply elasticities are sufficient statistics summarizing its macroeconomic effects for a large class of financial frictions. We demonstrate their usefulness for quantitative macroeconomic analysis in the context of models with household heterogeneity and illiquidity. They are crucial for policy questions ranging from the size of fiscal multipliers to the effectiveness of targeting the financial sector vs. households. Workhorse models imply different values of these elasticities, generating output responses to policies that can differ by orders of magnitude. We show how to construct empirical measures of these elasticities.

Keywords: financial frictions, sufficient statistics, HANK, monetary and fiscal policy

JEL code: E2, E6, H3, H6

[†]Federal Reserve Bank of St. Louis, contact email: yu-ting.chiang@stls.frb.org.

[‡]University of Warsaw and FAME|GRAPE, contact email: p.zoch@uw.edu.pl

We thank Fernando Alvarez, Adrien Auclert, Anmol Bhandari, Harris Dellas, Bill Dupor, Mikhail Golosov, Lars Hansen, Greg Kaplan, Ralph Luetticke, Kurt Mitman, B. Ravikumar, Matthew Rognlie, Wenting Song, Ludwig Straub, Vincent Sterk, Mikayel Sukiasyan, Harald Uhlig, Christian Wolf, and Mark Wright for valuable suggestions. Zoch acknowledges the support of National Center for Science (2019/35/N/HS4/02189). Views in this paper do not necessarily represent those of the Federal Reserve System.

1 Introduction

Financial intermediation is central to macroeconomics, influencing the transmission of policies and shocks through asset markets. However, models of financial intermediation often rely on various microfoundations with frictions governed by parameters that are intrinsically hard to measure, such as asset diversion rates or monitoring costs. The intricacy of these models makes it challenging to identify and quantify features of the intermediation process that are most relevant for aggregate outcomes. This poses an obstacle to integrating financial frictions into modern quantitative macro models, which are themselves becoming increasingly complex. Recent approaches address this complexity by identifying key features shared across a wide range of models for each "block" of the economy. For instance, a growing consensus suggests that households' intertemporal marginal propensities to consume effectively summarize their aggregate responses across various models (Auclert et al., 2023). Yet, no similar attempt has been made to derive a counterpart for the financial sector. Our paper addresses this issue.

Our main idea is a simple observation that financial intermediaries are, effectively, suppliers of assets: They take one type of asset, such as loans, and transform it into a different type, such as deposits. As they are suppliers of assets, their supply curves fully describe how they respond to changes in prices and quantities of assets. To the extent that we know the shape of their asset supply, details of the intermediation process are irrelevant.

We build on this observation to derive a set of asset supply elasticities that serve as sufficient statistics to summarize frictions in financial intermediation models. These elasticities allow us to incorporate financial frictions into state-of-the-art quantitative macroeconomic models while remaining agnostic about their microfoundations. We derive formulas to isolate the channels through which these elasticities affect aggregate outcomes. We measure these elasticities empirically and compare them to implicit assumptions in standard macroeconomic models. These assumptions significantly affect conclusions about policy issues ranging from the impact of government spending and tax cuts to the effect of asset purchase programs, with output responses varying by up to orders of magnitude. Our sufficient statistics provide empirical discipline on financial frictions that is necessary for analyzing these policies.

We derive these results in a general framework that nests models of financial intermediation with various microfoundations and allows for household heterogeneity and illiquidity. Households consume and save in different assets, with some assets being "liquid" and thus preferable. The financial sector issues liquid assets and holds illiquid capital, supplying liquidity to the economy under frictions. Production is subject to nominal rigidities. The government influences aggregate demand through spending, taxes, and transfers, as well as policies that transmit through the asset markets, such as interest rates, government debt issuance, and illiquid asset purchases.

We show there exists a simple structure that contains the class of intermediation frictions nested in our framework, including those originating from asset diversion (Gertler and Karadi, 2011), costly state verification (Bernanke et al., 1999), costly leverage (Cúrdia and Woodford, 2016), and collateral constraints (Kiyotaki and Moore, 1997). The structure describes how the financial sector's leverage responds to expected returns over different time horizons. The magnitude of responses is governed by two *leverage sensitivity* parameters, while the dependency on return horizons is controlled by a *forward-looking component*. This structure can be directly mapped to data, allowing us to summarize intermediation frictions without taking a strong stance on a specific microfoundation. Using this structure, we characterize a set of *liquidity supply elasticities* for the financial block of the model, describing how the level of intermediation responds to expected and realized returns over an infinite time horizon.

To show that these elasticities are sufficient statistics, we recast the economy into an intertemporal demand-and-supply system of goods and assets. In the system, the financial block interacts with other model components only through returns and quantities of assets, and therefore, its elasticities contain all relevant information up to first-order approximation. This financial block is "portable" in the sense that we can characterize and estimate it independently of other model components, such as the household and production blocks. This feature facilitates its integration into a wide range of quantitative macroeconomic models with broad applicability, including common representative and heterogeneous agent frameworks. Our representation of the economy employs a sequence space approach similar to that in Auclert et al. (2021), Wolf (2021b), Dávila and Schaab (2023), McKay and Wolf (2023), and Angeletos et al. (2023). While these works abstract away from financial intermediation, we show that the same tools are useful for understanding financial frictions, and these frictions are crucial for policy analysis.

Using the demand-and-supply representation, we show that the financial sector influences policy transmission through an *asset market channel*. The channel depends on the *cross-price elasticities* of liquidity supply with respect to returns on capital. Low elasticities indicate that intermediaries view liquid assets and capital as less substitutable; other things equal, an increase in liquidity (e.g., due to government debt issuance) leads to large increases in capital prices and raises aggregate demand through investment and consumption. Despite their crucial role in asset market transmission, standard macroeconomic models assume a broad range of values for these elasticities, from zero to infinity. These assumptions are often embedded in model setups or implicit in functional forms. Our approach makes these assumptions explicit, emphasizing the need for empirical measurement to validate or refine them.

We demonstrate how to empirically discipline these elasticities by estimating the leverage sensitivities and the forward-looking component for the U.S. banking sector. To address potential identification threats, we construct instruments from structural shocks using common proxies for monetary policy shocks (Bauer and Swanson, 2023), oil supply shocks (Baumeister and Hamilton, 2019), and intermediary net worth shocks (Ottonello and Song, 2022). Our estimates indicate that the U.S. banking sector's liquidity supply elasticities are twice as large as those implied by functional forms in standard financial intermediation models. These estimates provide useful target moments for calibrating financial frictions in quantitative models, including those studying how financial intermediation interacts with complex household consumption-saving behaviors, such as Lee et al. (2020), Fernández-Villaverde et al. (2020), Lee (2021), Mendicino et al. (2021), Cui and Sterk (2021), and Ferrante and Gornemann (2022). Our theoretical result implies that, for frictions nested in our framework, targeting the moments we estimate ensures financial intermediation can best represent empirical features relevant to the interaction.

We illustrate the quantitative importance of our sufficient statistics with two policy questions: (1) the size of government spending multiplier, and (2) a "Wall Street vs. Main Street" debate: Can asset purchase programs stimulate the economy more effectively than transferring resources to households with tax cuts? To study these questions, we specialize the household sector to a standard two-asset heterogeneous agent model, calibrated to match asset holdings and consumption responses in the data. Keeping all else equal, we vary the financial sector's liquidity supply elasticities from our empirical measures to values implied by common assumptions and compare their implications for the two policy questions.

The effects of policies vary significantly across specifications, with inelastic supply leading to stronger reactions due to asset market responses. Government spending multipliers can differ by up to a factor of two, with greater variation observed under a higher degree of debt financing as it has a larger impact on asset markets. The "Wall Street vs. Main Street" debate depends even more on the elasticities: Output responses to asset purchases differ by orders of magnitude, and the effects of tax cuts vary by a factor of three. Since our empirical elasticities are relatively high, they imply modest asset market responses relative to standard financial intermediation models. Consequently, these elasticities indicate that targeting households with tax cuts may be more effective in stimulating output than asset purchases, which rely heavily on the asset market channel. These results highlight the importance of our sufficient statistics in providing the necessary empirical discipline on financial frictions for policy analysis.

Finally, our approach has its limitations. While the elasticities we characterize are invariant to policies that take effect through prices and quantities of assets, they may not be invariant to macroprudential regulations. Such regulations directly affect intermediation frictions and change the financial sector's sensitivities to returns, thereby altering the asset supply system. Studying these regulations requires microfounded models, and our elasticities should only serve as target moments for calibration. Furthermore, our characterization relies on first-order approximation and cannot capture nonlinear dynamics like those in Brunnermeier and Sannikov (2014). Nevertheless, our approach allows us to incorporate financial frictions into quantitative macro models, addressing policy questions often precluded by simplifying assumptions in nonlinear macro-finance models. Moreover, by providing a theoretical foundation for the measurements needed to discipline financial frictions, our approach is a first step toward quantifying these frictions in nonlinear models.

2 Model

2.1 Households

Time is discrete, $t \in \{0, \dots, \infty\}$. Households are indexed by $i \in [0, 1]$ and have time separable preferences. Households derive utility from final goods consumption $c_{i,t}$, holdings of illiquid and liquid assets $a_{i,t}$ and $b_{i,t}$, and disutility from labor $h_{i,t}$. Illiquid and liquid assets pay real returns r_t^A and r_t^B, and trading of illiquid assets incurs portfolio adjustment costs. Preferences can be type-dependent and indexed by i. Each household solves the following maximization problem:

$$\max_{c_{i,t}, a_{i,t}, b_{i,t}} \mathbb{E} \sum_{t=0}^{\infty} \beta_i^t \left[u_i \left(c_{i,t}, a_{i,t}, b_{i,t} \right) - \nu_i(h_{i,t}) \right],$$

subject to budget constraints

$$a_{i,t} + b_{i,t} + c_{i,t} + \Phi(a_{i,t}, a_{i,t-1}, r_t^A) = (1 + r_t^A) a_{i,t-1} + (1 + r_t^B) b_{i,t-1} + y_{i,t} - \mathcal{T}_t(y_{i,t})$$

with borrowing constraints $a_{i,t} \geq \underline{a}$, $b_{i,t} \geq \underline{b}$ and real labor income $y_{i,t} = z_{i,t} \frac{W_t}{P_t} h_{i,t}$. The real income of households depends on idiosyncratic earnings shocks $z_{i,t}$, nominal wage per efficiency unit of labor, W_t, and the price of the final good, P_t. Households form expectations over idiosyncratic shocks $z_{i,t}$. Labor $h_{i,t}$ is taken as exogenous by each household and is determined by monopolistically competitive labor unions to be described shortly. Income tax is given by tax function $\mathcal{T}_t(y_{i,t})$.

Our formulation of households is general enough to encompass standard representative agent models ($z_{i,t} \equiv 1$ with no preference heterogeneity), assets-in-the-utility models, the spender-saver type two-agent models, and the Bewley-Hugget-Aiyagari-Imrohorglu type heterogeneous agent models. While representative agent models provide useful benchmarks for illustration, heterogeneous agent models generate consumption and asset allocation behaviors that are quantitatively crucial for aggregate responses to policies. For ease of exposition, we assume there are no aggregate shocks. However, up to first-order approximation, all results apply with the presence of aggregate shocks.

2.2 Production

Final goods production: A representative firm produces final goods y_t with capital k_{t-1} and differentiated types of labor, $h_{\ell,t}$, supplied by unions indexed by $\ell \in [0,1]$:

$$y_t = k_{t-1}^{\alpha} h_t^{1-\alpha}, \quad h_t = \left(\int h_{\ell,t}^{\frac{\varepsilon_W - 1}{\varepsilon_W}} d\ell \right)^{\frac{\varepsilon_W}{\varepsilon_W - 1}},$$

where $\varepsilon_W > 1$ is the elasticity of substitution between labor types. Given nominal wages $\{W_{\ell,t}\}$ and capital rental rate R_t, the firm chooses capital and labor to maximize profit:

$$\max_{k_{t-1}, \{h_{\ell,t}\}} P_t y_t - R_t k_{t-1} - \int W_{\ell,t} h_{\ell,t} d\ell.$$

Labor supply: Unions are monopolistically competitive. To supply labor $h_{\ell,t}$, each union combines labor from households: $h_{\ell,t} = \int z_{i,t} h_{i,\ell,t} di$, following an exogenous allocation rule, $h_{i,\ell,t} = l(z_{i,t}) h_{\ell,t}$ such that $\int z_{i,t} l(z_{i,t}) di = 1$. Given labor demand, unions set nominal wage growth $\pi_{W,\ell,t} := \frac{W_{\ell,t}}{W_{\ell,t-1}} - 1$ to maximize utilitarian welfare of households, subject to a wage adjustment cost:

$$\sum_{t=0}^{\infty} \int \beta_i^t \left[u_i \left(c_{i,t}, a_{i,t}, b_{i,t} \right) - \nu_i(h_{i,t}) - \frac{\kappa_W}{2} \pi_{W,\ell,t}^2 d\ell \right] di,$$

where $h_{i,t} = \int h_{i,\ell,t} d\ell$. Wage adjustment cost is borne as disutility by unions and does not affect the resource constraint; $\kappa_W > 0$ parameterizes the level of nominal wage rigidity. The symmetry between unions implies each household's nominal wages sum to $z_{i,t} W_t h_{i,t}$, where W_t is the ideal wage index. In Appendix D.2, we show our results do not depend on whether nominal rigidity takes the form of price or wage rigidity.

Capital: The aggregate capital stock has the following law of motion:

$$k_t = \left(1 - \delta + \Gamma\left(\iota_t\right)\right) k_{t-1}, \quad \iota_t := \frac{x_t}{k_{t-1}},$$

where x_t, ι_t denote the investment level and investment rate, δ is the depreciation rate, and $\Gamma(\cdot)$ captures capital adjustment cost. Let q_t denote the price of capital. Holding capital from periods t to $t+1$ earns a return on capital:

$$1 + r_{t+1}^K = \max_{\iota_{t+1}} \frac{R_{t+1}/P_{t+1} + q_{t+1}\left(1 + \Gamma\left(\iota_{t+1}\right) - \delta\right) - \iota_{t+1}}{q_t}. \tag{1}$$

2.3 The Financial Sector

Capital is held by a passive mutual fund and an intermediary:

$$k_t = k_t^F + k_t^B.$$

Mutual fund: The mutual fund holds capital k_t^F and intermediary net worth n_t. The fund constitutes households' total illiquid asset holdings with value $a_t = q_t k_t^F + n_t$. As a result, the return on illiquid assets is given by the value-weighted average of returns on capital r_t^K and returns on intermediary net worth r_t^N:

$$r_{t+1}^A = \frac{1}{a_t}(r_{t+1}^K q_t k_t^F + r_{t+1}^N n_t). \tag{2}$$

Intermediary: The intermediary can transform illiquid capital into liquid assets, thereby supplying liquidity to the economy. In each period, the intermediary issues liquid assets $\tilde{d}_t$ and holds capital and government debt, k_t^B and b_t^B. We assume government debts are perfect substitutes for liquid assets, and the *net liquidity supply* is given by $d_t \coloneqq \tilde{d}_t - b_t^B$.[1] The intermediary's ability to transform assets can, for example, represent its superior capability to manage loans. Without the intermediary, households would need to perform the task themselves and incur the cost $\Phi(\cdot)$.

The intermediation of assets is subject to frictions. Given net worth n_t, the intermediary can hold capital and supply liquidity with leverage Θ_t:

$$q_t k_t^B = \Theta_t n_t, \quad d_t = (\Theta_t - 1) n_t,$$

where the expression for d_t is implied by the intermediary's balance sheet, $q_t k_t^B = n_t + d_t$. The level of leverage is governed by a function $\Theta(\cdot)$, representing the intermediation frictions:

$$\Theta_t = \Theta(\{r_{s+1}^K, r_{s+1}^B\}_{s \geq t}). \tag{3}$$

Function $\Theta(\cdot)$ depends on the entire path of future returns $\{r_{s+1}^K, r_{s+1}^B\}_{s \geq t}$, which respectively stand for the intermediary's future investment opportunity and funding cost. This formulation of the frictions allows us to nest a class of financial intermediation models as special cases, along with a few useful extensions of these models. We

[1] Since liquid assets and government debt are perfect substitutes, whether households hold government debt directly or not is irrelevant.

discuss this nesting result in Section 3.1.

Given leverage Θ_t, the intermediary's balance sheet implies the return on net worth satisfies:

$$r^N_{t+1} = \Theta_t(r^K_{t+1} - r^B_{t+1}) + r^B_{t+1}.$$

We assume that the law of motion for net worth follows

$$n_t = G(\Theta_{t-1}, r^K_t, r^B_t)n_{t-1} + m. \tag{4}$$

Net worth evolves as a fraction $G(\cdot)$ of the past net worth plus an exogenous net worth inflow m. Together, they describe the net allocation of resources from the passive mutual fund. Function G depends on endogenous variables predetermined at time t: past leverage Θ_{t-1} and realized returns r^K_t, r^B_t. This formulation contains the common exogenous net worth specification as in Gertler and Kiyotaki (2010). Moreover, we show in Appendix B.3 that, for aggregate responses, this formulation is equivalent to a class of models with endogenous dividend and equity issuance decisions.

2.4 Government

The government sets a sequence of government purchases g_t, government debt b^G_t, liquid rate target r^B_t, total tax revenue T_t, and illiquid assets holdings a^G_t. The government debt is real debt, and monetary policy adjusts the nominal interest rate to keep the real liquid rate at its target for all $t > 0$. The liquid rate r^B_0 in period 0 is predetermined. The government collects tax revenue through the tax system $\mathcal{T}_t(y_{i,t}) = y_{i,t} - (1-\tau_t)y_{i,t}^{1-\lambda}$. Given $\{y_{i,t}\}$, tax rate τ_t is set such that $T_t = \int \mathcal{T}_t(y_{i,t})di$. The government faces budget constraints:

$$b^G_t - (1 + r^B_t)b^G_{t-1} = a^G_t - (1 + r^A_t)a^G_{t-1} + g_t - T_t. \tag{5}$$

2.5 Definition of Equilibrium

Given $\left\{g_t, b^G_t, r^B_t, T_t\right\}$, an equilibrium consists of prices $\left\{P_t, R_t, W_{\ell,t}, q_t, r^A_t, r^K_t\right\}$ and allocations $\{y_t, c_{i,t}, x_t, h_t, h_{i,\ell,t}, k_t, k^F_t, k^B_t, a_t, a^G_t, a_{i,t}, b_{i,t}, n_t, d_t\}$ such that: (1) households maximize utility subject to constraints; (2) firms maximize profit and investment rate maximizes the return on capital, (3) nominal wages maximize payoff of the labor unions; (4) the intermediary's capital holdings and liquidity supply is given by the intermediation frictions and the net worth process; (5) the illiquid return r^A is consistent with the balance sheet of the mutual fund; (6) the government budget constraint holds given the tax system, and (7) markets clear:

$$\int (c_{i,t} + \Phi_{i,t})\, di + x_t + g_t = y_t, \quad \int b_{i,t} di = d_t + b_t^G, \quad \int a_{i,t} di + a_t^G = n_t + q_t k_t^F,$$

where (i) in the goods market, output equals the total of consumption, investment, and government purchases; (ii) in the liquid asset market, households' liquid assets holdings equal the total supplied by the intermediary and the government; and (iii) in the illiquid asset market, the total of household and government's holdings of illiquid assets is equal to the value of assets held by the fund. The capital market clears when the total capital held by the intermediary and the fund equals the aggregate capital stock.

3 Financial Intermediation and Liquidity Supply

We characterize financial intermediation in this section. We show the formulation in Section 2 contains a large class of financial intermediation models with various objective functions and different constraints. Despite the dissimilarity, these models share a simple structure that describes how leverage responds to expected returns. This structure clarifies the distinct strengths and restrictions of the nested models and provides a concise summary of them. We use this unified structure to develop a liquidity supply system that describes how asset intermediation responds to returns. The structure allows us to systematically study how frictions in the financial sector interact with other model components to affect aggregate outcomes in Section 4.

Although we focus on the intermediary's role in transforming productive capital into liquid assets, our characterization applies more broadly to other types of transformation between different asset categories as long as the intermediation process is subject to the frictions nested in our framework.

3.1 Nesting Models of Financial Intermediation

We provide an overview of the nested models and lay out details in Appendix B.1.

Model 1, asset diversion (Gertler and Kiyotaki (2010), Gertler and Karadi (2011)):

An intermediary can divert a fraction $1/\theta$ of assets. If that happens, depositors force it into bankruptcy. To avoid asset diversion, intermediation is limited by the intermediary's continuation value $v_t(n_t) = \eta_t n_t$:

$$q_t k_t^B \leq \theta \eta_t n_t, \quad \eta_t = \Lambda_{t,t+1} \left(f + (1-f)\, \eta_{t+1}\right) \left[1 + r_{t+1}^B + \left(r_{t+1}^K - r_{t+1}^B\right) \theta \eta_t\right],$$

where $\Lambda_{t,t+1}$ denotes the intermediary's discount factor.[2]

Model 2, costly state verification (Bernanke et al. (1999)): Intermediaries receive idiosyncratic returns on assets, which depositors can only observe by incurring a monitoring cost. The intermediary's capital holdings are given by a function ψ^{BGG} that depends on the distribution of idiosyncratic returns and the monitoring cost:

$$q_t k_t^B = \psi^{BGG}\left(\frac{1 + r_{t+1}^K}{1 + r_{t+1}^B}\right) n_t.$$

Model 3, costly leverage (Uribe and Yue (2006), Chi et al. (2021) and Cúrdia and Woodford (2016)): Intermediaries need to incur a convex cost $\Upsilon(\psi_t) n_t$ that depends on leverage $\psi_t = \frac{q_t k_t^B}{n_t}$. Optimal leverage is linked to the spread between returns:

$$r_{t+1}^K - r_{t+1}^B = \Upsilon'\left(\psi_t\right).$$

Model 4, collateral constraint (similar to Kiyotaki and Moore (1997), Bianchi and Mendoza (2018), Ottonello et al. (2022)): Intermediation is limited by a fraction $\vartheta < 1$ of collateral value backing it. If the collateral value includes the value of capital and the associated return,[3] we have:

$$\left(1 + r_{t+1}^B\right) d_t \leq \vartheta \left(1 + r_{t+1}^K\right) q_t k_t^B.$$

These models of financial intermediation together with the law of motion for net worth (Equation 4) imply sequences of $\{q_t k_t^B, d_t\}$. The following lemma states that these models are nested by the setup described in Section 2.3.

Lemma 1 *Given $\{q_t k_t^B, d_t\}$ in model $j \in \{1, \ldots, 4\}$, there exists* $\Theta(\{r_{s+1}^K, r_{s+1}^B\}_{s \geq t})$

[2]We allow for any discount rate of the form $\Lambda_{t,t+1} = \Lambda(r_{t+1}^K, r_{t+1}^B)$.

[3]The exact form of constraints differs among models, depending on what can be pledged as collateral. We discuss different variations in Appendix B.1.

such that $q_t k_t^B = \Theta_t n_t$ *and* $d_t = (\Theta_t - 1)n_t$. *Moreover, around the steady state,*

$$\frac{\partial \Theta_t}{\partial r_{s+1}^K} = \gamma^{s-t}\ \bar{\Theta}_{r^K}, \quad \frac{\partial \Theta_t}{\partial r_{s+1}^B} = -\gamma^{s-t}\ \bar{\Theta}_{r^B}, \quad \forall s \geq t,$$

where $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}, \gamma \geq 0$ *are given by parameters in model* j *and steady-state variables.*

Proof. See Appendix B.1. □

Parameters $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}, \gamma$ describe how the financial sector's ability to intermediate assets and supply liquidity depends on expected returns: $\bar{\Theta}_{r^K}$ and $\bar{\Theta}_{r^B}$ are *leverage sensitivities* that govern how strongly Θ_t responds, and γ is a *forward-looking component* that controls how much Θ_t responds to returns at horizon $s-t$.

In asset diversion models, $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}, \gamma > 0$. Because $\gamma > 0$, intermediation is forward-looking in the sense that it depends on future returns over different horizons. However, these models impose a strict restriction on the leverage sensitivities $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}$: these parameters are fully determined by the steady-state levels of leverage and returns, and there is no extra flexibility in the microfounded model to control them.

By contrast, costly state verification and costly leverage models feature no forward-looking component, $\gamma = 0$, and intermediation does not respond to expected returns beyond the next period. Yet, unlike the rigid form imposed by asset diversion models, these models feature an extra degree of freedom to control the leverage sensitivities $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}$. These sensitivities are determined by the monitoring cost and the distribution of idiosyncratic returns for costly state verification models, and they are governed by the curvature of the leverage cost function for costly leverage models. Similarly, models with collateral constraints feature $\gamma = 0$, as changes in collateral value are captured by changes in r_{t+1}^K. However, the leverage sensitivities $\bar{\Theta}_{r^K}$ and $\bar{\Theta}_{r^B}$ are restricted by steady-state leverage and returns, similarly to asset diversion models.

Besides nesting models common in the literature, our formulation includes several generalizations of existing models. For example, we show that a costly leverage model augmented with the dynamic structure of asset diversion models can deliver both the flexibility of leverage sensitivities $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}$ and a forward-looking component $\gamma > 0$. We discuss these generalizations in Appendix B.2.

3.2 Liquidity Supply Elasticities

The intermediation frictions $\Theta(\{r^K_{t+s}, r^B_{t+s}\})$ and net worth process $G(\Theta_{t-1}, r^K_t, r^B_t)$ together determine the financial sector's ability to intermediate assets and supply liquidity. This ability is described by a liquidity supply function, $\mathcal{D}_t(\{r^K_s, r^B_s\}_{s=0}^{\infty})$, which governs how much liquidity d_t the intermediary can supply given returns $\{r^K_s, r^B_s\}_{s=0}^{\infty}$. Up to first-order approximation, how liquidity supply responds to changes in returns is characterized by its elasticities.

Proposition 1 *The cross-price semi-elasticities of liquidity supply around the steady state are given by:*

$$\frac{\partial \mathcal{D}_t/\partial r^K_s}{\mathcal{D}_t} = \begin{cases} \gamma^{s-t-1}\bar{\Theta}_{r^K}\left(\frac{1}{\Theta-1} + \gamma\Sigma(t)\right), & s > t, \\ \left(\bar{G}_{r^K} + \bar{\Theta}_{r^K}\Sigma(s)\right)\bar{G}^{t-s}, & s \leq t, \end{cases}$$

where $\Sigma(s) \coloneqq \bar{G}_{\Theta}\frac{1-(\gamma\bar{G})^s}{1-\gamma\bar{G}}$, *and* $\bar{G}, \bar{G}_{\Theta}, \bar{G}_{r^K}$ *are the steady-state values and derivatives of function* G*. The own-price semi-elasticities* $\frac{\partial \mathcal{D}_t/\partial r^B_s}{\mathcal{D}_t}$ *are given by the same formula with* $\bar{\Theta}_{r^K}$ *and* $\bar{G}_{r^K}$ *replaced by* $-\bar{\Theta}_{r^B}$ *and* $\bar{G}_{r^B}$*.*

Proof. See Appendix A.1. □

The intermediation frictions determine how the financial sector's liquidity supply responds to changes in returns. For example, the cross-price elasticities $\frac{\partial \mathcal{D}_t/\partial r^K_s}{\mathcal{D}_t}$ are positive and increasing in $\bar{\Theta}_{r^K}$. If cross-price elasticities are high, the financial sector is willing to provide much more liquidity in response to an increase in r^K_s. In other words, only a small decrease in r^K_s is necessary for the financial sector to increase its holdings of government debt (therefore reducing its net liquidity supply) by a given amount. In this sense, capital and liquid assets are more substitutable with high cross-price elasticities.

The formula in Proposition 1 describes how liquidity supply at time t responds to changes in returns at time s. If $s > t$, an increase in r^K_s directly increases liquidity supply through relaxing intermediation friction Θ_t with sensitivity $\bar{\Theta}_{r^K}$. Moreover, it relaxes frictions in all periods before t with decay rate γ, and increases liquidity supply through the accumulation of net worth. Function $\Sigma(t)$ summarizes the accumulative effect. On the other hand, an increase in past return r^K_s with $s \leq t$ has no direct effect on Θ_t. It affects liquidity supply only through the propagation of net worth: Net worth at time s increases directly by $\bar{G}_{r^K}$ and the accumulation from all periods before is given by $\Sigma(s)$; both propagate to period t with rate $\bar{G}$.

In the next section, we show that these liquidity supply elasticities are sufficient statistics summarizing how the financial sector affects aggregate outcomes. They cannot be estimated nonparametrically without any structure, as they are infinite dimensional objects. Proposition 1 is useful because it provides one such structure, containing a large class of models with distinct microfoundations but governed by only a few parameters. Since these parameters describe a structural relationship between empirically observable objects such as leverage and returns, we can obtain an empirical summary of this class of models by estimating these parameters directly. Moreover, these elasticities are policy invariant to the extent that the policies under consideration affect the financial sector through changes in returns. By focusing on these elasticities, we can study a large set of macroeconomic policies without taking a stance on the exact microfoundation of the underlying frictions.

4 Aggregate Responses to Policies

We now use the financial sector's asset supply system to study how it interacts with the rest of the economy to determine aggregate responses to policies.

4.1 A Demand-and-Supply Representation

We recast the economy as a demand-and-supply system of goods and assets. Given government policies and key aggregate variables, we solve the optimization problem for each agent to obtain their aggregate behavior along the transition path. Our result in Section 3.2 shows how the *financial block* of the economy implies a liquidity supply function, $\mathcal{D}_t$, that summarizes how financial intermediaries respond to aggregate conditions through $\{r_s^K, r_s^B\}$. The same logic applies to the *household block* of the model: Given a sequence of output, taxes, returns, and the initial asset position, we can solve the households' consumption-saving problem to obtain a consumption function, $\mathcal{C}_t$, and liquidity demand function, $\mathcal{B}_t$.[4] Similarly, we obtain an investment

[4]We define the consumption function, $\mathcal{C}_t$, to include both final goods consumed by the households, $c_{i,t}$, and the portfolio adjustment cost, $\Phi_t(a_{i,t}, a_{i,t-1})$.

function, $\mathcal{X}_t$, from the *production block*. Lemma 2 represents the equilibrium of the model as that of a demand-and-supply system.

Lemma 2 *There exist functions $\mathcal{C}_t, \mathcal{B}_t$, and $\mathcal{X}_t$, such that, given government policies $\left\{g_s, T_s, r_s^B, b_s^G\right\}_{s=0}^{\infty}$, the equilibrium output and returns on capital $\left\{y_s, r_s^K\right\}_{s=0}^{\infty}$ solve:*

$$\mathcal{C}_t(\{y_s, r_s^A; r_s^B, T_s\}_{s=0}^{\infty}) + \mathcal{X}_t(\{y_s, r_s^K\}_{s=0}^{\infty}) + g_t = y_t,$$

$$\mathcal{B}_t(\{y_s, r_s^A; r_s^B, T_s\}_{s=0}^{\infty}) = \mathcal{D}_t(\{r_s^K, r_s^B\}_{s=0}^{\infty}) + b_t^G,$$

where

$$r_t^A = \mathcal{R}_t^A\big(\{r_s^K; r_s^B, y_s\}_{s=0}^{\infty}; \mathcal{D}_{t-1}(\{r_s^K; r_s^B\}_{s=0}^{\infty})\big).$$

Function $\mathcal{R}_t^A$ corresponds to the accounting identity in Equation 2, and government asset holdings $\left\{a_t^G\right\}$ satisfy the government budget constraint in Equation 5. Moreover, functions $\mathcal{C}_t, \mathcal{B}_t$, and $\mathcal{X}_t$ do not depend on Θ and G.

Proof. See Appendix A.2. □

The two main equations in Lemma 2 correspond to the goods market and the liquid asset market clearing conditions, where we drop the illiquid asset market clearing condition as it is redundant by Walras' law. Given government policies, an equilibrium is described by sequences $\left\{y_t, r_t^K\right\}_{t=0}^{\infty}$ such that (1) demand for final goods equals output produced, and (2) liquidity demand equals liquidity supply.

This demand-and-supply formulation allows us to identify key features from each block of the model: The financial block enters the system *only* through the liquidity supply function $\mathcal{D}_t$. As a result, *all* relevant properties of the financial sector are contained in $\mathcal{D}_t$. Up to first-order approximation, the liquidity supply elasticities characterized in Proposition 1 are sufficient statistics that summarize how financial frictions affect aggregate outcomes. The exact microfoundations behind the frictions do not matter as long as they generate the same liquidity supply. Moreover, since $\mathcal{D}_t$ can be described independently of the household and production sectors, our result is compatible with a large class of quantitative macroeconomic models, including standard representative agent frameworks and models that emphasize realistic household consumption-saving behaviors, such as the TANK and HANK models.

Assumptions about the household sector are summarized by the consumption function $\mathcal{C}_t$ and the liquidity demand function $\mathcal{B}_t$. Lemma 2 shows that the consumption function, $\mathcal{C}_t$, plays a key role in the goods market, sharing the same emphasis with a large literature, such as Auclert et al. (2023), Auclert et al. (2021), and Wolf (2021a). However, our result also highlights households' liquidity demand $\mathcal{B}_t$ as an important feature besides their consumption responses.[5] In Appendix D.1, we characterize several canonical household specifications nested in our framework to illustrate their distinct implications on liquidity demand, including limiting cases ranging from perfectly elastic to perfectly inelastic. Together, households' liquidity demand interacts with the financial sector's liquidity supply to determine how policies affect aggregate outcomes $\{y_t, r_t^K\}$ through asset markets.

4.2 Aggregate Responses

We study first-order aggregate responses to government policies around the steady state. We focus on policies such that $\{dg_t, dT_t, dr_t^B, db_t^G, da_t^G\}_{t=0}^{\infty}$ converge to zero as $t \to \infty$ and the equilibrium in which aggregate responses converge to zero as $t \to \infty$. We use a column vector $\boldsymbol{y}$ to represent $\{y_t\}_{t=0}^{\infty}$ and $d\boldsymbol{y}$ for its first-order deviation; notation for $\boldsymbol{T}, \boldsymbol{b}^G$, $\boldsymbol{g}$ is similar. We use $\boldsymbol{r}^K$ to represent $\{r_{t+1}^K\}_{t=0}^{\infty}$ and $d\boldsymbol{r}^K$ for its first-order deviation; notation for $\boldsymbol{r}^B$ follows the same convention. The sequences of returns start from period 1 because r_0^B is predetermined and r_0^K can be expressed as a function of output and expected returns, $r_0^K(\boldsymbol{y}, \boldsymbol{r}^K)$, as defined in Appendix A.3.

We characterize the equilibrium in two steps. First, we study the asset market responses by solving for returns on capital $d\boldsymbol{r}^K$ that satisfy the liquid asset market clearing condition, given policies and output $d\boldsymbol{y}$. We then use the solution for $d\boldsymbol{r}^K$ in the goods market clearing condition to find the equilibrium output responses $d\boldsymbol{y}$.

Excess Liquidity and Asset Markets Responses

To study asset market responses, we define a notion of *excess liquidity* as the difference

[5] Aguiar et al. (2021) emphasize an aggregate asset demand function, but all assets are perfect substitutes in their economy. Auclert et al. (2023) study a two-asset economy, but liquidity demand does not affect aggregate outcomes. This feature can be understood as an assumption about the financial sector, as we show in Proposition 2.

between liquidity supply and demand:

$$\mathcal{E}_t(\boldsymbol{y}, \boldsymbol{r}^K, \boldsymbol{r}^B, \boldsymbol{T}, \boldsymbol{b}^G) \coloneqq \mathcal{D}_t(r_0^K(\boldsymbol{y}, \boldsymbol{r}^K), \boldsymbol{r}^K, \boldsymbol{r}^B) + b_t^G - \mathcal{B}_t(\boldsymbol{y}, \boldsymbol{r}^A(\boldsymbol{r}^K, \boldsymbol{r}^B, \boldsymbol{y}), \boldsymbol{r}^B, \boldsymbol{T}),$$

where $\boldsymbol{r}^A(\boldsymbol{r}^K, \boldsymbol{r}^B, \boldsymbol{y})$ denotes functions $\mathcal{R}_t^A(\cdot)$ in vector form, representing the accounting identity in Equation 2. We use $\boldsymbol{\epsilon}_{r^K}$ to denote derivatives of $\mathcal{E}$ with respect to $\boldsymbol{r}^K$, where $\boldsymbol{\epsilon}_{r^K}(t,s)$ represents how excess liquidity in period t responds to r_{s+1}^K. Other derivatives follow the same convention.

An equilibrium in the liquid asset market is reached when excess liquidity is zero. Proposition 2 shows by how much returns on capital need to adjust to clear the liquid asset market in response to shifts in excess liquidity due to policies and output.

Proposition 2 *In equilibrium, returns on capital satisfy*

$$d\boldsymbol{r}^K = -\boldsymbol{\epsilon}_{r^K}^{-1} \underbrace{\left(d\boldsymbol{b}^G + \boldsymbol{\epsilon}_T d\boldsymbol{T} + \boldsymbol{\epsilon}_{r^B} d\boldsymbol{r}^B + \boldsymbol{\epsilon}_y d\boldsymbol{y} \right)}_{\textit{shifts in excess liquidity}}. \tag{6}$$

Moreover, if $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B} \to \infty$ *with* $\bar{\Theta}_{r^B}/\bar{\Theta}_{r^K} \to \varsigma$*, then* $d\boldsymbol{r}^K \to \varsigma d\boldsymbol{r}^B$.

Proof. See Appendix A.4. □

To understand the result, consider a special case where households' liquidity demand $\mathcal{B}_t$ is perfectly inelastic with respect to returns. In this case, $\boldsymbol{\epsilon}_{r^K}$ is determined by the financial sector's cross-price elasticities of liquidity supply (Proposition 1). Furthermore, suppose the financial sector features a constant net worth $n_t = m$ and no forward-looking component, $\gamma = 0$. In this case, Proposition 1 implies

$$\boldsymbol{\epsilon}_{r^K}^{-1} = (m\bar{\Theta}_{r^K})^{-1}\boldsymbol{I},$$

and Proposition 2 implies a positive shift in excess liquidity leads to a decrease in dr_{t+1}^K proportional to $\bar{\Theta}_{r^K}^{-1}$. For example, consider a shift in excess liquidity, say, due to an increase in b_t^G. Since household liquidity demand is fixed, the financial sector will need to absorb the increase in liquidity by holding more government debt and decreasing its net liquidity supply. This requires a decrease in expected returns dr_{t+1}^K, accompanied by an increase in capital price. This mechanism is stronger when the leverage sensitivity $\bar{\Theta}_{r^K}$ is low (liquidity supply inelastic), and capital and liquid assets are less substitutable for the financial sector.

Two polar assumptions about the financial sector's liquidity supply provide important benchmarks. On the one hand, when liquidity supply is perfectly elastic ($\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B} \to \infty$), assets are perfect substitutes for the financial sector. The perfect link between asset markets allows the government to fully control returns on capital r^K_{t+1} through monetary policy r^B_{t+1}. Changes in government debt, tax, and output have no effect on r^K_{t+1}, and households' liquidity demand $\mathcal{B}_t$ plays no role in determining the equilibrium outcome. On the other hand, when liquidity supply is perfectly inelastic, $\bar{\Theta}_{r^K} = \bar{\Theta}_{r^B} = G = 0$, asset market responses are determined entirely by households' liquidity demand $\mathcal{B}_t$. These limiting cases are important benchmarks because they are common assumptions in models studying fiscal and monetary policies. As we discuss in Appendix D.3, the perfectly elastic benchmark corresponds closely to the assumption in Auclert et al. (2023), and the perfectly inelastic benchmark corresponds to Kaplan et al. (2018). These assumptions about the liquidity supply lead to drastically different policy implications, as we show in Section 6.

Two polar cases of households' liquidity demand also provide a useful contrast. If $\mathcal{B}_t$ is perfectly elastic with respect to returns, features of the financial sector have no effects on aggregate outcomes. On the contrary, if $\mathcal{B}_t$ is perfectly inelastic, asset market responses are determined entirely by features of the financial sector. From these special cases, we see that asset market responses must be determined by the joint properties of households' liquidity demand and the financial sector's liquidity supply. However, as we show in Section 5, when households' consumption-saving behaviors are calibrated to match standard moments in the microdata, their liquidity demand is orders of magnitude less elastic than our measures of the financial sector's liquidity supply. As a result, quantitatively, asset market responses in our model will be mostly determined by features of the financial sector, represented by their cross-price elasticities we derived in Proposition 1.

Aggregate Output Responses

We combine asset market responses with the goods market clearing condition to solve for output responses. To understand the goods market, we define *aggregate demand*, Ψ_t, as the total of consumption, investment, and government spending:

$$\Psi_t(\boldsymbol{y}, \boldsymbol{r}^K, \boldsymbol{r}^B, \boldsymbol{T}, \boldsymbol{g}) \coloneqq \mathcal{C}_t(\boldsymbol{y}, r^A(\boldsymbol{y}, \boldsymbol{r}^K, \boldsymbol{r}^B), \boldsymbol{r}^B, \boldsymbol{T}) + \mathcal{X}_t(\boldsymbol{y}, \boldsymbol{r}^K) + g_t,$$

where $\boldsymbol{\Psi}_{r^K}$ is a matrix of derivatives and $\boldsymbol{\Psi}_{r^K}(t,s)$ represents how aggregate demand in period t responds to r^K_{s+1}. Other derivatives are defined similarly.

Equilibrium in the goods market requires aggregate output to equal aggregate demand. By totally differentiating the aggregate functions in the goods market clearing condition and using the expression for $d\boldsymbol{r}^K$ from Proposition 2, we obtain the following expression for output:

Theorem 1 *Given $\{d\boldsymbol{g}, d\boldsymbol{T}, d\boldsymbol{r}^B, d\boldsymbol{b}^G\}$, the output response is given by:*

$$d\boldsymbol{y} = \underbrace{\left(\mathbf{I} - \boldsymbol{\Psi}_y - \boldsymbol{\Omega}\boldsymbol{\epsilon}_y\right)^{-1}}_{\text{(3) modified Keynesian cross}} \Big(\underbrace{d\boldsymbol{g} + \boldsymbol{\Psi}_T d\boldsymbol{T} + \boldsymbol{\Psi}_{r^B} d\boldsymbol{r}^B}_{\text{(1) goods market channel}} + \underbrace{\boldsymbol{\Omega}\left(d\boldsymbol{b}^G + \boldsymbol{\epsilon}_T d\boldsymbol{T} + \boldsymbol{\epsilon}_{r^B} d\boldsymbol{r}^{\boldsymbol{B}}\right)}_{\text{(2) asset market channel}} \Big),$$

where $\boldsymbol{\Omega} := \boldsymbol{\Psi}_{r^K}(-\boldsymbol{\epsilon}_{r^K}^{-1})$.[6]

Proof. See Appendix A.5. □

Government policies affect output through three channels. (1) The goods market channel shows how government purchases, tax, and liquid rate directly affect aggregate demand in the goods market, capturing the standard Keynesian logic. (2) The asset market channel describes how government debt, tax, and liquid rate affect aggregate demand through asset markets. (3) A modified Keynesian cross captures the feedback between aggregate demand and income.

The transmission through asset markets depends on an *asset market propagation matrix* $\boldsymbol{\Omega}$, which represents how shifts in excess liquidity affect aggregate demand. It consists of two components. First, matrix $-\boldsymbol{\epsilon}_{r^K}^{-1}$ describes by how much shifts in excess liquidity due to policies affect returns on capital $\boldsymbol{r}^K$, representing the asset market responses in Proposition 2. Second, matrix $\boldsymbol{\Psi}_{r^K}$ describes how changes in $\boldsymbol{r}^K$ affect aggregate demand. For example, a negative entry $\boldsymbol{\Psi}_{r^K}(t,s)$ with $s > t$ reflects how a decrease in expected returns r^K_{s+1} leads to an increase in consumption and investment at time t (e.g., through an increase in current capital price).

Asset market propagation matrix $\boldsymbol{\Omega}$ is shaped by features of the financial sector: Intermediation frictions represented by $\bar{\Theta}_{r^K}$ determine the cross-price elasticities of the financial sector's liquidity supply, $\mathcal{D}_t$. Liquidity supply affects the response of

[6] We use $(\mathbf{I}-\boldsymbol{\Psi}_y-\boldsymbol{\Omega}\boldsymbol{\epsilon}_y)^{-1}$ to denote a generalized inverse matrix $\mathcal{M}$ such that $\mathcal{M}(\mathbf{I}-\boldsymbol{\Psi}_y-\boldsymbol{\Omega}\boldsymbol{\epsilon}_y) = \mathbf{I}$, see Auclert et al. (2023) for details.

excess liquidity through $\boldsymbol{\epsilon}_{r^K}$, and eventually shapes the asset market propagation through $\boldsymbol{\Omega}$.

Besides governing the asset market channel, the asset market propagation matrix also modifies the feedback between aggregate income and demand through the Keynesian cross. An increase in aggregate income not only affects aggregate demand directly but also shifts excess liquidity through $\boldsymbol{\epsilon}_y$. For example, when households receive higher income, they can increase their liquidity demand, reduce excess liquidity, and lower aggregate demand through asset market responses. Therefore, the same force that increases aggregate demand through the asset market channel can dampen the Keynesian cross feedback.

5 Taking the Model to the Data

We now take the model to the data to prepare for a quantitative assessment of how the financial sector affects aggregate responses to policies. We estimate parameters governing the intermediation frictions with bank balance sheet data and yield curves on Treasury and corporate bonds. We combine the estimates with a calibrated net worth process to measure the liquidity supply elasticities, while performing extensive checks in the appendix to argue for the robustness of our quantitative result. Finally, we calibrate the rest of the model and discuss its quantitative implications on household liquidity demand.

5.1 Empirical Summary of Intermediation Frictions

To provide an empirical summary of the intermediation frictions, we estimate key parameters of financial intermediation $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}, \gamma$ with identified structural shocks.

We assume the following empirical counterpart of the mapping between the intermediary's leverage and expected returns in Lemma 1:

$$d\Theta_t = \sum_{h=1}^{\infty} \gamma^{h-1} \Big(\bar{\Theta}_{r^K} \mathbb{E}_t[dr^K_{t+h}] - \bar{\Theta}_{r^B} \mathbb{E}_t[dr^B_{t+h}] \Big) + \upsilon_t, \tag{7}$$

where υ_t represents either measurement errors or exogenous shocks to the level of intermediation. We provide an overview of how we measure $d\Theta_t$, $\mathbb{E}_t[dr^K_{t+h}]$, and

$\mathbb{E}_t[dr^B_{t+h}]$ for the U.S. banking sector below and describe details in Appendix C.1.

Leverage: We take the market value of equity and the liquid asset positions of bank-holding companies come from the CRSP and Call Report data. For each quarter, we aggregate the market value of equity for n_t and the net liquid asset supply (liquid liabilities minus liquid assets) for d_t. Liquid liabilities and assets on the banking sector balance sheet include deposits in checkable, time, savings accounts, money market fund shares, and government liabilities such as cash, reserve, and Treasury debt. The effective leverage is calculated as

$$\text{effective leverage } (\Theta_t) \coloneqq 1 + \frac{\text{net supply of liquid assets } (d_t)}{\text{market value of bank equity } (n_t)}.$$

Returns: We use the yield curves on U.S. Treasury bonds to construct liquid rates over different horizons, $\mathbb{E}_t[dr^B_{t+h}]$. Since the banking sector's liquidity supply is given by their issuance of liquid assets net of government debt, these yields captured the relevant margin of adjustment for liquidity supply. For expected returns on capital, $\mathbb{E}_t[dr^K_{t+h}]$, we rely on the yield curves of high-quality market corporate bonds (grade A and above). To better represent returns on the banking sector's asset holdings, we adjust the yield curve proportionally so that their fluctuations are similar to Moody's BAA bond yield index of the corresponding horizon. Nominal yields are converted to real yields using inflation expectations data from the Cleveland Fed.

Threats to Identification

The main threat to identification is that the residual υ_t in Equation 7 may contain exogenous "leverage shocks" that affect the banking sector's ability to sustain their leverage, given the expected returns. For example, such shocks can represent changes in macroprudential policies or financial intermediaries' business strategies. These changes affect how much assets the financial sector can intermediate per unit of net worth. Exogenous changes to the idiosyncratic risk profile in Bernanke et al. (1999) studied in Christiano et al. (2014) would also appear in the residual υ_t. Because these shocks to the level of intermediation affect expected returns in general equilibrium, they introduce an omitted variable bias. As a result, our estimation requires instrumental variables for identification with the presence of these shocks.

Identification Strategy

A valid instrument I_t needs to satisfy the exclusion restriction: $\mathbb{E}[\upsilon_t \times I_t] = 0$, and the relevance condition: $\mathbb{E}\big[I_t \times \mathbb{E}_t[dr^K_{t+h}]\big],\ \mathbb{E}\big[I_t \times \mathbb{E}_t[dr^B_{t+h}]\big] \neq 0$. Intuitively, we need the instrument to move with expected returns, but to be uncorrelated with exogenous changes in the banking sector's ability to maintain its leverage under a certain level of expected returns.

We use three shock proxies to construct such instruments: (i) the high-frequency monetary policy shock proxies constructed by Bauer and Swanson (2023), (ii) the oil supply shock proxies constructed by Baumeister and Hamilton (2019), and (iii) the high-frequency intermediaries net worth shock proxies constructed by Ottonello and Song (2022). We argue that these proxies are unlikely to be driven by a change in macroprudential policy, a change in the banking sector's business strategy, or shocks to the banking sector as in Christiano et al. (2014): These changes are unlikely to happen exactly during the short windows around the FOMC announcement or the release of earnings reports, and they are not likely to comove with the oil supply.

With these shock proxies, we construct returns variations $\mathbb{E}_t[d\check{r}^K_{t+h}]$ and $\mathbb{E}_t[d\check{r}^B_{t+h}]$ as our instrument from an SVAR model: We first recover structural shocks corresponding to the three proxies by assuming that only these shocks can affect the proxies contemporaneously. We then extract variations in forward rates that are attributed to these shocks. To the extent that the proxies are not affected contemporaneously by events that change the banking sector's leverage given expected returns, return variations $\mathbb{E}_t[d\check{r}^K_{t+h}]$ and $\mathbb{E}_t[d\check{r}^B_{t+h}]$ satisfy the exclusion restriction. To check for the relevance condition, we show that return variations generated by the three structural shocks account for approximately 20% of the total variations in expected returns, as reported in Appendix C.1.

Estimation Results

We estimate parameters $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}, \gamma$ using the generalized method of moments with moment conditions of the form: $\mathbb{E}[\upsilon_t \times I_t] = 0$. We consider two alternative specifications of I_t. First, as a baseline case, we assume that υ_t consists purely of measurement errors and use $I_t \in \big\{\mathbb{E}_t[dr^K_{t+h}],\ \mathbb{E}_t[dr^B_{t+h}]\big\}$ in our estimation. If we had a linear model, this would correspond to ordinary least squares regression. Second, to address the threat to identification that υ_t contains exogenous shocks to leverage, we use variation in expected returns generated by the identified shocks: $I_t \in \left\{\mathbb{E}_t[\check{dr}^K_{t+h}],\ \mathbb{E}_t[\check{dr}^B_{t+h}]\right\}$. We use forward rates for horizons 1, 5, 10, and 30 years in both specifications. The estimation results are reported in Table 1, and details are provided in Appendix C.1.

Table 1: Estimation of Intermediation Frictions

	(1) baseline	(2) IV
size of cross-price, $\bar{\Theta}_{r^K}$	27.58	23.73
	(13.49)	(16.31)
size of own-price, $\bar{\Theta}_{r^B}$	22.51	25.78
	(16.73)	(21.46)
forward-looking, γ	0.94	0.94
	(0.03)	(0.06)
Observations	252	252
J-test (p-value)	10.09 (0.12)	7.37 (0.28)

Note: Estimation uses iterative GMM for optimal weighting matrix. Standard errors use heteroskedastic and autocorrelation consistent estimators. Sample period: January 1999 to December 2019, monthly observation.

The first column of Table 1 presents estimates from our baseline specification and that with instrumental variables. Estimates of $\bar{\Theta}_{r^K}$ and $\bar{\Theta}_{r^B}$ imply the effective leverage of the banking sector increases by around 25 percentage points in response to one percentage point increase in the spread between the two returns. The forward-looking component γ is converted to quarterly frequency since we use a quarterly frequency calibration for our quantitative results. Our estimate for γ is around 0.94, which implies a "half-life" of around two and a half years: the response to a spread increase two and a half years ahead is half as strong as the response to a change next quarter. The total response of banks' effective leverage is a discounted sum of responses to all future spreads.

In Appendix C.1, we carry out two sets of robustness checks. First, to alleviate concerns that any specific proxy may contain information about υ_t and violate the exclusion restriction, we exclude each of the three proxies and construct return variations with the other two proxies. The results are largely similar to those in Table 1.

Second, we generalize our empirical specification to measure the extent to which our estimates vary with the aggregate state of the economy. This provides us with useful information to gauge the situation under which our result is useful. We do not find statistically significant results in support of state dependency, as the standard errors for the state dependency parameter are large.

5.2 Implied Liquidity Supply Elasticities

We calculate the liquidity supply elasticities using our estimates for $\bar{\Theta}_{r^K}$, $\bar{\Theta}_{r^B}$, and γ and the formula in Proposition 1. To complete the calculation, we need information about steady-state returns and leverage and to specify the net worth process.

Steady-state returns and leverage: We set liquid rate r^B equal to 0%, consistent with the average one-year Treasury real yield in our sample. We target r^K at 3.5% per annum, corresponding to the average real yield on BAA bonds. The average effective leverage $\bar{\Theta}$ is 4.

Net worth process: We assume the intermediary net worth follows:

$$G\left(\Theta_{t-1}, r_t^K, r_t^B\right) = (1-f)\left[1 + r_t^B + \left(r_t^K - r_t^B\right)\Theta_{t-1}\right].$$

The net worth process features a constant payout rate f from the leveraged return. We set $f = 0.06$, a value that falls in the common range in the literature. In Appendix E.3, we allow for a fully flexible G function and demonstrate that our policy conclusion is insensitive to a wide range of values for $\bar{G}_\Theta$, $\bar{G}_{r^K}$, $\bar{G}_{r^B}$, and $\bar{G}$. As we show in Appendix B.3, these robustness checks imply that our policy conclusion is also robust to alternative net worth specifications that feature endogenous equity issuance as in Karadi and Nakov (2021) and Akinci and Queralto (2022).

Figure 1 shows the semi-elasticities, $\frac{\partial \mathcal{D}_t/\partial r_s^K}{\mathcal{D}_t}$ and $\frac{\partial \mathcal{D}_t/\partial r_s^B}{\mathcal{D}_t}$, implied by Proposition 1, using our estimates for $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}, \gamma$ and the calibrated net worth process. Each line represents how liquidity supply responds to an increase in returns in a different period s. The cross-price elasticities show that liquidity supply increases before period s, reflecting the forward-looking component. After period s, liquidity supply drops sharply but remains elevated due to propagation through net worth. The size of the initial response $\frac{\partial \mathcal{D}_0/\partial r_1^K}{\mathcal{D}_0}$ implies an 8.05% increase in liquidity supply in response to a one percentage point increase in r_1^K. The own-price elasticities have a similar pattern with the opposite sign.

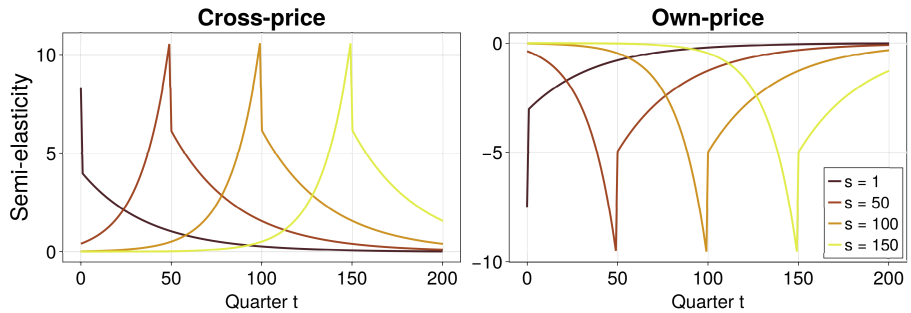

Figure 1: Semi-elasticities of liquidity supply. Each line corresponds to a different period s and shows semi-elasticity of liquidity supply in quarter t with respect to r_s^K and r_s^B.

Benchmarks: Common Assumptions in Macro Models

To put our measures of the liquidity supply elasticities into context, we compare them to three common assumptions in workhorse macroeconomic models that feature, respectively, (1) a perfectly inelastic liquidity supply, (2) a perfectly elastic liquidity supply, and (3) asset diversion frictions as in Gertler-Karadi-Kiyotaki.

The first two benchmarks correspond respectively to the assumptions in Kaplan et al. (2018) and Auclert et al. (2023), as we discussed in Section 4.2. These polar cases are useful benchmarks because they are currently some of the most popular models for quantitative analysis of monetary and fiscal policy.

The third benchmark (GKK and its variants) constitutes the majority of macroeconomic models with financial frictions. However, as discussed in Section 3.1, GKK imposes a tight restriction on the liquidity supply elasticities, linking them to the intermediary's steady-state returns and leverage:[7]

$$\bar{\Theta}_{r^K} = \frac{\bar{\Theta}(\bar{\Theta}-1)}{1+r^K}, \quad \bar{\Theta}_{r^B} = \frac{\bar{\Theta}(\bar{\Theta}-1)}{1+r^B}, \quad \gamma = \frac{(1-f)(1+r^B+\left(r^K-r^B\right)\bar{\Theta})^2}{(1+r^K)(1+r^B)}.$$

Given the steady-state returns and leverage in our sample, the values of these parameters are, respectively, 11.9, 12.0, and 0.998. The semi-elasticities implied by these parameters are half of those in our baseline, for example, $\frac{\partial \mathcal{D}_0/\partial r_1^K}{\mathcal{D}_0}$ equals 3.96.

[7]The formula depends on the intermediary's discount factor, which we use $1/(1+r_{t+1}^K)$ as a baseline. Formulas for other alternatives are provided in Appendix B.1.

The differences between our empirical measures and these benchmarks are important as they lead to diverging policy conclusions, as we show in Section 6.

5.3 Production, Government, and Households

Production: The elasticity of output with respect to capital α is set to 0.35. Depreciation rate δ is 5.58% yearly. Capital production function is $\Gamma(\iota_t) = \bar{\iota}_1 \iota_t^{1-\kappa_I} + \bar{\iota}_2$, where $\bar{\iota}_1, \bar{\iota}_2$ generate steady-state investment-to-capital ratio equal to δ and the price of capital equal to 1. We set $\kappa_I = 0.5$ so that the elasticity of investment to capital price is 2. Unions allocate labor uniformly among households: $l(z_{i,t}) \equiv 1/\int z_{i,t} di$. Since monetary policy targets real liquid rates, the slope of the wage Phillips curve does not matter for output responses. Therefore, the exact values of the elasticity of substitution between labor varieties, ε_W, and nominal wage rigidity, κ_W, are inconsequential.

Government: We set steady-state net tax revenue to 15% of output and the tax system's progressivity parameter, λ, to 0.18. Net liquid assets supplied by the government (and held by the private sector) is 21% of steady-state output, consistent with liquid asset positions in the data, as shown in Appendix C. We assume the government holds no illiquid assets in the steady state. The level of government purchases implied by the government budget constraint is 15% of output.

Households

Preferences: There are two types of households, indexed by s with population share μ_s. Period utility functions have the following form:

$$u_s(c) - \nu_s(h) = \frac{c^{1-\sigma_s} - 1}{1-\sigma_s} - \varsigma \frac{h^{1+\frac{1}{\varphi}}}{1+\frac{1}{\varphi}}, \quad \sigma_s \geq 0, \ \varphi \geq 0.$$

We set the intertemporal elasticity of substitution, $1/\sigma_s$, to $\{1/2, 2\}$ for $s \in \{1, 2\}$, following a simplified version of Aguiar et al. (2020). The Frisch labor supply elasticity, φ, is set to 1. Parameter ς is set so that steady-state average hours worked equal one-third.

Income process: We use a discrete-time version of the income process described in Kaplan et al. (2018), which targets eight moments of the male-earnings distribution from

Guvenen et al. (2015). Income process is the same for both household types.

Assets: Adjustment cost of illiquid assets is similar to Auclert et al. (2021):

$$\Phi_t(a_{i,t}, a_{i,t-1}, r_t^A) = \frac{\chi_1}{\chi_2} \left| \frac{a_{i,t} - (1 + r_t^A)a_{i,t-1}}{a_{i,t-1} + \chi_0} \right|^{\chi_2} [a_{i,t-1} + \chi_0].$$

We set χ_0 to 0.1 and assume asset positions cannot be negative: $\underline{a} = \underline{b} = 0$.

Parameters of households that we calibrate internally include the discount rates β_s of both types, the share of agents with high intertemporal elasticity of substitution μ_2, and two parameters of the adjustment cost function χ_1 and χ_2. We target five moments: the steady-state ratios of liquid and illiquid assets to GDP, the shares of wealthy (WHtM) and poor hand-to-mouth (PHtM) agents (25% and 15%), and the first quarter marginal propensity to consume out of $500 transfer (MPC) (20%). Table 1 shows that the model replicates target moments and reports calibrated parameter values.[8]

Table 2: Households Calibration

Target Moments	Model	Data	Parameter	Value
Liquid assets to GDP	0.60	0.55	β_1	0.983
Illiquid assets to GDP	3.36	3.43	β_2	0.943
Poor Hand-to-Mouth	15%	9 - 17%	μ_2	0.176
Wealthy Hand-to-Mouth	25%	12 - 33%	χ_1	23.34
First quarter MPC	20%	15 - 25%	χ_2	2.0154

Data Source: See Appendix C.2 for liquid assets and illiquid assets positions; shares of HtM households: Table 3 in Kaplan et al. (2014); MPC: Kaplan and Violante (2022).

Our household sector features a canonical two-asset heterogeneous agent model calibrated to match standard targets. Households' consumption responses to an increase in disposable income (MPC) are large, as commonly emphasized in the literature. However, the same portfolio adjustment frictions that generate illiquidity and large consumption responses for WHtM households also imply that these households face difficulties in adjusting asset positions in response to returns. As a result, with the portfolio adjustment cost calibrated to match empirically plausible MPC, households' consumption responses also inform us about their liquidity demand with respect to returns.

[8] Households' illiquid assets positions equal gross illiquid assets minus liabilities. For example, higher mortgage lending from banks to households will lower households' illiquid asset position and increase intermediary capital holdings.

Implied Liquidity Demand

Figure 2 compares household liquidity demand from our calibration to our estimates of the financial sector's liquidity supply, along with the excess liquidity supply. Each line represents responses to an increase in r_s^K, taking into account its effect on the sequence of illiquid returns $\{r_s^A\}$.

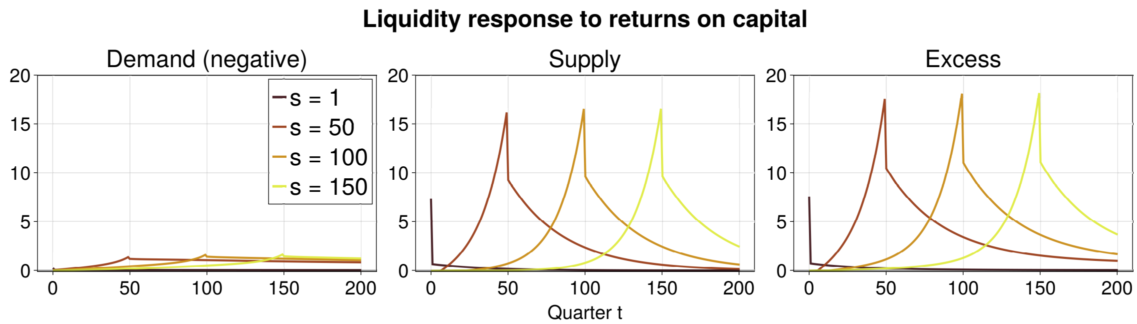

Figure 2: Entries of $-\tilde{\mathbf{B}}_{r^K}, \mathbf{D}_{r^K}$, and $\boldsymbol{\epsilon}_{r^K}$ matrices (see Appendix A.6 for their definitions). Each line corresponds to a different period s and shows a response of liquidity demand, liquidity supply, or excess liquidity in quarter t with respect to r_s^K

Liquidity demand responses are an order of magnitude smaller than liquidity supply responses. This implication is consistent with existing empirical evidence, such as Gabaix et al. (2024), which suggests that insensitivity and inertia in asset allocations are prominent features of the household sector, including even ultra-rich households. As discussed in Section 4.2, the contrast between the elasticities of liquidity demand and supply has an important implication: When liquidity demand is inelastic with respect to returns, asset market responses are mostly determined by the financial sector through its cross-price elasticities of liquidity supply. Since workhorse macro models feature a wide range of assumptions on these elasticities, they generate substantially different conclusions for a variety of policy questions.

6 Policy Implications

Our sufficient statistics impose empirical discipline on assumptions about the financial sector and have crucial policy implications. We use two policy questions to demonstrate its importance. The first question concerns the government spending multipliers: How do changes in government spending affect aggregate output? The second question is central to the Wall Street vs. Main Street debate: Can asset purchases stimulate aggregate output more effectively than tax cuts?

Through our sufficient statistics, we systematically compare policy implications of common assumptions about the financial sector, as discussed in Section 5.2, and contrast them with our empirical measures. We calculate aggregate responses to policies under these assumptions, keeping all else equal in our calibrated model.

6.1 Government Spending Multiplier

We show that the size of the government spending multiplier depends crucially on the implicit assumptions about the financial sector. Consider the government implementing a policy path $\{d\hat{b}^G_t, d\hat{g}_t, d\hat{r}^B_t, d\hat{a}^G_t, d\hat{T}_t\}$ with

$$d\hat{g}_t = \eta^t s_0, \quad d\hat{b}^G_t = \rho_{b^G}(d\hat{b}^G_{t-1} + d\hat{g}_t), \quad d\hat{r}^B_t = d\hat{a}^G_t = 0,$$

and tax revenue $d\hat{T}_t$ is set to satisfy the budget constraint. Parameter s_0 controls the size of spending, decaying at rate η; the extent of deficit financing is governed by ρ_{b^G}. We set $\eta = 0.5$ so that most spending is completed in one year; size s_0 is such that debt peaks at 1% of annual GDP. We use $\rho_{b^G} = 0.95$ as a baseline, which implies government debt peaks in one year and is reduced to half in five years. As we discuss below, the main message remains the same for different levels of deficit financing. Figure 3 shows the resulting policy paths.

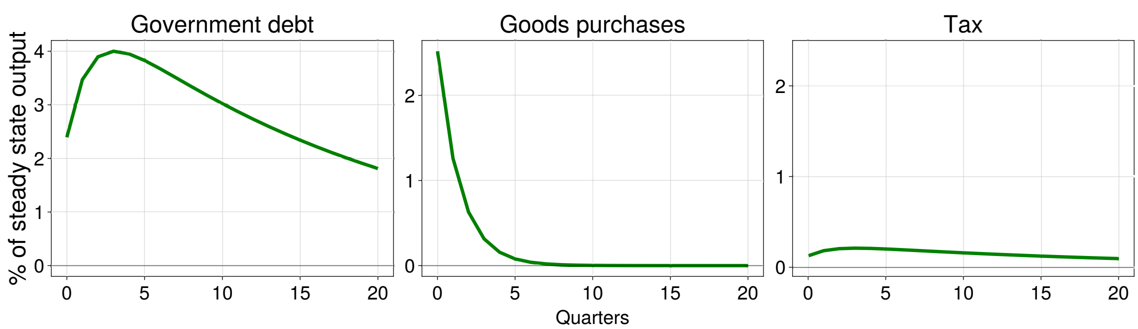

Figure 3: The path of government debt, good purchases, and taxes.

Output Responses

Figure 4 shows how output responses to government spending depend on assumptions about the financial sector. Each line represents a version of our model with a different specification of the financial sector's liquidity supply elasticities. The blue and black lines indicate responses with perfectly inelastic and elastic supply. The red line represents output responses with elasticities implied by our empirical estimates. Yellow shades from dark to light represent models with decreasing values for leverage sensitivity $\bar{\Theta}_{r^K}$ from our empirical estimate ($\bar{\Theta}_{r^K} = 25$) to that implied by a Gertler-Karadi-Kiyotaki type model ($\bar{\Theta}_{r^K} = 12$).

Output responses on impact differ substantially across different assumptions about the financial sector, ranging from 2.3% to 4.5% of steady-state output, with inelastic liquidity supply associated with stronger responses. The impact government spending multiplier ranges from 1.1 to 1.9, and the cumulative multiplier from 0.5 to 2.5.[9]

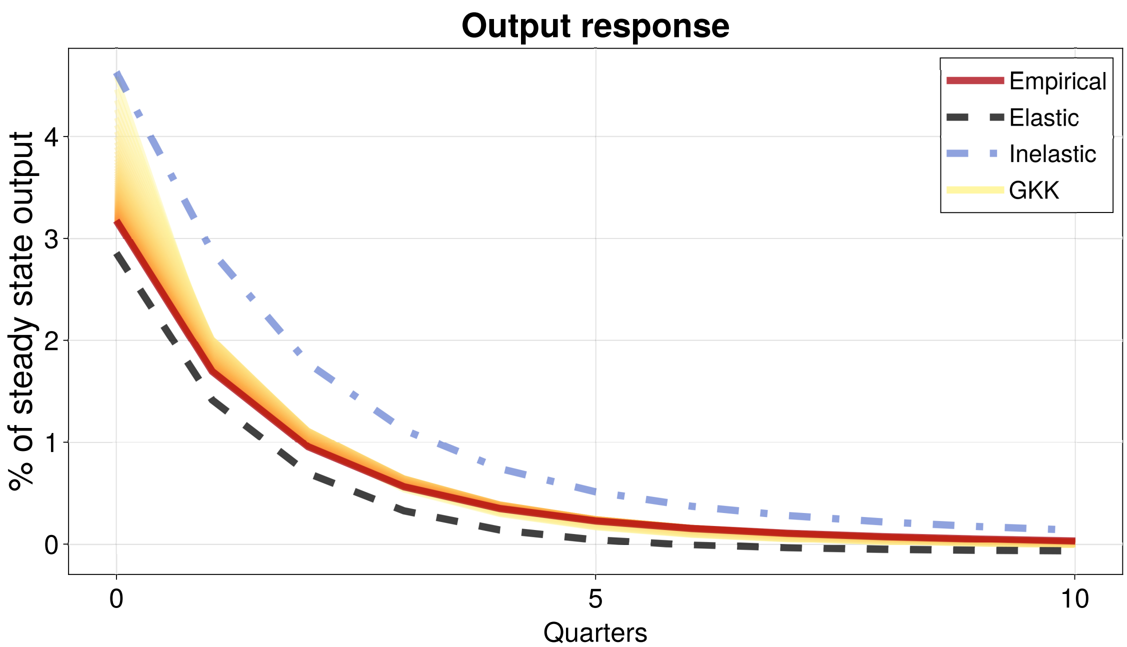

Figure 4: Output response to government goods purchases with $\eta = 0.5, \rho_{b^G} = 0.95$. Red: empirical estimate ($\bar{\Theta}_{r^K} = 25$). Black: perfectly elastic liquidity supply ($\bar{\Theta}_{r^K} \to \infty$). Blue: perfectly inelastic liquidity supply ($\mathcal{D}_{r^K} = 0$). Dark to light yellow: high to low $\bar{\Theta}_{r^K}$ (from $\bar{\Theta}_{r^K} = 25$ to $\bar{\Theta}_{r^K} = 12$ in GKK). Figure 14 in Appendix E.2 shows the responses of consumption and investment.

[9]The cumulative multiplier is calculated as $\sum_{t=0}^{\infty}(1 + r^B)^{-t} dy_t / \sum_{t=0}^{\infty}(1 + r^B)^{-t} dg_t$. If we use r^K instead, the corresponding numbers are 0.7 and 2.5

Decomposition

To understand why assumptions about the financial sector lead to widely different outcomes, we use Theorem 1 to decompose output responses into the goods market channel, the asset market channel, and the modified Keynesian cross, as shown in Figure 5.

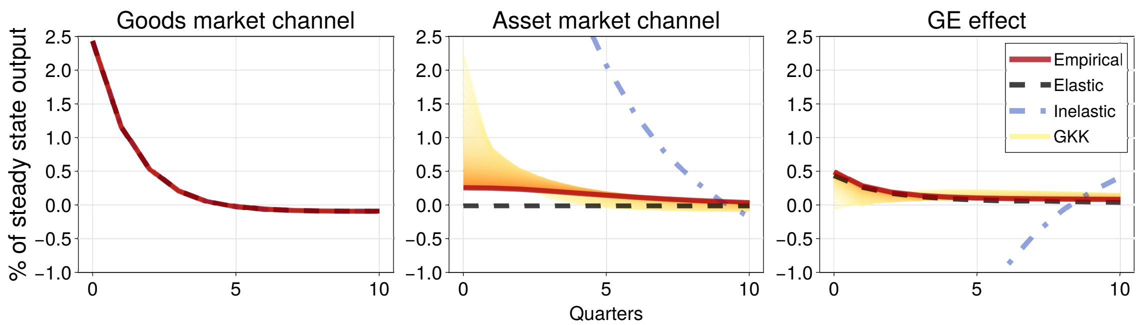

Figure 5: Decomposition of output responses to government goods purchases with $\eta = 0.5$, $\rho_{bG} = 0.95$ using Theorem 1. GE effect: the difference between total output responses and the sum of the goods market and asset market channels. Red: empirical estimate ($\bar{\Theta}_{r^K} = 25$). Black: perfectly elastic liquidity supply ($\bar{\Theta}_{r^K} \to \infty$). Blue: perfectly inelastic liquidity supply ($\mathcal{D}_{r^K} = 0$). Dark to light yellow: high to low $\bar{\Theta}_{r^K}$ (from $\bar{\Theta}_{r^K} = 25$ to $\bar{\Theta}_{r^K} = 12$ in GKK). Figure 15 in Appendix E.2 shows a version with the full range of responses.

The goods market channel reflects direct responses of aggregate demand to $d\boldsymbol{g}$ and $d\boldsymbol{T}$. Since these responses do not depend on the financial sector, all specifications generate the same outcome.

By contrast, the same policy generates significant differences in the asset markets due to different assumptions about the financial sector, as shown in the middle panel. As government debt and taxes shift excess liquidity, variation in the financial sector's liquidity supply elasticities generates a wide range of responses through the asset market propagation matrix $\boldsymbol{\Omega}$. By changing $\bar{\Theta}_{r^K}$ from our empirical estimate to that imposed by a GKK model, aggregate demand rises by an order of magnitude, and the contrast is even more drastic between the perfectly elastic and inelastic benchmarks.

Finally, we present the modified Keynesian cross as the general equilibrium (GE) effect, which is the difference between total output responses and the sum of the first two channels. A prominent aspect of this channel is a strong dampening response when liquidity supply is perfectly inelastic (blue line): While shifts in excess liquidity generate strong aggregate demand through the asset market channel, increases in aggregate income shift up liquidity demand, reduce excess liquidity, and dampen the total output responses. The dampening force reduces the output response on impact by up to 50%. The same dampening force is also present in other specifications, but it is dominated by the standard Keynesian feedback when liquidity supply elasticities are high.

The Size of Multipliers

While Figure 4 shows output responses to government spending with a specific path of government debt and taxes, Figure 6 shows that our main message holds for a wide range of debt financing schemes, as parameterized by ρ_{b^G}.

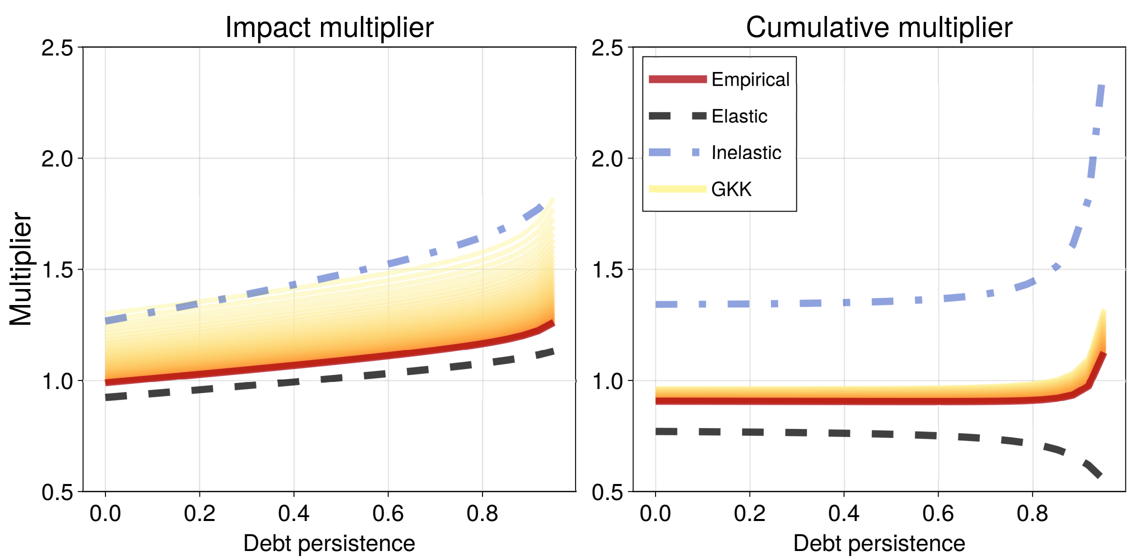

Figure 6: Impact and cumulative government spending multipliers for $\eta = 0.5$ and $\rho_{b^G} \in [0, 0.95]$. Impact multiplier: dy_0/dg_0. Cumulative multiplier: $\sum_{t=0}^{\infty}(1+r^B)^{-t}dy_t / \sum_{t=0}^{\infty}(1+r^B)^{-t}dg_t$. Red: empirical estimate ($\bar{\Theta}_{r^K} = 25$). Black: perfectly elastic liquidity supply ($\bar{\Theta}_{r^K} \to \infty$). Blue: perfectly inelastic liquidity supply ($\mathcal{D}_{r^K} = 0$). Dark to light yellow: high to low $\bar{\Theta}_{r^K}$ (from $\bar{\Theta}_{r^K} = 25$ to $\bar{\Theta}_{r^K} = 12$ in GKK).

Across all degrees of debt financing, the government spending multiplier is always significantly larger when the financial sector features an inelastic liquidity supply (blue line) than an elastic supply (black line). In both cases, our model converges to canonical HANK frameworks that are commonly used for analyzing fiscal and monetary policies, as we discussed in Section 4.2. The only difference between the two is their assumptions about the financial sector. Yet, they generate government spending impact multipliers that vary by almost a factor of two as shifts in excess liquidity generate different asset market responses. The differences are even larger for cumulative multipliers. Even in the case without debt financing ($\rho_{b^G} = 0$), assumptions about the financial sector matter as taxes imposed by the government shift households' liquidity demand. With a greater degree of debt financing, assumptions about the financial sector lead to larger variations in the multipliers, as the issuance of government debt generates larger shifts in excess liquidity. Figure 16 in Appendix E.1 shows a similar when we fix $\rho_{b^G} = 0$ and vary the persistence of spending, η, instead.

6.2 Asset Purchases vs. Tax Cuts

Our second policy question concerns the Wall Street vs. Main Street debate: Can asset purchases stimulate aggregate output more effectively than tax cuts? We compare two alternative policies for the government.

The first policy, given by $\{d\tilde{b}^G_t, d\tilde{g}_t, d\tilde{r}^B_t, d\tilde{a}^G_t, d\tilde{T}_t\}$, features an asset purchase program in which the government issues liquid debt $d\tilde{b}^G_t$ and holds illiquid asset $d\tilde{a}^G_t$ of the same value: $d\tilde{a}^G_t = d\tilde{b}^G_t$. The government keeps $d\tilde{g}_t = d\tilde{r}^B_t = 0$, and adjusts tax revenue $d\tilde{T}_t$ to satisfy its budget constraint. This policy implies net asset purchases:[10]

$$d\Delta_t \coloneqq d\tilde{a}^G_t - (1 + r^A) d\tilde{a}^G_{t-1}.$$

For the second policy, given by $\{d\check{b}^G_t, d\check{g}_t, d\check{r}^B_t, d\check{a}^G_t, d\check{T}_t\}$, the government keeps illiquid asset holdings at $d\check{a}^G_t = 0$ and pays out $d\Delta_t$ as tax cuts: $d\check{T}_t \coloneqq d\tilde{T}_t - d\Delta_t$. It maintains the same path for debt, $d\check{b}^G_t = d\tilde{b}^G_t$, and $d\check{g}_t = d\check{r}^B_t = 0$.

To parameterize the policy paths, we assume the government debt follows

$$d\tilde{b}^G_t = \rho_{b^G}(d\tilde{b}^G_{t-1} + s_t), \quad s_t = \eta^t s_0.$$

As in Section 6.1, we set $\rho_{b^G} = 0.95$ and $\eta = 0.5$, so that government debt peaks at 1% of annual GDP in one year, and we calculate the implied paths for asset purchases and tax cuts. Figure 7 shows the paths for the two policies for government debt, net asset purchases ($d\Delta_t$ and 0), and tax revenue ($d\tilde{T}_t$ and $d\check{T}_t$).

[10]Since $a^G = 0$ in the steady state and net asset purchases are given by $\Delta_t \coloneqq \tilde{a}^G_t - (1 + r^A_t)\tilde{a}^G_{t-1}$, and the formula for $d\Delta_t$ follows from a first-order approximation with $dr^A_t a^G = 0$.

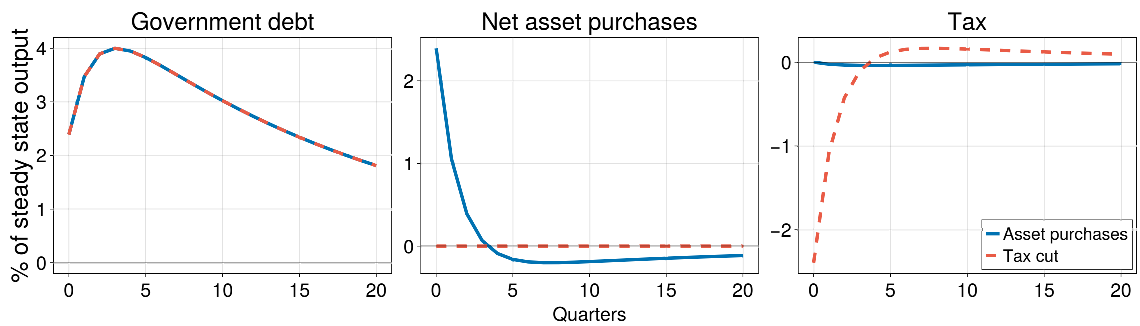

Figure 7: The path of government debt, net asset purchases, and taxes.

Output Responses

Figure 8 compares how different assumptions about the financial sector's liquidity supply elasticities affect output responses to asset purchases and tax cuts.

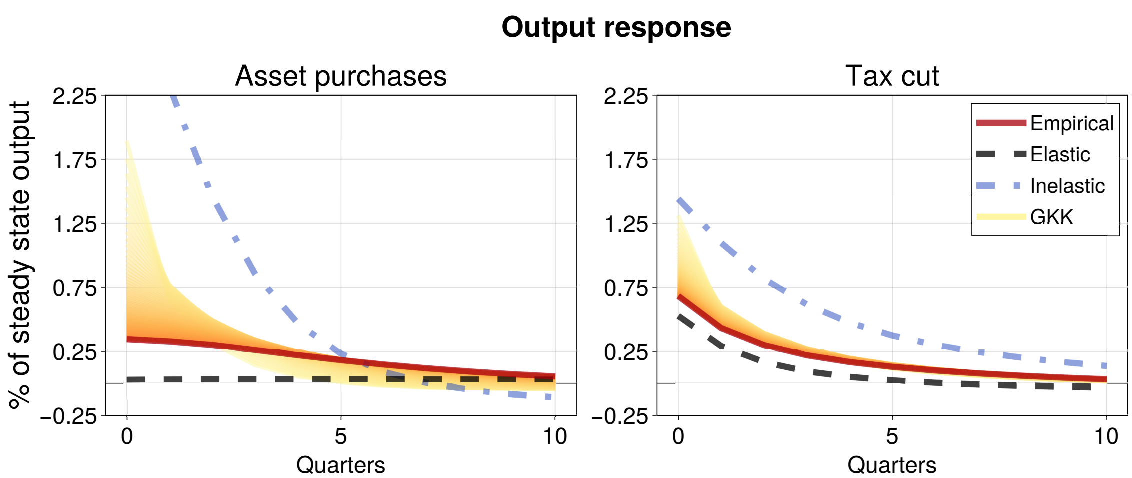

Figure 8: Output responses to asset purchases and tax cuts with $\eta = 0.5, \rho_{b^G} = 0.95$. Red: empirical estimate ($\bar{\Theta}_{r^K} = 25$). Black: perfectly elastic liquidity supply ($\bar{\Theta}_{r^K} \to \infty$). Blue: perfectly inelastic liquidity supply ($\mathcal{D}_{r^K} = 0$). Dark to light yellow: high to low $\bar{\Theta}_{r^K}$ (from $\bar{\Theta}_{r^K} = 25$ to $\bar{\Theta}_{r^K} = 12$ in GKK). Appendix E.2 shows the full range of output responses in Figure 19 and the responses of consumption and investment in Figures 17 and 18.

For asset purchases, output responses on impact differ by orders of magnitudes, ranging from 0.03% to more than 3.2% between perfectly elastic and inelastic liquidity supply; for tax cuts, output responses differ almost by a factor of three, ranging from 0.5% to 1.4%. Despite the difference being substantial for both policies, the effects of asset purchases are noticeably more sensitive to assumptions about the financial sector than tax cuts. To understand the contrast between the two policies, we decompose output responses into the three channels using Theorem 1.

Decomposition

Figure 9 shows the decomposition for asset purchases and tax cuts, with each column representing a channel. The left column shows the goods market channel. Asset purchases have little effect through this channel as their impact on tax revenue is limited. Tax cuts, on the contrary, generate substantial responses due to households' high MPCs. The effect of tax cuts emphasizes the importance of household heterogeneity and illiquidity. These features break Ricardian equivalence and generate substantial effects from deficit-financed tax cuts. However, since households' MPCs do not depend on the financial sector, all specifications generate the same outcome through the goods market channel.

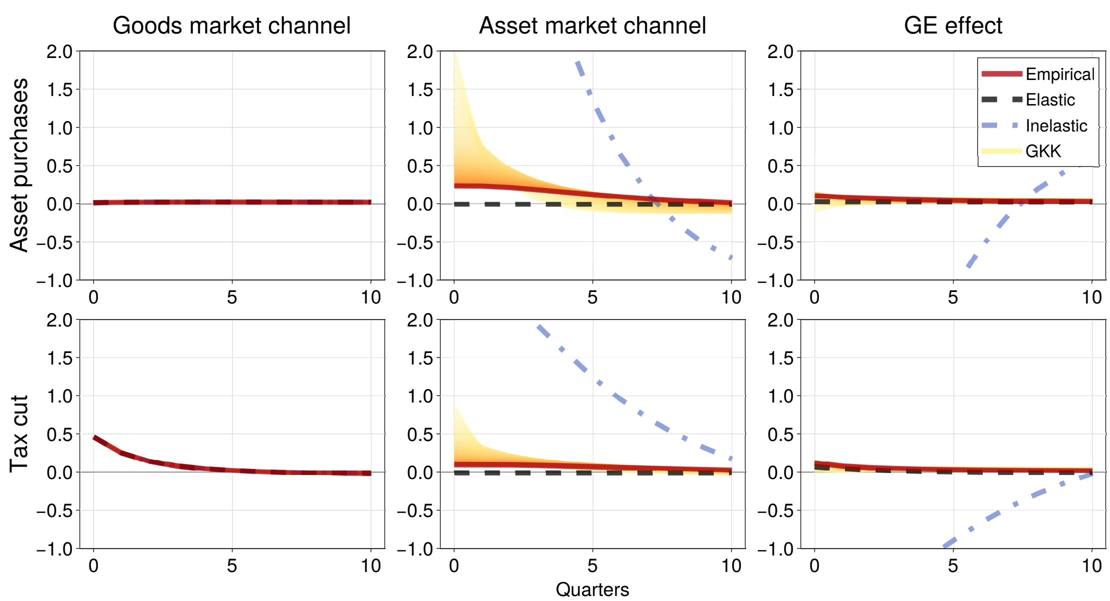

Figure 9: Decomposition of output responses using Theorem 1. The first row shows decomposition for asset purchases; the second row shows decomposition for tax cuts, both policies with $\eta = 0.5, \rho_{bG} = 0.95$. GE effect: the difference between total output responses and the sum of the goods market and asset market channels. Red: empirical estimate ($\bar{\Theta}_{r^K} = 25$). Black: perfectly elastic liquidity supply ($\bar{\Theta}_{r^K} \to \infty$). Blue: perfectly inelastic liquidity supply ($\mathcal{D}_{r^K} = 0$). Dark to light yellow: high to low $\bar{\Theta}_{r^K}$ (from $\bar{\Theta}_{r^K} = 25$ to $\bar{\Theta}_{r^K} = 12$ in GKK). Figure 20 in Appendix E.2 shows a version with the full range of responses.

By contrast, through the asset market channel, assumptions about the financial sector generate a wide range of outcomes in response to shifts in excess liquidity generated by the two policies. Comparing the two policies, the asset purchase program generates stronger effects through asset markets as it generates larger shifts in excess liquidity. However, as it relies more on the propagation through asset markets, it is more sensitive to the assumption about the financial sector than tax cuts, explaining the wider range of output responses for asset purchases in Figure 8.

Relative Effects of Asset Purchases vs. Tax Cuts

To study how our conclusions about the effectiveness of the two policies depend on assumptions about the financial sector, we calculate the difference in output responses to asset purchases and tax cuts:

$$d\boldsymbol{y}^{asset} - d\boldsymbol{y}^{tax} = \underbrace{\left(\mathbf{I} - \boldsymbol{\Psi}_y - \boldsymbol{\Omega}\boldsymbol{\epsilon}_y\right)^{-1}}_{\text{(3) modified Keynesian cross}} \Big(\underbrace{\boldsymbol{\Psi}_T d\boldsymbol{\Delta}}_{\text{(1) goods market}} + \underbrace{\boldsymbol{\Omega}\ \boldsymbol{\epsilon}_T d\boldsymbol{\Delta}}_{\text{(2) asset market}} \Big).$$

The difference is positive when asset purchases have a stronger effect on output and vice versa. As the difference depends only on $d\boldsymbol{\Delta}$, our conclusion about the relative effects does not hinge on our assumption about $d\boldsymbol{g}, d\boldsymbol{r}^B$, as long as both policies feature the same paths. Therefore, we effectively control for responses due to other policy variables by focusing on the difference.

Figure 10 represents the difference in output responses. On one end, when the financial sector features perfectly inelastic liquidity supply, asset purchases have a stronger effect than tax cuts: the difference in output response amounts to 1.8% of steady-state output on impact. A Gertler-Karadi-Kiyotaki type model gives a qualitatively similar prediction: Asset purchases are more effective in stimulating output, with a 0.6% difference. On the other end, when the financial sector features a perfectly elastic liquidity supply, the asset market channel vanishes, the effect of tax cuts dominates through the goods market channel, and the difference is -0.5% on impact. Compared to this benchmark, our empirical estimate generates a non-negligible response to asset purchases. Still, our estimates of these elasticities are relatively high, which implies modest asset market responses compared to the goods market channel. As a result, they predict that policies targeting households directly, such as tax cuts, can stimulate output more effectively than policies that rely mostly on the asset market channel, such as asset purchases.

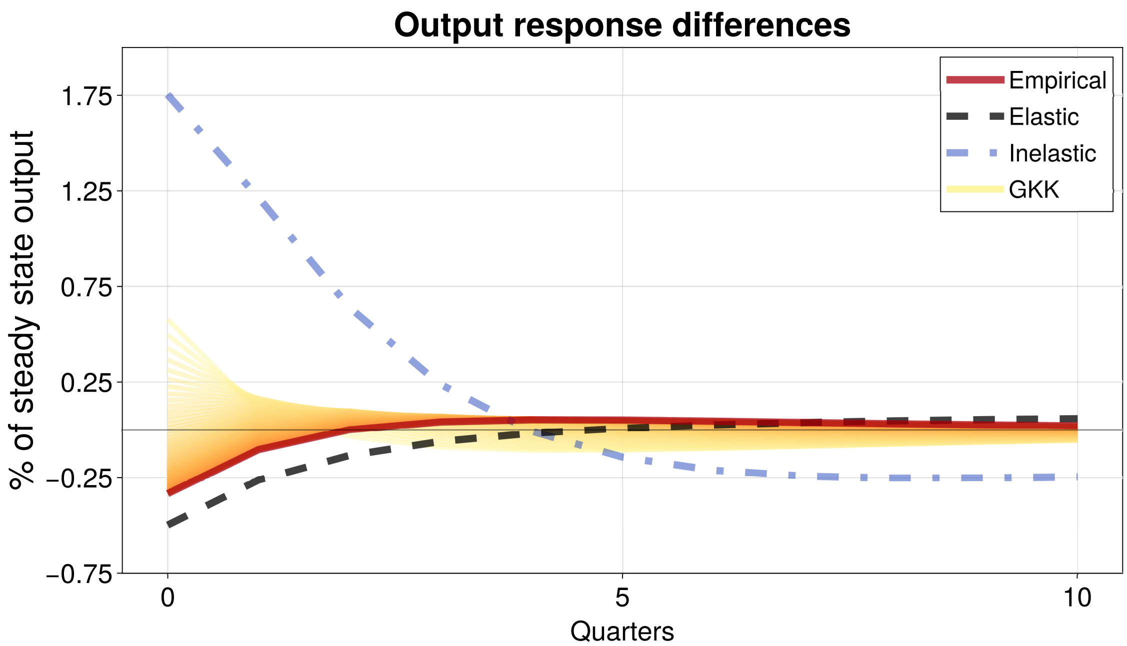

Figure 10: Difference between output response to asset purchases and tax cuts. Positive values mean that responses to asset purchases are larger. Red: empirical estimate ($\bar{\Theta}_{r^K} = 25$). Black: perfectly elastic liquidity supply ($\bar{\Theta}_{r^K} \to \infty$). Blue: perfectly inelastic liquidity supply ($\mathcal{D}_{r^K} = 0$). Dark to light yellow: high to low $\bar{\Theta}_{r^K}$ (from $\bar{\Theta}_{r^K} = 25$ to $\bar{\Theta}_{r^K} = 12$ in GKK).

7 Conclusion

In this paper, we show that, for a large class of macro models with financial frictions, the financial sector's asset supply elasticities provide sufficient information about the underlying financial frictions. By focusing on these elasticities, we can strengthen empirical discipline on these frictions. Such discipline is highly policy-relevant, as we demonstrated in our policy analysis. Moreover, with minimal assumptions on the detailed microfoundation, we can integrate this class of financial intermediation models into state-of-the-art quantiative macro models with rich features such as household heterogeneity and illiquidity. This integration allows us to study a set of important policy questions that standard macro-finance models with a simplified household sector are not suited for.

Generalizing the financial sector's asset supply system to include various types of assets and intermediaries is a natural step for a comprehensive framework to study the transmission of policies and shocks through asset markets. Normative analysis based on such an asset supply system will allow us to understand how characteristics of the financial sector should shape a wide range of government policies, such as open market operations, quantitative easing and tightening, and Operation Twist. We leave these topics for future research.

References

Aguiar, Mark A, Mark Bils, and Corina Boar, "Who Are the Hand-to-Mouth?," Working Paper 26643, National Bureau of Economic Research January 2020.

Aguiar, Mark, Manuel Amador, and Cristina Arellano, "Micro Risks and Pareto Improving Policies with Low Interest Rates," Staff Report 625, Federal Reserve Bank of Minneapolis June 2021.

Akinci, Ozge and Albert Queralto, "Credit spreads, financial crises, and macroprudential policy," *American Economic Journal: Macroeconomics*, 2022, *14* (2).

Angeletos, George-Marios, Chen Lian, and Christian K Wolf, "Can Deficits Finance Themselves?," Technical Report, National Bureau of Economic Research 2023.

Auclert, Adrien, Bence Bardóczy, Matthew Rognlie, and Ludwig Straub, "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models," *Econometrica*, 2021, *89* (5), 2375–2408.

_ **, Matthew Rognlie, and Ludwig Straub**, "The Intertemporal Keynesian Cross," Technical Report 2023.

Bauer, Michael D and Eric T Swanson, "A reassessment of monetary policy surprises and high-frequency identification," *NBER Macroeconomics Annual*, 2023, *37* (1), 87–155.

Baumeister, Christiane and James D Hamilton, "Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks," *American Economic Review*, 2019, *109* (5), 1873–1910.

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist, "The financial accelerator in a quantitative business cycle framework," in "Handbook of Macroeconomics," Vol. 1, Elsevier, December 1999, chapter 21.

Bianchi, Javier and Enrique G Mendoza, "Optimal time-consistent macroprudential policy," *Journal of Political Economy*, 2018, *126* (2).

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

Chi, Chun-Che, Stephanie Schmitt-Grohé, and Martín Uribe, "Optimal Bank Reserve Remuneration and Capital Control Policy," Working Paper 29473, National Bureau of Economic Research November 2021.

Christiano, Lawrence J., Roberto Motto, and Massimo Rostagno, "Risk Shocks," *American Economic Review*, January 2014, *104* (1), 27–65.

Cui, Wei and Vincent Sterk, "Quantitative easing with heterogeneous agents," *Journal of Monetary Economics*, 2021, *123.*

Cúrdia, Vasco and Michael Woodford, "Credit frictions and optimal monetary policy," *Journal of Monetary Economics*, 2016, *84.*

den Heuvel, Skander J. Van, "The welfare cost of bank capital requirements," *Journal of Monetary Economics*, 2008, *55* (2).

Dávila, Eduardo and Andreas Schaab, "Optimal Monetary Policy with Heterogeneous Agents: Discretion, Commitment, and Timeless Policy," Working Paper 30961, National Bureau of Economic Research February 2023.

Fernández-Villaverde, Jesús, Samuel Hurtado, and Galo Nuño, "Financial frictions and the wealth distribution," Working Papers 2013, Banco de España 2020.

Ferrante, Francesco and Nils Gornemann, "Devaluations, Deposit Dollarization, and Household Heterogeneity," *International Finance Discussion Paper*, 2022, (1336).

Gabaix, Xavier, Ralph SJ Koijen, Federico Mainardi, Sangmin Oh, and Motohiro Yogo, "Asset demand of us households," Technical Report, National Bureau of Economic Research 2024.

Gertler, Mark and Nobuhiro Kiyotaki, "Chapter 11 - Financial Intermediation and Credit Policy in Business Cycle Analysis," in "in" Handbook of Monetary Economics, Elsevier, 2010.

— **and Peter Karadi**, "A model of unconventional monetary policy," *Journal of monetary Economics*, 2011, *58* (1).

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song, "What do data on millions of US workers reveal about life-cycle earnings risk?," Technical Report, National Bureau of Economic Research 2015.

Kaplan, Greg and Giovanni L. Violante, "The Marginal Propensity to Consume in Heterogeneous Agent Models," NBER Working Papers 30013, National Bureau of Economic Research, Inc 2022.

— **, Benjamin Moll, and Giovanni L Violante**, "Monetary policy according to HANK," *American Economic Review*, 2018, *108* (3).

— **, Giovanni L. Violante, and Justin Weidner**, "The Wealthy Hand-to-Mouth," *Brookings Papers on Economic Activity*, 2014, *45.*

Karadi, Peter and Anton Nakov, "Effectiveness and addictiveness of quantitative easing," *Journal of Monetary Economics*, 2021, *117* (C).

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

Lee, Donggyu, "Quantitative Easing and Inequality," mimeo, Federal Reserve Bank of New York 2021.

Lee, Seungcheol, Ralph Luetticke, and Morten Ravn, "Financial Frictions: Macro vs Micro Volatility," CEPR Discussion Papers 15133 2020.

McKay, Alisdair and Christian K. Wolf, "Optimal Policy Rules in HANK," Technical Report 2023.

Mendicino, Caterina, Lukas Nord, and Marcel Peruffo, "Distributive Effects of Banking Sector Losses," Technical Report 2021.

Ottonello, Pablo and Wenting Song, "Financial intermediaries and the macroeconomy: Evidence from a high-frequency identification," Technical Report, National Bureau of Economic Research 2022.

— , **Diego J Perez, and Paolo Varraso**, "Are collateral-constraint models ready for macroprudential policy design?," *Journal of International Economics*, 2022, *139.*

Plagborg-Møller, Mikkel and Christian K. Wolf, "Local Projections and VARs Estimate the Same Impulse Responses," *Econometrica*, 2021, *89* (2).

Uribe, Martin and Vivian Z. Yue, "Country spreads and emerging countries: Who drives whom?," *Journal of International Economics*, June 2006, *69* (1).

Wolf, Christian K., "Interest Rate Cuts vs. Stimulus Payments: An Equivalence Result," NBER Working Papers 29193, National Bureau of Economic Research, Inc August 2021.

— , "The Missing Intercept: A Demand Equivalence Approach," NBER Working Papers 29558, National Bureau of Economic Research, Inc December 2021.

A Proofs and Derivations

A.1 Proof of Proposition 1

Proof. To derive the response of liquidity supply to changes in returns, totally differentiating $d_t = (\Theta_t - 1)\, n_t$ and evaluating at the steady state gives

$$d\mathcal{D}_t = d\Theta_t \bar{n} + \left(\bar{\Theta} - 1\right) dn_t,$$

where the net worth process follows

$$dn_t = \bar{G} dn_{t-1} + (\bar{G}_\Theta d\Theta_{t-1} + \bar{G}_{r^K} dr_t^K + \bar{G}_{r^B} dr_t^B) n.$$

Consider changes in returns dr_s^K, dr_s^B at time s, and let $dr_t^K = dr_t^B = 0$, $\forall t \neq s$. Because Θ_{t-1} responds only to dr_s^K when $t \leq s$ and Θ_{-1} is pre-determined, we have $\frac{d\Theta_{t-u-1}}{dr_s^K} = 0$, for $u \leq t - s - 1$ or $u \geq t$, and

$$dn_t = \begin{cases} \sum_{u=0}^{t-1} \bar{G}_\Theta \bar{G}^u \frac{d\Theta_{t-u-1}}{dr_s^K} n \; dr_s^K, & s > t, \\ \sum_{u=t-s}^{t-1} \bar{G}_\Theta \bar{G}^u \frac{d\Theta_{t-u-1}}{dr_s^K} n \; dr_s^K, & s \leq t. \end{cases}$$

For class of intermediation frictions described in Section 3, Lemma 1 implies $\frac{d\Theta_{t-u-1}}{dr_s^K} = \gamma^{s-t+u}\bar{\Theta}_{r^K}, \forall t > u \geq t-s$. Substitute the expression and let $\sigma(s) \coloneqq \frac{1-(\gamma\bar{G})^s}{1-\gamma\bar{G}} \times \mathbf{1}_{\{s\geq 0\}}$, we have

$$dn_t = \begin{cases} \bar{G}_\Theta \gamma^{s-t}\sigma(t)\bar{\Theta}_{r^K} n dr_s^K, & s > t, \\ \bar{G}_\Theta \bar{G}^{t-s}\sigma(s)\bar{\Theta}_{r^K} n dr_s^K, & s \leq t. \end{cases}$$

Use $D_t = (\Theta_t - 1)n_t$, then

$$\frac{\partial \mathcal{D}_t / \partial r_s^K}{\mathcal{D}_t} = \begin{cases} \gamma^{s-t-1}\bar{\Theta}_{r^K}\left(\frac{1}{\Theta-1} + \gamma\Sigma(t)\right), & s > t, \\ \left(\bar{G}_{r^K} + \bar{\Theta}_{r^K}\Sigma(s)\right)\bar{G}^{t-s}, & s \leq t, \end{cases}$$

where $\Sigma(t) \coloneqq \bar{G}_\Theta \sigma(t)$

□

A.2 Proof of Lemma 2

Proof. We define the aggregate functions respectively for the household, production, and financial blocks of the model. Because these aggregate functions incorporate the optimality conditions for each block, sequences that satisfy market clearing given these functions represent an equilibrium.

Households

The solution of the household problem defines a set of mappings from after-tax income and returns, $\left\{y_{i,t} - \mathcal{T}(y_{i,t}), r_t^A, r_t^B\right\}_{s=0}^{\infty}$, to the optimal consumption, savings in each type of asset, and the adjustment cost for each household i.

From the firm's problem, we have $\frac{W_t}{P_t}h_t = (1-\alpha)\,y_t$. Because labor unions are identical, $h_{lt} = h_t$, and the labor demand rule implies $h_{i,t} = l(z_{i,t})h_t$ and $y_{i,t} = z_{i,t}l(z_{i,t})(1-\alpha)y_t$. Given the tax system, after-tax income for household i is given by:

$$y_{i,t} - \mathcal{T}(y_{i,t}) = (1-\tau_t)\left(z_{i,t}l(z_{i,t})(1-\alpha)y_t\right)^{1-\lambda}.$$

The tax rate τ_t consistent with tax revenue T_t satisfies:

$$\int y_{i,t} di - T_t = (1-\tau_t)\big((1-\alpha)y_t\big)^{1-\lambda} \int \big(z_{i,t} l(z_{i,t})\big)^{1-\lambda} di.$$

Therefore,

$$1-\tau_t = \frac{(1-\alpha)y_t - T_t}{\big((1-\alpha)y_t\big)^{1-\lambda} \int \big(z_{i,t} l(z_{i,t})\big)^{1-\lambda} di},$$

and individual after-tax income is given by:

$$y_{i,t} - \mathcal{T}(y_{i,t}) = \frac{\big(z_{i,t} l(z_{i,t})\big)^{1-\lambda}}{\int \big(z_{i,t} l(z_{i,t})\big)^{1-\lambda} di} \Big((1-\alpha)y_t - T_t\Big).$$

As a result, the optimal policy rules of individual households can be expressed as functions of the idiosyncratic state $\{z_{i,s}\}$ and $\left\{y_s, r_s^A; r_s^B, T_s\right\}_{s=0}^{\infty}$. Aggregation across individuals given the initial distribution of assets and productivity gives us the aggregate assets and consumption demand functions: $\mathcal{A}_t\left(\left\{y_s, r_s^A; r_s^B, T_s\right\}_{s=0}^{\infty}\right)$, $\mathcal{B}_t\left(\left\{y_s, r_s^A; r_s^B, T_s\right\}_{s=0}^{\infty}\right)$ and $\mathcal{C}_t\left(\left\{y_s, r_s^A; r_s^B, T_s\right\}_{s=0}^{\infty}\right)$, where we define the consumption function to include the adjustment cost:

$$\mathcal{C}_t\left(\left\{y_s, r_s^A; r_s^B, T_s\right\}_{s=0}^{\infty}\right) \coloneqq \int c_{i,t} + \Phi(a_{i,t}, a_{i,t-1}, r_t^A) di.$$

Production

To obtain the investment function use the law of motion for capital to get the investment ratio

$$\frac{x_t}{k_{t-1}} = \Gamma^{-1}\left(\frac{k_t - (1-\delta)\, k_{t-1}}{k_{t-1}}\right) \eqqcolon \iota(k_t, k_{t-1})$$

and use this in the first order condition with respect to ι_t, we have

$$q_t = \frac{1}{\Gamma'\left(\iota(k_t, k_{t-1})\right)} \eqqcolon \hat{q}\left(k_t, k_{t-1}\right)$$

All the above result in

$$1 + r_{t+1}^K = \frac{\alpha \frac{y_{t+1}}{k_t} + \hat{q}\left(k_{t+1}, k_t\right)\left(\frac{k_{t+1}}{k_t}\right) - \iota(k_{t+1}, k_t)}{\hat{q}\left(k_t, k_{t-1}\right)},$$

which can be solved to obtain capital in each period as a function of the path of

$\{y_s, r_s^K\}$, given initial capital k_{-1}: $\mathcal{K}_t\left(\left\{y_s, r_s^K\right\}_{s=0}^{\infty}\right)$.

We then use the law of motion for capital again to back out the investment function $\mathcal{X}_t\left(\left\{y_s, r_s^K\right\}_{s=0}^{\infty}\right)$. Moreover, we can express capital price as $q_t := \mathcal{Q}_t\left(\left\{y_s, r_s^K\right\}_{s=0}^{\infty}\right)$.

The Financial Sector

The liquidity supply functions given returns, $\mathcal{D}_t(\left\{r_s^K, r_s^B\right\}_{s=0}^{\infty})$, are defined as in Section 3. For the function $\mathcal{R}_t^A(\cdot)$, we using Equation 2. Define $L_t := d_t/\left(q_t k_t\right)$ to be the liquidity transformation ratio, which represents the share of capital held as liquid assets. The accounting identity in Equation 2 can be written as:

$$1 + r_{t+1}^A = \frac{(1 + r_{t+1}^K) - (1 + r_{t+1}^B)L_t}{1 - L_t}.$$

Using functions $\mathcal{D}_t\left(\left\{r_s^K, r_s^B\right\}_{s=0}^{\infty}\right)$, $\mathcal{Q}_t\left(\left\{y_s, r_s^K\right\}_{s=0}^{\infty}\right)$, and $\mathcal{K}_t\left(\left\{y_s, r_s^K\right\}_{s=0}^{\infty}\right)$, we can write L_t as $\mathcal{L}_t\left(\left\{y_s, r_s^K; \mathcal{D}_t\right\}_{s=0}^{\infty}\right)$, and $r_{t+1}^A = \mathcal{R}_{t+1}^A$ where

$$\mathcal{R}_{t+1}^A\left(\left\{r_s^K, r_s^B, y_s; \mathcal{D}_t\right\}_{s=0}^{\infty}\right) := \frac{\left(1 + r_{t+1}^K\right) - \left(1 + r_{t+1}^B\right)\mathcal{L}_t\left(\left\{y_s, r_s^K; \mathcal{D}_t\right\}_{s=0}^{\infty}\right)}{1 - \mathcal{L}_t\left(\{y_s, r_s^K; \mathcal{D}_t\}_{s=0}^{\infty}\right)} - 1 \quad (8)$$

Market Clearing

From the definition of the aggregate functions, the goods market clearing and liquid asset market clearing conditions are given by

$$\mathcal{C}_t(\{y_s, r_s^A; r_s^B, T_s\}_{s=0}^{\infty}) + \mathcal{X}_t(\{y_s, r_s^K\}_{s=0}^{\infty}) + g_t = y_t, \quad (9)$$

$$\mathcal{B}_t\left(\left\{y_s, r_s^A; r_s^B, T_s\right\}_{s=0}^{\infty}\right) = \mathcal{D}_t\left(\left\{r_s^K, r_s^B\right\}_{s=0}^{\infty}\right) + b_t^G. \quad (10)$$

Given $\left\{g_s, T_s, r_s^B, b_s^G\right\}_{s=0}^{\infty}$, let $\{y_s, r_s^K\}_{s=0}^{\infty}$ be a sequence that satisfies Equations 8, 10, and 10. We solve a_s^G from Equation 5, so the government budget constraint is satisfied. Because the aggregate functions for households are derived under household budget constraints, by the Walras law, the illiquid asset market clears

$$\mathcal{A}_t\left(\left\{y_s, r_s^A, r_s^B; T_s\right\}\right) = \mathcal{Q}_t\left(\left\{y_s, r_s^K\right\}_{s=0}^{\infty}\right)\mathcal{K}_t\left(\left\{y_s, r_s^K\right\}_{s=0}^{\infty}\right) - \mathcal{D}_t\left(\left\{r_s^K, r_s^B\right\}_{s=0}^{\infty}\right) - a_t^G.$$

Therefore, sequence $\{y_s, r_s^K\}_{s=0}^{\infty}$ constitute an equilibrium. □

A.3 Time 0 Returns

We express r_0^K as a function of output and expected returns by noting that

$$1 + r_0^K = \frac{\alpha \frac{y_0}{k_{-1}} + \hat{q}\left(k_0, k_{-1}\right)\left(\frac{k_0}{k_{-1}}\right) - \iota(k_0, k_{-1})}{\hat{q}\left(k_{-1}, k_{-2}\right)},$$

where only y_0 and k_0 are not predetermined. From the proof of Lemma 2, we have $k_0 = \mathcal{K}_0\left(\left\{y_s, r_{s+1}^K\right\}_{s=0}^{\infty}\right)$. This allows us to write r_0^K as a function of $\left\{y_s, r_s^K\right\}_{s=0}^{\infty}$.

A.4 Proof of Proposition 2.

Proof. Recall the definition of excess liquidity supply

$$\mathcal{E}_t(\boldsymbol{y}, \boldsymbol{r}^K, \boldsymbol{r}^B, \boldsymbol{T}, \boldsymbol{b}^G) := \mathcal{D}_t(r_0^K(\boldsymbol{y}, \boldsymbol{r}^K), \boldsymbol{r}^K, \boldsymbol{r}^B) + b_t^G - \mathcal{B}_t(\boldsymbol{y}, r^A(\boldsymbol{y}, \boldsymbol{r}^K, \boldsymbol{r}^B), \boldsymbol{r}^B, \boldsymbol{T}).$$

Liquid asset market clears if $\mathcal{E}_t(\boldsymbol{y}, \boldsymbol{r}^K, \boldsymbol{r}^B, \boldsymbol{T}, \boldsymbol{b}^G) = 0$. By totally differentiating this condition in every period we have

$$\boldsymbol{\epsilon}_{r^K} d\boldsymbol{r}^{\boldsymbol{K}} + \boldsymbol{\epsilon}_y d\boldsymbol{y} + \boldsymbol{\epsilon}_T d\boldsymbol{T} + d\boldsymbol{b}^{\boldsymbol{G}} + \boldsymbol{\epsilon}_{r^B} d\boldsymbol{r}^{\boldsymbol{B}} = \boldsymbol{0},$$

where $\boldsymbol{\epsilon}_{r^K} := \mathbf{D}_{r^K} - \tilde{\mathbf{B}}_{r^K}$, $\boldsymbol{\epsilon}_{r^B} := \mathbf{D}_{r^B} - \tilde{\mathbf{B}}_{r^B}$, $\boldsymbol{\epsilon}_y := \mathbf{D}_y - \tilde{\mathbf{B}}_y$, $\boldsymbol{\epsilon}_T := -\tilde{\mathbf{B}}_T$, and the matrices are defined in Appendix A.6. Rearrange and left-multiply by the inverse of $-\boldsymbol{\epsilon}_{r^K}$ to obtain Equation 6.

For the second part of Proposition 2, note that if $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B} \to \infty$ and $\bar{\Theta}_{r^B}/\bar{\Theta}_{r^K} \to \varsigma$, Proposition 1 implies

$$\frac{\partial \mathcal{D}_t}{\partial r_s^K}\frac{1}{\bar{\Theta}_{r^K}} \to \begin{cases} \Sigma(s) G^{t-s}(\bar{\Theta} - 1)n, & s \leq t, \\ \gamma^{s-t-1}\Big(1 + (\bar{\Theta} - 1)\gamma\Sigma(t)\Big)n, & s > t, \end{cases}$$

$$\frac{\partial \mathcal{D}_t}{\partial r_s^B}\frac{1}{\bar{\Theta}_{r^K}} \to \begin{cases} -\varsigma\Sigma(s) G^{t-s}(\bar{\Theta} - 1)n, & s \leq t, \\ -\varsigma\gamma^{s-t-1}\Big(1 + (\bar{\Theta} - 1)\gamma\Sigma(t)\Big)n, & s > t. \end{cases}$$

We can write it as $\mathbf{D}_{r^K}\frac{1}{\bar{\Theta}_{r^K}} \to \mathbf{D}_{\infty,r},\ \mathbf{D}_{r^B}\frac{1}{\bar{\Theta}_{r^K}} \to -\varsigma\mathbf{D}_{\infty,r}$, where

$$\mathbf{D}_{\infty,r} \coloneqq \begin{cases} \Sigma(s)G^{t-s}(\bar{\Theta}-1)n, & s \le t \\ \gamma^{s-t-1}\Big(1+(\bar{\Theta}-1)\gamma\Sigma(t)\Big)n, & s > t. \end{cases}$$

Assume that first derivatives of $\mathcal{B}_t$ are bounded. Divide the linearized liquid asset market clearing condition by $\bar{\Theta}_{r^K}$. As $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B} \to \infty$ with $\bar{\Theta}_{r^B}/\bar{\Theta}_{r^K} \to \varsigma$, for all bounded sequences $\left\{d\boldsymbol{y}, d\boldsymbol{r}^K, d\boldsymbol{r}^B, d\boldsymbol{b}^G\right\}$, the limit of the liquid asset market clearing condition is

$$\left(\mathbf{I} - \mathbf{B}_{r^A}\frac{r^K - r^B}{(1-L)^2}\frac{L}{d}\right)\mathbf{D}_r^{\infty}\left(d\boldsymbol{r}^K - \varsigma d\boldsymbol{r}^B\right) = \mathbf{0},$$

where $L = d/(qk)$. The condition is satisfied for $d\boldsymbol{r}^K = \varsigma d\boldsymbol{r}^B$. □

A.5 Proof of Theorem 1.

Proof. The aggregate demand is defined as

$$\Psi_t(\boldsymbol{y}, \boldsymbol{r}^K, \boldsymbol{r}^B, \boldsymbol{T}, \boldsymbol{g}) \coloneqq \mathcal{C}_t(\boldsymbol{y}, r^A(\boldsymbol{y}, \boldsymbol{r}^K, \boldsymbol{r}^B), \boldsymbol{r}^B, \boldsymbol{T}) + \mathcal{X}_t(\boldsymbol{y}, \boldsymbol{r}^K) + g_t.$$

Goods market clears if $\Psi_t(\boldsymbol{y}, \boldsymbol{r}^K, \boldsymbol{r}^B, \boldsymbol{T}, \boldsymbol{g}) = y_t$. By totally differentiating this condition in every period we have

$$\boldsymbol{\Psi}_{r^K}d\boldsymbol{r}^{\boldsymbol{K}} + \boldsymbol{\Psi}_y d\boldsymbol{y} + \boldsymbol{\Psi}_T d\boldsymbol{T} + d\boldsymbol{b}^{\boldsymbol{G}} + \boldsymbol{\Psi}_{r^B}d\boldsymbol{r}^{\boldsymbol{B}} + d\boldsymbol{g} = d\boldsymbol{y}$$

where $\boldsymbol{\Psi}_{r^K} \coloneqq \tilde{\mathbf{C}}_{r^K} + \mathbf{X}_{r^K},\ \boldsymbol{\Psi}_{r^B} \coloneqq \tilde{\mathbf{C}}_{r^B},\ \boldsymbol{\Psi}_y \coloneqq \tilde{\mathbf{C}}_y + \mathbf{X}_y,\ \boldsymbol{\Psi}_T \coloneqq \tilde{\mathbf{C}}_T$, and the matrices are defined in Appendix A.6.

Let $\boldsymbol{\Omega} \coloneqq \boldsymbol{\Psi}_{r^K}(-\boldsymbol{\epsilon}_{r^K}^{-1})$, and use Proposition 2 to write

$$\boldsymbol{\Omega}\left(\boldsymbol{\epsilon}_y d\boldsymbol{y} + \boldsymbol{\epsilon}_T d\boldsymbol{T} + d\boldsymbol{b}^{\boldsymbol{G}} + \boldsymbol{\epsilon}_{r^B}d\boldsymbol{r}^{\boldsymbol{B}}\right) + \boldsymbol{\Psi}_y d\boldsymbol{y} + \boldsymbol{\Psi}_T d\boldsymbol{T} + d\boldsymbol{b}^{\boldsymbol{G}} + \boldsymbol{\Psi}_{r^B}d\boldsymbol{r}^{\boldsymbol{B}} + d\boldsymbol{g} = d\boldsymbol{y}.$$

Finally, rearrange it as

$$d\boldsymbol{y} = \left(\mathbf{I} - \boldsymbol{\Psi}_y - \boldsymbol{\Omega}\ \boldsymbol{\epsilon}_y\right)^{-1} \times \Big(\ d\boldsymbol{g} + \boldsymbol{\Psi}_T d\boldsymbol{T} + \boldsymbol{\Psi}_{r^B}d\boldsymbol{r}^B + \boldsymbol{\Omega}\big(d\boldsymbol{b}^G + \boldsymbol{\epsilon}_T d\boldsymbol{T} + \boldsymbol{\epsilon}_{r^B}d\boldsymbol{r}^{\boldsymbol{B}}\big)\Big),$$

which is the formula in Theorem 1. □

A.6 Additional Derivations: Linearized equilibrium conditions

We evaluate derivatives of aggregate functions $\mathcal{X}_t(\cdot), \mathcal{B}_t(\cdot), \mathcal{C}_t(\cdot), \mathcal{D}_t(\cdot),\ \mathcal{R}_t^A(\cdot)$ at the steady state and represent them as matrices. We use the following notation: $d\boldsymbol{r}^B$ represents $\{dr_{s+1}^B\}_{s=0}^\infty$ as a column vector. The same convention applies to other rates of return. We use $d\boldsymbol{y}$ to represent $\{dy_s\}_{s=0}^\infty$ as a column vector, and similar for other variables that are not rates of return.

Production

Linearization of the formula for return on capital results in

$$d\boldsymbol{r}^{\boldsymbol{K}} + \frac{\left(1+r^K\right)\bar{q}'}{k}(\mathbf{I}-\mathbf{S}_{-1})d\boldsymbol{k} = \frac{\alpha}{k}\mathbf{S}_{+1}d\boldsymbol{y} - \frac{\alpha y}{k^2}d\boldsymbol{k} + \frac{\bar{q}'+\bar{q}-\bar{\iota}'}{k}(\mathbf{S}_{+1}-\mathbf{I})d\boldsymbol{k}$$

which allows us to express $d\boldsymbol{k}$ as

$$d\boldsymbol{k} = \mathbf{K}_y d\boldsymbol{y} + \mathbf{K}_{r^K} d\boldsymbol{r}^{\boldsymbol{K}},$$

where $\mathbf{K}_y \coloneqq \Xi^{-1}\frac{\alpha}{k}\mathbf{S}_{+1},\ \mathbf{K}_{r^K} \coloneqq -\Xi^{-1}$, and

$$\Xi \coloneqq \frac{\alpha y}{k^2}\mathbf{I} + \frac{\left(1+r^K\right)\bar{q}'}{k}\left(\mathbf{I}-\mathbf{S}_{-1}\right) - \frac{\bar{q}'+\bar{q}-\bar{\iota}'}{k}(\mathbf{S}_{+1}-\mathbf{I}),$$

where

$$\mathbf{S}_{+1} \coloneqq \begin{bmatrix} 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \quad \mathbf{S}_{-1} \coloneqq \begin{bmatrix} 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Linearizing the expression for $\iota(k_t, k_{t-1})$ from the proof of Lemma 2, we have $d\boldsymbol{x} = (\bar{\iota}'(\mathbf{I}-\mathbf{S}_{-1})+\bar{\iota})d\boldsymbol{k}$, and we can write

$$d\boldsymbol{x} = \mathbf{X}_y d\boldsymbol{y} + \mathbf{X}_{r^K} d\boldsymbol{r}^{\boldsymbol{K}},$$

where $\mathbf{X}_y \coloneqq (\bar{\iota}'(\mathbf{I}-\mathbf{S}_{-1})+\bar{\iota})\mathbf{K}_y,\ \mathbf{X}_{r^K} \coloneqq (\bar{\iota}'(\mathbf{I}-\mathbf{S}_{-1})+\bar{\iota})\mathbf{K}_{r^K}$.

Linearizing $\hat{q}\left(k_t, k_{t-1}\right)$ from the proof of Lemma 2, we have $d\boldsymbol{q} = \frac{\bar{q}'}{k}\left(\mathbf{I}-\mathbf{S}_{-1}\right)d\boldsymbol{k}$, and we can write it as

$$d\boldsymbol{q} = \mathbf{Q}_y d\boldsymbol{y} + \mathbf{Q}_{r^K} d\boldsymbol{r}^{\boldsymbol{K}},$$

where $\mathbf{Q}_y \coloneqq \frac{\bar{q}'}{k}\left(\mathbf{I} - \mathbf{S}_{-1}\right)\mathbf{K}_y,\ \mathbf{Q}_{r^K} \coloneqq \frac{\bar{q}'}{k}\left(\mathbf{I} - \mathbf{S}_{-1}\right)\mathbf{K}_{r^K}.$

Besides these matrices, the time 0 return on capital response, dr_0^K, can be expressed as

$$dr_0^K = \alpha \frac{1}{\bar{k}} dy_0 + (1 - \delta)\, dq_0.$$

In a matrix form, we can write

$$dr_0^K = \frac{\alpha}{\bar{k}} \mathbf{e}_1^\intercal d\boldsymbol{y} + (1 - \delta)\left(\mathbf{q}_y^\intercal d\boldsymbol{y} + \mathbf{q}_{r^K}^\intercal d\boldsymbol{r}^K\right), \tag{11}$$

where $\mathbf{q}_y^\intercal, \mathbf{q}_{r^K}^\intercal$ are row vectors from the first rows of $\mathbf{Q}_y$, $\mathbf{Q}_{r^K}$, describing how the price of capital at time 0 depends on output and return on capital. $\mathbf{e}_1^\intercal$ is a row vector with 1 as its first entry, and zeros elsewhere

Liquidity Supply

Financial intermediation in the economy is characterized as derivatives of the liquidity supply function

$$\mathcal{D}_t(r_0^K(\boldsymbol{y}, \boldsymbol{r}^K), \boldsymbol{r}^K, \boldsymbol{r}^B).$$

Let $\mathbf{D}_{r^K}$ be a matrix of total derivatives of $\mathcal{D}_t(\cdot)$ with respect to rates of return on capital $\boldsymbol{r}^K$. Its $(t+1, s+1)$ entry is a total derivative of $\mathcal{D}_t(\cdot)$ with respect to r_{s+1}^K. $\mathbf{D}_{r^B}$ is defined similarly. Notice the difference in timing for rows and columns. Entry $(t+1, s+1)$ of $\mathbf{D}_y$ is a total derivative of $\mathcal{D}_t(\cdot)$ with respect to y_s.

The formulas from Proposition 2 imply the matrices have the following form, with minor modifications using equation 11 to incorporate the dependence of time-0 return on capital, r_0^K, on future returns on capital:

$$\mathbf{D}_{r^K} = d \times \left(\mathcal{D}_{r^K} + \mathbf{n}_0(1-\delta)\mathbf{q}_{r^K}^\intercal\right),$$

$$\mathbf{D}_{r^B} = d \times \mathcal{D}_{r^B},$$

$$\mathbf{D}_y = d \times \mathbf{n}_0\left(\frac{\alpha}{k}\mathbf{e}_1^\intercal + (1-\delta)\mathbf{q}_y^\intercal\right),$$

where d is steady state liquidity supply and where the (t+1,s) elements of $\mathcal{D}_{r^K}, \mathcal{D}_{r^B}$ are $\frac{\partial \mathcal{D}_t / \partial r_s^K}{\mathcal{D}_t}$ and $\frac{\partial \mathcal{D}_t / \partial r_s^B}{\mathcal{D}_t}$ from Proposition 1. $\mathbf{D}_{r^K}$ is modified to capture the response of dr_0^K to future returns of capital, $d\boldsymbol{r}^k$, where $\mathbf{n}_0$ is a column vector that traces the propagation of net worth with its $t+1$ element being $\bar{G}_{r^K} G^t$. Similarly, $\mathbf{D}_y$ captures the response of dr_0^K to output, $d\boldsymbol{y}$.

Illiquid asset return

Before discussing linearization of the household side of the economy, we provide formulas that allow us to express dr_t^A as a function of other variables. For dr_0^A, we have $dr_0^A = dr_0^K/(1-L)$ where $L = d/qk$ is the steady state value of L_t.

For matrices that relate $\{dr_{s+1}^A\}_{s=0}^{\infty}$ to $\{dr_{s+1}^K, dr_{s+1}^B, dy_s\}_{s=0}^{\infty}$, Equation 8 implies

$$\mathbf{R}_{r^K}^A = \frac{1}{1-L}\mathbf{I} + \frac{r^K - r^B}{(1-L)^2}\mathbf{L}_{r^K}, \quad \mathbf{R}_{r^B}^A = \frac{L}{1-L}\mathbf{I} + \frac{r^K - r^B}{(1-L)^2}\mathbf{L}_{r^B}, \quad \mathbf{R}_y^A = \frac{r^K - r^B}{(1-L)^2}\mathbf{L}_y.$$

From $L_t = \frac{d_t}{q_t k_t}$, we have

$$\mathbf{L}_{r^K} = -\frac{L}{q}\mathbf{Q}_{r^K} - \frac{L}{k}\mathbf{K}_{r^K} + \frac{L}{d}\mathbf{D}_{r^K}, \quad \mathbf{L}_{r^B} = \frac{L}{d}\mathbf{D}_{r^B}, \quad \mathbf{L}_y = -\frac{L}{q}\mathbf{Q}_y - \frac{L}{k}\mathbf{K}_y + \frac{L}{d}\mathbf{D}_y.$$

Households

Let $\mathbf{C}_{r^A}$ be a matrix, whose $(t+1, s)$ element is a partial derivative of $\mathcal{C}_t$ with respect to r_s^A. We use the same convention for $\mathbf{C}_{r^B}$ Similarly, $\mathbf{C}_y$ is a matrix of partial derivatives of $\mathcal{C}_t$ with respect to aggregate output. Its $(t+1, s+1)$ elements is a partial derivative of $\mathcal{C}_t$ with respect to y_s. $\mathbf{C}_T$ is defined analogously. Similarly, we define all matrices that contain derivatives of $\mathcal{B}$.

While these matrices capture responses to all returns $\{r_{s+1}^A, \forall s \geq 0\}$, they miss the response to r_0^A. To capture the response, define the following matrices

$$\mathbf{B}_{r_0^A} := \begin{bmatrix} \frac{\partial \mathcal{B}_0}{\partial r_0^A} & 0 & 0 & \cdots \\ \frac{\partial \mathcal{B}_1}{\partial r_0^A} & 0 & 0 & \cdots \\ \frac{\partial \mathcal{B}_2}{\partial r_0^A} & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}, \quad \mathbf{B}_{r_0^A} := \begin{bmatrix} \frac{\partial \mathcal{C}_0}{\partial r_0^A} & 0 & 0 & \cdots \\ \frac{\partial \mathcal{C}_1}{\partial r_0^A} & 0 & 0 & \cdots \\ \frac{\partial \mathcal{C}_2}{\partial r_0^A} & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix},$$

and use Equation 11 to define

$$\tilde{\mathbf{B}}_{r_0^A,y} := \frac{1}{1-L}\mathbf{B}_{r_0^A} \times \left[\frac{\alpha}{\bar{k}}\mathbf{e}_1^\intercal + (1-\delta)\,\mathbf{q}_y^\intercal\right], \quad \tilde{\mathbf{B}}_{r_0^A,r^K} := \frac{1}{1-L}\mathbf{B}_{r_0^A} \times (1-\delta)\,\mathbf{q}_{r^K}^\intercal,$$

$$\tilde{\mathbf{C}}_{r_0^A,y} := \frac{1}{1-L}\mathbf{C}_{r_0^A} \times \left[\frac{\alpha}{\bar{k}}\mathbf{e}_1^\intercal + (1-\delta)\,\mathbf{q}_y^\intercal\right], \quad \tilde{\mathbf{C}}_{r_0^A,r^K} := \frac{1}{1-L}\mathbf{C}_{r_0^A} \times (1-\delta)\,\mathbf{q}_{r^K}^\intercal.$$

With these matrices capturing the effect of $d\boldsymbol{y}$ and $d\boldsymbol{r^K}$ on consumption and asset demand through dr_0^A, we define the full consumption responses as:

$$\tilde{\mathbf{C}}_y := \mathbf{C}_y + \mathbf{C}_{r^A}\mathbf{R}_y^A + \tilde{\mathbf{C}}_{r_0^A,y}, \qquad \tilde{\mathbf{C}}_{r^K} := \mathbf{C}_{r^A}\mathbf{R}_{r^K}^A + \tilde{\mathbf{C}}_{r_0^A,r^K},$$

$$\tilde{\mathbf{C}}_{r^B} := \mathbf{C}_{r^B} + \mathbf{C}_{r^A}\mathbf{R}_{r^B}^A, \qquad \tilde{\mathbf{C}}_T := \mathbf{C}_T.$$

Similarly, the full liquid asset demand responses are defined as:

$$\tilde{\mathbf{B}}_y := \mathbf{B}_y + \mathbf{B}_{r^A}\mathbf{R}_y^A + \tilde{\mathbf{B}}_{r_0^A,y}, \qquad \tilde{\mathbf{B}}_{r^K} := \mathbf{B}_{r^A}\mathbf{R}_{r^K}^A + \tilde{\mathbf{B}}_{r_0^A,r^K},$$

$$\tilde{\mathbf{B}}_{r^B} := \mathbf{B}_{r^B} + \mathbf{B}_{r^A}\mathbf{R}_{r^B}^A, \qquad \tilde{\mathbf{B}}_T := \mathbf{B}_T.$$

With these matrices, we can construct the matrices that represent the responses of excess liquidity $\mathcal{E}_t$ and aggregate demand $\mathbf{\Psi}_t$:

$$\boldsymbol{\epsilon}_{r^K} := \mathbf{D}_{r^K} - \tilde{\mathbf{B}}_{r^K},\ \boldsymbol{\epsilon}_{r^B} := \mathbf{D}_{r^B} - \tilde{\mathbf{B}}_{r^B},\ \boldsymbol{\epsilon}_y := \mathbf{D}_y - \tilde{\mathbf{B}}_y,\ \boldsymbol{\epsilon}_T := -\tilde{\mathbf{B}}_T$$

and

$$\mathbf{\Psi}_{r^K} \coloneqq \tilde{\mathbf{C}}_{r^K} + \mathbf{X}_{r^K},\ \mathbf{\Psi}_{r^B} \coloneqq \tilde{\mathbf{C}}_{r^B},\ \mathbf{\Psi}_y \coloneqq \tilde{\mathbf{C}}_y + \mathbf{X}_y,\ \mathbf{\Psi}_T \coloneqq \tilde{\mathbf{C}}_T.$$

B Nested Models and Extensions

B.1 Nested Models of Financial Frictions

We show how our framework nests some commonly used models of financial frictions by appropriately choosing the financial constraint $\Theta\left(\left\{r_{s+1}^B, r_{s+1}^K\right\}_{s\geq t}\right)$. We also demonstrate that in all these models financial frictions result in $\Theta_t(\cdot)$ that has the special structure we use in Lemma 1.

Gertler-Karadi-Kiyotaki

In Gertler and Kiyotaki (2010) and Gertler and Karadi (2011) there is a continuum of banks indexed by $j \in [0,1]$. Bank activity is subject to an agency problem. Every period, after receiving returns on assets and paying depositors, bank j exits with probability f and transfers its retained earnings as dividends to its owners. At the same time, a new bank enters and receives some initial net worth to operate with. Conditional on surviving, bank j chooses how much loans $l^B_{j,t}$ and deposits $d_{j,t}$ to issue. Banks cannot issue equity. Moreover, an agency problem constrains the amount of deposits they can issue. After obtaining funding from depositors and investing in assets (loans), bank j can divert fraction $1/\theta$ of assets and run away. If this happens, depositors force it into bankruptcy and bank j has to close. The largest amount of funding an intermediary can receive from depositors depends on the franchise value $v_{j,t}\left(n_{j,t}\right)$, where $n_{j,t}$ is net worth — bank j must be better off continuing instead of running away. The optimization problem is:

$$v_{j,t}\left(n_{j,t}\right) = \max_{\left\{l^B_{j,t+s}, d_{j,t+s}, n_{j,t+s+1}\right\}_{s=0}^{\infty}} \sum_{s=1}^{\infty} \Lambda_{t,t+s} \left(1-f\right)^{s-1} f n_{j,t+s}$$

subject to

$$l^B_{j,t} \leq \theta_t v_{j,t}\left(n_{j,t}\right), \quad n_{j,t} + d_{j,t} = l^B_{j,t}, \quad n_{j,t+1} = \left(1 + r^K_{t+1}\right) l^B_{j,t} - \left(1 + r^B_{t+1}\right) d_{j,t}.$$

The first constraint is the incentive compatibility constraint resulting from the agency problem. $\Lambda_{t,t+s}$ is the discount factor used by banks.[11] We can write the value function in a recursive form:

$$v_{j,t}\left(n_{j,t}\right) = \max_{l^B_{j,t}, d_{j,t}, n_{j,t+1}} \Lambda_{t,t+1} \left(f n_{j,t+1} + \left(1-f\right) v_{j,t+1}\left(n_{j,t+1}\right)\right).$$

Guess linearity: $v_{j,t}\left(n_{j,t}\right) = \eta_{j,t} n_{j,t}$. Define $\psi_{j,t} := l^B_{j,t}/n_{j,t}$. Bellman equation is

$$\begin{aligned} \eta_{j,t} n_{j,t} = & \max_{\psi_{j,t}} \Lambda_{t,t+1} \left(f + \left(1-f\right) \eta_{j,t+1}\right) \left[1 + r^B_{t+1} + \left(r^K_{t+1} - r^B_{t+1}\right) \psi_{j,t}\right] n_{j,t} \\ & + \lambda_{j,t} \left[\eta_{j,t} - \frac{1}{\theta} \psi_{j,t}\right] n_{j,t}. \end{aligned}$$

The guess that $v_{j,t}\left(n_{j,t}\right) = \eta_{j,t} n_{j,t}$ is verified if $\lambda_{j,t} < 1$.

[11] Gertler and Kiyotaki (2010) and Gertler and Karadi (2011) use the representative household's discount factor. With perfect foresight, it corresponds to r^B_{t+1}.

By complementarity slackness $\lambda_{j,t}\left[\eta_{j,t}-\frac{1}{\theta}\psi_{j,t}\right]=0$ and we can write

$$\eta_{j,t}n_{j,t}=\max_{\psi_{j,t}}\Lambda_{t,t+1}\left(f+\left(1-f\right)\eta_{j,t+1}\right)\left[1+r^B_{t+1}+\left(r^K_{t+1}-r^B_{t+1}\right)\psi_{j,t}\right]n_{j,t}.$$

If the incentive compatibility constraint is binding, we have

$$\eta_{j,t}=\Lambda_{t,t+1}\left(f+\left(1-f\right)\eta_{j,t+1}\right)\left[1+r^B_{t+1}+\left(r^K_{t+1}-r^B_{t+1}\right)\eta_{j,t}\theta\right].$$

Let $\Theta_{j,t}=\theta\eta_t$, then we have

$$\Theta_{j,t}=\frac{\Lambda_{t,t+1}\left(f\theta+\left(1-f\right)\Theta_{j,t+1}\right)\left(1+r^B_{t+1}\right)}{1-\Lambda_{t,t+1}\left(f\theta+\left(1-f\right)\Theta_{j,t+1}\right)\left(r^K_{t+1}-r^B_{t+1}\right)}. \tag{12}$$

Because all banks face the same rates of return and use the same discount rate, $\Theta_{j,t}$ is identical for all j. Therefore, we use Θ_t to denote all $\Theta_{j,t}$, and it follows that $l^B_{j,t}=\Theta_t n_{j,t}$. Given any discount rate $\Lambda_{s-1,s}=\Lambda(r^K_s,r^B_s)$, we can iterate Equation 12 forward and write $\Theta_t=\Theta\left(\left\{r^B_{s+1},r^K_{s+1}\right\}_{s\geq t}\right)$. Aggregating individual banks $\int_0^1 l^B_{j,t}dj=q_tk^B_t$ and $\int_0^1 n_{j,t}dj=n^B_t$ we obtain

$$q_tk^B_t=\Theta\left(\left\{r^B_{s+1},r^K_{s+1}\right\}_{s\geq t}\right)n_t$$

which coincides with the solution to the bank's problem described in Section 2.3.

We obtain the expressions for $\bar{\Theta}_{r^K},\bar{\Theta}_{r^B}$ and γ by differentiating Equation 12 with respect to returns and evaluating the resulting expression at the steady state. Depending on assumptions on bankers' discount rates, we have:

If $\Lambda_{s-1,s}=1/\left(1+r^K_s\right)$,

$$\bar{\Theta}_{r^K}=\frac{\bar{\Theta}(\bar{\Theta}-1)}{1+r^K},\quad \bar{\Theta}_{r^B}=\frac{\bar{\Theta}(\bar{\Theta}-1)}{1+r^B},\quad \gamma=\frac{(1-f)(1+r^B+\left(r^K-r^B\right)\bar{\Theta})^2}{(1+r^K)(1+r^B)}.$$

If $\Lambda_{s-1,s}=1/\left(1+r^B_s\right)$,

$$\bar{\Theta}_{r^K}=\frac{\bar{\Theta}^2}{1+r^B},\quad \bar{\Theta}_{r^B}=\frac{1+r^K}{1+r^B}\frac{\bar{\Theta}^2}{1+r^B},\quad \gamma=\frac{(1-f)(1+r^B+\left(r^K-r^B\right)\bar{\Theta})^2}{(1+r^B)^2}.$$

If $\Lambda_{s-1,s}=1/(1+\tilde{r})$ for some constant $\tilde{r}$ (e.g., $\Lambda_{s-1,s}=\beta$ in Lee et al. (2020)),

$$\bar{\Theta}_{r^K}=\frac{\bar{\Theta}^2}{1+r^B},\quad \bar{\Theta}_{r^B}=\frac{\bar{\Theta}(\bar{\Theta}-1)}{1+r^B},\quad \gamma=\frac{(1-f)(1+r^B+\left(r^K-r^B\right)\bar{\Theta})^2}{(1+\tilde{r})(1+r^B)}.$$

Bernanke, Gertler, Gilchrist (1999)

In Bernanke et al. (1999) financial frictions arise because of "costly state verification". In their model, there is a continuum of entrepreneurs that need to finance capital purchases. Realized returns are idiosyncratic and cannot be observed by the lenders unless they incur a monitoring cost. This creates a link between entrepreneurs' capital expenditures, their net worth, and the spread between the expected return on capital and the safe rate. Entrepreneurs face a constant probability of exit f and consume their retained earnings upon exiting. We can interpret entrepreneurs as banks and map this model to our framework. The key condition in Bernanke et al. (1999) is Equation 3.8 (p. 1353)

$$q_t k_t^B = \psi\left(\frac{1+r_{t+1}^K}{1+r_{t+1}^B}\right) n_t$$

with $\psi'(\cdot) > 0$ and $\psi(1) = 1$.[12] If we define $\Theta\left(\left\{r_{s+1}^K, r_{s+1}^B\right\}_{s\geq t}\right) := \psi\left(\frac{1+r_{t+1}^K}{1+r_{t+1}^B}\right)$, the solution to the bank's problem described in Section 2.3 and dynamics of bank net worth will coincide with the one in Bernanke et al. (1999). Notice that here the financial friction at time t depends only on r_{t+1}^K and r_{t+1}^B and not on returns more than one period ahead. In this model

$$\bar{\Theta}_{r^K} = \psi'\left(\frac{1+r^K}{1+r^B}\right)\frac{1}{1+r^B}, \quad \bar{\Theta}_{r^B} = \psi'\left(\frac{1+r^K}{1+r^B}\right)\frac{1+r^K}{(1+r^B)^2}, \quad \gamma = 0.$$

Costly leverage

Uribe and Yue (2006), Chi et al. (2021) and Cúrdia and Woodford (2016) consider reduced form financial frictions. They assume that banks need to incur a resource cost that depends on the level of financial intermediation. Since the marginal cost of intermediation is increasing in the scale of intermediation, there will be a link between the leverage ratio and the spread between returns on assets held by banks and deposits. Our framework allows us to nest these models without any modification to the framework if we assume that this cost is borne in units of utility or that it is rebated back lump-sum to the bank. We need to make this change to ensure that the law of motion for n_t in Equation 4 remains the same. More specifically, assume that

[12] There is no aggregate uncertainty in our framework, and this explains why there is no expectation operator in front of r_{t+1}^K.

the bank maximizes

$$r^N_{t+1} n_t = \max_{k^B_t, d_t} r^K_{t+1} q_t k^B_t - r^B_{t+1} d_t - \Upsilon_t \left(\frac{q_t k^B_t}{n_t} \right) n_t + \bar{\Upsilon}_t$$

subject to balance sheet $q_t k^B_t = d_t + n_t$.

Here $\Upsilon_t \left(\frac{q_t k^B_t}{n_t} \right) n_t$ captures costs related to financial intermediation. $\bar{\Upsilon}_t$ is the lump-sum rebate, equal to intermediation costs in equilibrium (alternatively we can assume that the cost is in disutility). Assume it is strictly increasing in the leverage ratio $\psi_t := q_t k^B_t / n_t$. First order condition is

$$r^K_{t+1} - r^B_{t+1} = \Upsilon'_t \left(\frac{q_t k^B_t}{n_t} \right),$$

which can be rewritten as

$$q_t k^B_t = \Upsilon'^{-1}_t \left(r^K_{t+1} - r^B_{t+1} \right) n_t.$$

If we define $\Theta \left(\left\{ r^K_{s+1}, r^B_{s+1} \right\}_{s \geq t} \right) := \Upsilon'^{-1}_t \left(r^K_{t+1} - r^B_{t+1} \right)$, then the solution to the bank's problem described in Section 2.3 will be the same as the one to the problem stated above. Note that Θ_t does not depend on returns more than one period in the future. Moreover, since $\Upsilon_t \left(\frac{q_t k^B_t}{n_t} \right) n_t = \bar{\Upsilon}_t$, $r^N_{t+1} n_t$ is the same as in section. In this model

$$\bar{\Theta}_{r^K} = \frac{1}{\Upsilon'' \left(\frac{qk^B}{n} \right)}, \quad \bar{\Theta}_{r^B} = \frac{1}{\Upsilon'' \left(\frac{qk^B}{n} \right)}, \quad \gamma = 0.$$

Collateral constraints

Consider a collateral constraint in which banks can pledge a fraction $\vartheta < 1$ of the value of their capital holdings along with returns on their capital. The highest possible level of net liquid asset issuance d_t satisfies

$$\left(1 + r^B_{t+1} \right) d_t \leq \vartheta \left(1 + r^K_{t+1} \right) q_t k^B_t.$$

By using the balance sheet, we can rewrite it as

$$q_t k^B_t \leq \frac{1 + r^B_{t+1}}{1 + r^B_{t+1} - \vartheta \left(1 + r^K_{t+1} \right)} n_t. \tag{13}$$

We can map it to our framework by defining

$$\Theta\left(\left\{r_{s+1}^{K}, r_{s+1}^{B}\right\}_{s\geq t}\right) := \frac{1+r_{t+1}^{B}}{1+r_{t+1}^{B}-\vartheta\left(1+r_{t+1}^{K}\right)}.$$

Taking derivatives and evaluating at the steady-state, we have

$$\bar{\Theta}_{r^K} = \frac{\vartheta\bar{\Theta}}{1+r^B-\vartheta\left(1+r^K\right)}, \quad \bar{\Theta}_{r^B} = -\frac{1+r^K}{1+r^B}\frac{\vartheta\bar{\Theta}}{1+r^B-\vartheta\left(1+r^K\right)}, \quad \gamma = 0,$$

where $\vartheta = (1-\frac{1}{\bar{\Theta}})\frac{1+r^B}{1+r^K}$ is linked to the steady-state returns and leverage.

Comparsion to Kiyotaki and Moore (1997)

Kiyotaki and Moore (1997) assume only the value of capital next period can be pledged as collateral. The constraint is

$$\left(1+r_{t+1}^{B}\right)d_t \leq \vartheta q_{t+1}k_t.$$

Using the bank balance sheet, we have

$$q_t k_t^B \leq \frac{1+r_{t+1}^{B}}{1+r_{t+1}^{B}-\vartheta\frac{q_{t+1}}{q_t}}n_t.$$

The constraint differs from the one in Equation 13 in that $1+r_{t+1}^{K}$ in the denominator is replaced by $\frac{q_{t+1}}{q_t}$. This form of collateral constraint is not nested in our framework exactly because $\frac{q_{t+1}}{q_t}$ is generally a function both returns on capital $\{r_s^K\}$ and output $\{y_s\}$. Yet, we expect the two collateral constraints to generate similar dynamics when most of the changes in $1+r_{t+1}^{K}$ are driven by capital gain $\frac{q_{t+1}}{q_t}$.

Current-value collateral constraints

An alternative form of collateral constraint assumes that liquidity supplied by the bank needs to be below the current value of capital: $d_t \leq \vartheta q_t k_t^B$. Using $d_t = q_t k_t^B - n_t$, we have

$$q_t k_t^B \leq \frac{1}{1-\vartheta}n_t.$$

This type of constraint is similar to that in Bianchi and Mendoza (2018) and behaves exactly as a regulatory constraint in Van den Heuvel (2008). See Ottonello et al. (2022) for a related discussion. In this case, $\bar{\Theta}_{r^K} = 0, \bar{\Theta}_{r^B} = 0, \gamma = 0$.

B.2 Generalization of Nested Models

GKK + Costly Leverage

Suppose that bankers in GKK solve the following problem:

$$v_{j,t}(n_{j,t}) = \max_{\Psi_{j,t}} \Lambda_{t,t+1}\big(f n_{j,t+1} + (1-f)\, v_{j,t+1}(n_{j,t+1})\big) - \Upsilon(\Psi_{j,t}) v_{j,t}(n_{j,t}),$$

$$\text{s.t.} \quad q_t k^B_{j,t} = \Psi_{j,t} n_{j,t},\; n_{j,t+1} = \big(1 + r^B_{t+1} + \left(r^K_{t+1} - r^B_{t+1}\right) \Psi_{j,t}\big) n_{j,t}.$$

In this problem, instead of assuming that the banker's leverage is constrained by their continuation value, they need to incur some reduced-form leverage cost, as in the costly leverage model.

Guess linearity $v_t(n_{j,t}) = \eta_t n_{j,t}$. The Bellman equation reduces to:

$$\eta_t = \max_{\Psi_{j,t}} \Lambda_{t,t+1}(f + (1-f)\, \eta_{t+1})\big(1 + r^B_{t+1} + \left(r^K_{t+1} - r^B_{t+1}\right) \Psi_{j,t}\big) - \Upsilon(\Psi_{j,t})\eta_t.$$

Solving the optimality condition, we can write the solution as $\Psi_{j,t} = \psi(\eta_{t+1}, r^K_{t+1}, r^B_{t+1})$ for some function $\psi(\cdot)$. Define $\Theta_t = \psi(\eta_{t+1}, r^K_{t+1}, r^B_{t+1})$. Since η_t follows a first-order difference equation with a terminal condition at $t \to \infty$, we can write $\Theta_t = \Theta(\{r^K_s, r^B_s\}_{s>t})$, and up to first order approximation, $d\Theta_t$ has the structure described in Lemma 1, and $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}$ and γ will depend on an extra parameter $\Upsilon''(\bar{\Psi})$.

Optimal Dividend Choice

Consider an individual bank solving the following optimal dividend payout problem:

$$v_{j,t}\left(n_{j,t}\right) = \max_{\delta_{j,t}} \varsigma(\delta_{j,t} n_{j,t}) + \beta v_{j,t+1}\left(n_{j,t+1}\right), \quad \text{s.t.}$$
$$q_t k^B_{j,t} = \vartheta(\delta_{j,t}) n_{j,t}, \quad n_{j,t+1} = \big(1 + r^B_{t+1} + \left(r^K_{t+1} - r^B_{t+1}\right) \vartheta(\delta_{j,t}) - \delta_{j,t}\big) n_{j,t}.$$

The bank chooses dividend payout rate $\delta_{j,t}$ and derives payoff over dividend $\delta_{j,t} n$ with utility function $\varsigma(\delta_{j,t} n) = \frac{1}{1-\gamma}(\delta_{j,t} n)^{1-\gamma}$. Function $\vartheta(\cdot)$ captures how dividend payout affects the bank's ability to leverage. One example is $\vartheta(\delta_{j,t}) = \vartheta(1 - \delta_{j,t})$, which says that the part of net worth scheduled to be paid out as dividend cannot be pledged to obtain funding. More generally, $\vartheta(\cdot)$ can capture the signaling effects of dividend payout.

Guess $v_{j,t}\left(n_{j,t}\right)=\frac{\eta_{j,t}}{1-\gamma}n_{j,t}^{1-\gamma}$, then the Bellman equation reduces to:

$$\eta_{j,t}=\max_{\delta_{j,t}}\left(\delta_{j,t}\right)^{1-\gamma}+\beta\eta_{j,t+1}\Big(\left(1+r_{t+1}^{B}\right)+\left(r_{t+1}^{K}-r_{t+1}^{B}\right)\vartheta(\delta_{j,t})-\delta_{j,t}\Big)^{1-\gamma}$$

Optimality requires:

$$\delta_{j,t}^{-\gamma}+\beta\eta_{j,t+1}\Big(\left(1+r_{t+1}^{B}\right)+\left(r_{t+1}^{K}-r_{t+1}^{B}\right)\vartheta(\delta_{j,t})-\delta_{j,t}\Big)^{-\gamma}(\vartheta'(\delta_{j,t})-1)=0.$$

Solving the optimality condition, we can write the solution as $\delta_{j,t}=\varrho(\eta_{j,t+1},r_{t+1}^{K},r_{t+1}^{B})$ for some function $\varrho(\cdot)$. Go back to the Bellman equation, we have:

$$\begin{aligned}\eta_{j,t}&=\varrho(\eta_{j,t+1},r_{t+1}^{K},r_{t+1}^{B})^{1-\gamma}\\&\quad+\beta\eta_{j,t+1}\Big(1+r_{t+1}^{B}+\left(r_{t+1}^{K}-r_{t+1}^{B}\right)\vartheta(\varrho(\eta_{j,t+1},r_{t+1}^{K},r_{t+1}^{B}))-\varrho(\eta_{j,t+1},r_{t+1}^{K},r_{t+1}^{B})\Big)^{1-\gamma}.\end{aligned}$$

Define $\Theta_t=\vartheta(\varrho(\eta_{j,t+1},r_{t+1}^{K},r_{t+1}^{B}))$. Since $\eta_{j,t}$ follows a first-order difference equation with a terminal condition at $t\to\infty$, we can write $\Theta_t=\Theta(\{r_s^K,r_s^B\}_{s>t})$ for some function $\Theta(\cdot)$, and up to first order approximation, $d\Theta_t$ has the structure described in Lemma 1, and $\bar{\Theta}_{r^K},\bar{\Theta}_{r^B}$ and γ will be controlled by γ, $\varsigma''(\bar{\delta})$, and β.

General Form

The two examples above belong to a class of problems of the form:

$$\begin{aligned}&v_{j,t}(n_{j,t})=\max_{\Theta_{j,t}}\varsigma(\Theta_{j,t},r_{t+1}^{K},r_{t+1}^{B})\times(\zeta(n_{j,t})+v_{j,t}(n_{j,t}))+\beta v_{j,t+1}(n_{j,t+1}),\\&q_tk_{j,t}^{B}=\Theta_{j,t}n_{j,t},\quad n_{j,t+1}=\Gamma(\Theta_{j,t},r_{t+1}^{K},r_{t+1}^{B})n_{j,t},\end{aligned}$$

where $\zeta(x)=\frac{1}{1-\gamma}x^{1-\gamma}$. The solution $\Theta_{j,t}$ of these problems has the structure in Lemma 1, $\Theta_{j,t}=\Theta(\{r_s^K,r_s^B\})$. Moreover, the solution does not depend on the individual $n_{j,t}$. Therefore, the individual net worth evolution $n_{j,t}$ can combined with appropriate net worth injection at the aggregate level to generate an aggregate net worth process $n_{t+1}=G(\Theta_{t-1},r_t^K,r_t^B)n_t+m$ consistent with the framework given by Equation 4.

B.3 Endogenous Equity and Dividend decision

This section shows that our formulation of the net worth process allows us to study an important class of models with endogenous equity and dividend decisions. These models are not nested by our formulation of net worth in Equation 4, however, we show that they imply liquidity supply elasticities of the same form as Proposition 1, up to a reparameterization of $\bar{G}_{r^K}$, $\bar{G}_{r^B}$, and $\bar{G}_{\Theta}$. As a result of Lemma 2, Proposition 2 and Theorem 1, this class of models is equivalent to the class of model described by Equation 3 and Equation 4 as far as aggregate responses to policies are concerned.

We consider models with endogenous equity injection studied by Karadi and Nakov (2021) and Akinci and Queralto (2022). These models solve a version of the Gertler-Karadi-Kiyotaki model augmented with optimal equity injections (equivalently, net dividend payout). We use a similar notation as in Appendix B.1. A surviving bank chooses how much equity to issue e_t, subject to a cost function $C(\xi_t)\,n_t$, where $\xi_t \coloneqq e_t/n_t$ is the ratio of equity issuance to net worth. Banks make this choice in period t before they observing $v_{j,t+1}(n_{t+1})$. The optimal equity issuance to net worth ratio solves the following problem:

$$
\begin{aligned}
v_{j,t}(n_t) &= \max_{\xi_t} \Lambda_{t,t+1}\left(f\tilde{n}_{t+1} + (1-f)\left[\mathbb{E}_t\left[v_{t+1}\left(\tilde{n}_{t+1}+\xi_t n_t\right)\right] - C(\xi_t)\,n_t\right]\right) \\
\tilde{n}_{t+1} &= \left(1+r^K_{t+1}\right) l_t - \left(1+r^B_{t+1}\right) d_t,
\end{aligned}
$$

where $\tilde{n}_{t+1}$ represents net worth resulting from bank profits and $\tilde{n}_{t+1}+\xi_t n_t$ represents total net worth at $t+1$.

The program has a linear value function: $v_t(n_t) = \eta_t n_t$ for some η_t. First order condition implies $\mathbb{E}_t[\eta_{t+1}] = C'(\xi_t)$, which we can write as

$$
\frac{e_t}{n_t} = \xi\left(\mathbb{E}_t[\eta_{t+1}]\right)
$$

where $\xi(\cdot) \coloneqq C'^{-1}(\cdot)$. Thus η_t follows

$$
\begin{aligned}
\eta_t = {} & \Lambda_{t,t+1}\left(f + (1-f)\,\eta_{t+1}\right)\left(\left(r^K_{t+1} - r^B_{t+1}\right)\theta\eta_t + \left(1+r^B_{t+1}\right)\right) \\
& + \Lambda_{t,t+1}\left(\eta_{t+1}\xi\left(\mathbb{E}_t[\eta_{t+1}]\right) - C\left(\xi\left(\mathbb{E}_t[\eta_{t+1}]\right)\right)\right).
\end{aligned}
$$

The law of motion for aggregate net worth is

$$n_{t+1} = (1-f)\left(\left(r^K_{t+1} - r^B_{t+1}\right)\theta\eta_t + \left(1 + r^B_{t+1}\right) + \xi\left(\mathbb{E}_t[\eta_{t+1}]\right)\right) n_t + m.$$

Use $\Theta_t = \theta\eta_t$ and define $\tilde{\Theta}_t \coloneqq \mathbb{E}_{t-1}[\Theta_t]$ to write the law of motion as

$$n_t = (1-f)\left(\left(r^K_t - r^B_t\right)\Theta_{t-1} + \left(1 + r^B_t\right) + \xi\left(\frac{\tilde{\Theta}_t}{\theta}\right)\right) n_{t-1} + m.$$

In perfect foresight equilibrium $\tilde{\Theta}_t = \Theta_t$ for all $t \geq 1$, but not for $t = 0$. As the law of motion depends on both Θ_{t-1} and $\tilde{\Theta}_t$, it cannot be reduced to the law of motion in Equation 4. Yet, as we now show, this does not change the structure of liquidity supply in Proposition 1.

Elasticities of Liquidity Supply

Due to endogenous equity and dividend decisions, the net worth process takes the following generalized form:

$$\begin{aligned} \Theta_t &= \Theta(\{r^K_{s+1}, r^B_{s+1}\}_{s \geq t}), \\ n_t &= H(\Theta_{t-1}, \tilde{\Theta}_t, r^K_t, r^B_t) n_{t-1} + m \end{aligned}$$

Totally differentiating the net worth process, we have

$$dn_t = \bar{H} dn_{t-1} + (\bar{H}_\Theta d\Theta_{t-1} + \bar{H}_{\tilde{\Theta}} d\tilde{\Theta}_t + \bar{H}_{r^K} dr^K_t + \bar{H}_{r^B} dr^B_t) n,$$

where $\bar{H}$ is the value of the H function evaluated at the steady state and $\bar{H}_{r^K}, \bar{H}_\Theta, \bar{H}_{\tilde{\Theta}}, \bar{H}_{r^B}$ are derivatives evaluated at the steady state

Consider changes in returns dr^K_s, dr^B_s in some period s, and let $dr^K_t = dr^B_t = 0$, $\forall t \neq s$. Moreover, define $d\check{n}_t, d\tilde{n}_t$ to be the changes in net worth due to changes in $d\Theta_{t-1}$ and $d\tilde{\Theta}_t$:

$$d\check{n}_t = \bar{H} d\check{n}_{t-1} + \bar{H}_\Theta d\Theta_{t-1} n, \quad d\tilde{n}_t = \bar{H} d\tilde{n}_{t-1} + \bar{H}_{\tilde{\Theta}} d\tilde{\Theta}_t n,$$

then

$$dn_t = d\check{n}_t + d\tilde{n}_t + \bar{H}^{t-s} \mathbf{1}_{\{s \leq t\}} (\bar{H}_{r^K} dr^K_s + \bar{H}_{r^B} dr^B_s).$$

From the proof of Proposition 1, we have

$$d\check{n}_t = \begin{cases} \bar{H}_\Theta \gamma^{s-t}\sigma(t)\bar{\Theta}_{r^K} n dr_s^K, & s > t, \\ \bar{H}_\Theta \bar{H}^{t-s}\sigma(s)\bar{\Theta}_{r^K} n dr_s^K, & s \le t. \end{cases}$$

where $\sigma(s) = \frac{1-(\gamma\bar{H})^s}{1-\gamma\bar{H}} \times \mathbf{1}_{\{s\ge 0\}}$.

As for $d\tilde{n}_t$, because $\tilde{\Theta}_t$ responds only to dr_s^K when $t+1 \le s$ and $\tilde{\Theta}_0$ is pre-determined, we have $\frac{d\tilde{\Theta}_{t-u}}{dr_s^K} = 0$, for $u \le t-s$ or $u \ge t$, and

$$d\tilde{n}_t = \begin{cases} \sum_{u=0}^{t-1} \bar{H}_{\tilde{\Theta}} \bar{H}^u \frac{d\tilde{\Theta}_{t-u}}{dr_s^K} n \; dr_s^K, & s > t+1, \\ \sum_{u=t-s+1}^{t-1} \bar{H}_{\tilde{\Theta}} \bar{H}^u \frac{d\tilde{\Theta}_{t-u}}{dr_s^K} n \; dr_s^K, & s \le t+1. \end{cases}$$

For class of intermediation frictions described in Section 3, Lemma 1 implies $\frac{d\tilde{\Theta}_{t-u}}{dr_s^K} = \gamma^{s-t+u-1}\bar{\Theta}_{r^K}, \forall t > u \ge t-s+1$. Substitution gives

$$d\tilde{n}_t = \begin{cases} \bar{H}_{\tilde{\Theta}} \gamma^{s-t-1}\sigma(t)\bar{\Theta}_{r^K} n dr_s^K, & s > t+1, \\ \bar{H}_{\tilde{\Theta}} \bar{H}^{t-s+1}\sigma(s-1)\bar{\Theta}_{r^K} n dr_s^K, & s \le t+1. \end{cases}$$

Putting the results together, and using $dD_t = d\Theta_t n + (\bar{\Theta}-1)dn_t$, we have

$$\frac{\partial \mathcal{D}_t/\partial r_s^K}{\mathcal{D}_t} = \begin{cases} \gamma^{s-t-1}\bar{\Theta}_{r^K}\left(\frac{1}{\bar{\Theta}-1} + \gamma\bar{H}_\Theta\sigma(t) + \bar{H}_{\tilde{\Theta}}\sigma(t)\right), & s > t+1 \\ \bar{\Theta}_{r^K}\left(\frac{1}{\bar{\Theta}-1} + \gamma\bar{H}_\Theta\sigma(t) + \bar{H}\bar{H}_{\tilde{\Theta}}\sigma(s-1)\right), & s = t+1 \\ \left(\bar{H}_{r^K} + \bar{\Theta}_{r^K}(\bar{H}_\Theta\sigma(s) + \bar{H}\bar{H}_{\tilde{\Theta}}\sigma(s-1))\right)\bar{H}^{t-s}, & s \le t. \end{cases}$$

Using the definition of $\sigma(s)$, we have $\sigma(s-1) = (\bar{H}\gamma)^{-1}(\sigma(s)-1))$, it follows that the liquidity supply elasticities have the identical form as in Proposition 1:

$$\frac{\partial \mathcal{D}_t/\partial r_s^K}{\mathcal{D}_t} = \begin{cases} \gamma^{s-t-1}\bar{\Theta}_{r^K}\big(\frac{1}{\bar{\Theta}-1} + \gamma\tilde{G}_\Theta\sigma(t)\big), & s > t, \\ \big(\tilde{G}_{r^K} + \bar{\Theta}_{r^K}\tilde{G}_\Theta\sigma(s)\big)\tilde{G}^{t-s}, & s \le t, \end{cases}$$

where $\tilde{G}_{r^K} := \bar{H}_{r^K} - \bar{\Theta}_{r^K}\gamma^{-1}\bar{H}_{\tilde{\Theta}}, \tilde{G}_\Theta := \bar{H}_\Theta + \gamma^{-1}\bar{H}_{\tilde{\Theta}}$, and $\tilde{G} := \bar{H}$. Similar steps result in the formula for $\frac{\partial \mathcal{D}_t/\partial r_s^B}{\mathcal{D}_t}$.

C Data, Estimation, and Calibration Moments

Data Source

For data on the banking sector, we obtain market values of bank holding companies from CRSP and link them to the Call Report data for their balance sheets. A cleaned version of the Call Report data is provided by Drechsler, Savov, and Schnabl on their website. For balance sheets of the rest of the economy, we use data from the Financial Accounts of the United States (FoF), available on FRED.

For expected returns, we obtain U.S. Treasury debt yields from the U.S. Treasury's website. For corporate bond yields, we obtain high-quality market (HQM) yields from the U.S. Treasury's website and Moody's BAA bond yields from FRED. To construct real yields, we use inflation expectations data from the Cleveland Fed.

For proxies of identified shocks, we use proxies for monetary policy shocks from Michael Bauer's website, oil shocks from Christiane Baumeister's website, and intermediary net worth shocks from Wenting Song's website.

We use data from January 1998 to December 2019 as our sample periods, during which all data are available. For our estimation, we drop the first year due to the construction of our instrumental variable, which we discuss below.

C.1 Estimation of Intermediation Frictions

C.1.1 Variable Construction

We describe how we construct variables in our estimation from our data source. We refer to variables from these datasets with their variable names.

Leverage ($d\Theta_t$):

- We use variables from the CRSP, Call Report, and FoF data to calculate the banking sector's effective leverage.
- *market value of bank equity:* For the market value of bank net worth, we use the variable "TCAP" from CRSP. We aggregate the value of all stocks with id "kypermno" under each "permco." We link the CRSP data to the Call Report data to CRSP with a cross-walk between "bhcid" and "permco."

- *liquid assets:* We include the following variables from the Call Report data: “cash,” “fedfundsrepoasset,” “securities”. Variable “securities” contains Treasury, Agency, and corporate debt. To seperate holding of Agency, and corporate debt, we use the aggregate FoF series for Private Depository Institutions to construct the following adjustment factor

$$adj_t := \frac{\text{cash} + \text{reserves} + \text{fed fund repo asset} + \text{treasury}}{\text{cash} + \text{reserves} + \text{fed fund repo asset} + \text{treasury} + \text{agency} + \text{muni}},$$

  where series ids are given by: cash - FL703025005, reserves - FL713113003, fed fund repo asset - FL702050005, treasury - LM703061105, agency -LM703061705, muni - LM703062005. We construct banks' liquid assets holdings as the sum of ‘cash,” “fedfundsrepoasset,” and “securities” from the Call Report multiplied by the adjustment factor adj_t.

- *liquid liabilities:* We include the following variables from the Call Report data: “deposits,” “foreigndep,” “fedfundsrepoliab.”

- The Call Report and FoF data are available at the quarterly frequency. We extend the measure of effective leverage, Θ_t, to the monthly frequency by interpolating quarterly observations of balance sheet items and time-aggregating daily market value of bank equity to monthly.

- *effective leverage:* We construct the effective leverage of the banking sector as

$$\Theta_t := 1 + \frac{\textit{liquid liabilities} - \textit{liquid assets}}{\textit{market value of bank equity}}.$$

- We calculate deviations of effective leverage from the steady state, $d\Theta_t$, as the deviation of effective leverage from a quadratic time trend. Figure 11 (top-left panel) shows the detrended effective leverage with the sample mean added back.

Expected returns ($\mathbb{E}_t[dr^K_{t+h}]$, $\mathbb{E}_t[dr^B_{t+h}]$):

- We use yields on Treasury debt and HQM corporate bonds, which are available daily for maturities of 6 months and 1, 2, 3, 5, 7, 10, 20, and 30 years, aggregating observations to a monthly frequency by calculating averages.

- We construct real yields by subtracting expected inflation from nominal yields, using data from the Cleveland Fed.

- We calculate spreads between HQM and Treasuries. We adjust the spreads between HQM and Treasuries with a constant factor so that at the 30-year maturity, the spread corresponds to the spread between Moody's BAA bond yields (series BAA from FRED) and Treasuries, which we think better reflects the expected returns on the banking sector's asset holdings. We obtain the adjustment factor as the coefficient from regressing the 30-year BAA-Treasury spread on the 30-year HQM-Treasury spread.

- We calculate deviations of real Treasury yields from a quadratic trend, and we add back the means. We do the same with the spreads. Figure 11 (top-right and bottom-left panels) shows the resulting yields and the spreads.

- We calculate (detrended) real yields on capital as a sum of detrended real Treasury yields and detrended spreads. Figure 11 (bottom-right panel) shows the real yields on capital.

- Finally, we use the yield curves to obtain forward rates used in our empirical specification. We extend the yields between the maturities we observe with a left-continuous step function and calculate the implied forward rates for all horizons from the yield curves. For each horizon h, we construct $\mathbb{E}_t[dr^K_{t+h}]$ and $\mathbb{E}_t[dr^B_{t+h}]$ as the deviation of h-quarters-ahead forward rates from their averages over time.

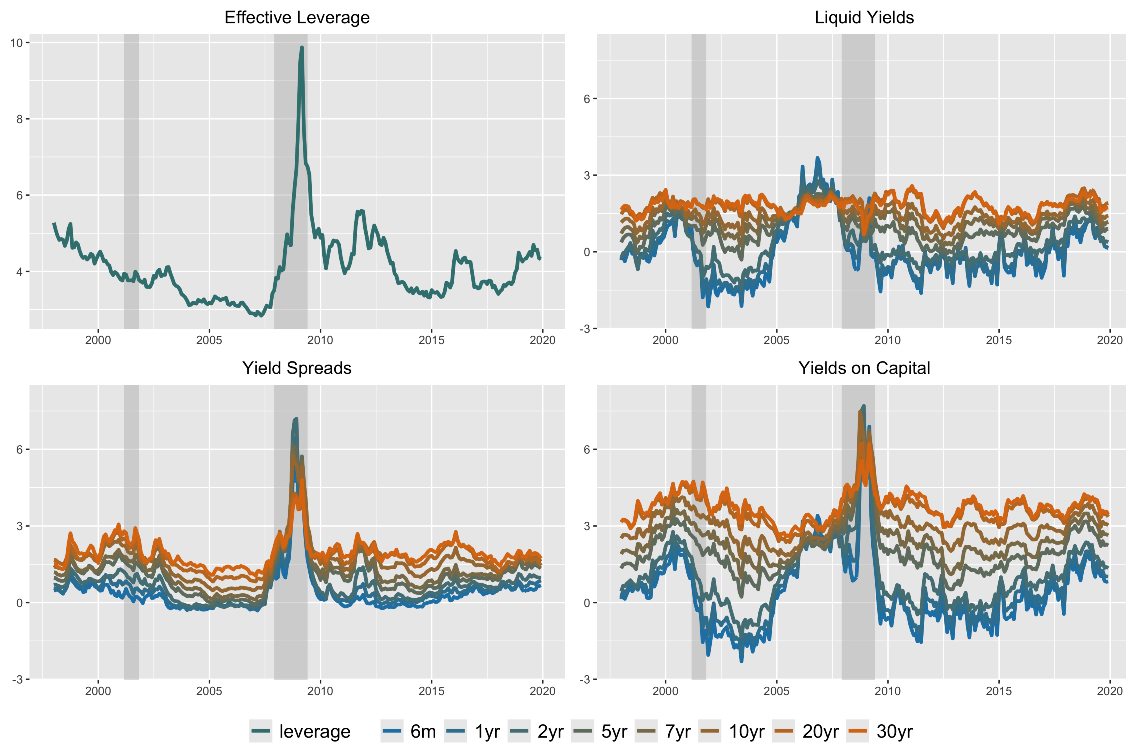

Figure 11: Top-left: Effective leverage calculated as $1+(\textit{liquid liabilities}-\textit{liquid assets})/\textit{market value of bank equity}$; top-right: real yields on Treasuries, calculated as nominal yields net of inflation expectations from the Cleveland Fed; bottom-left: the spreads between High-Quality Market (HQM) Corporate Bonds and Treasury yields at various maturities, adjusted by BAA-30yr Treasury spread. These series are detrended by subtracting a quadratic trend. Bottom-right: Real yields on capital, calculated as the sum of detrended real yields on Treasuries and spreads.

Shock proxies:

- For monetary policy shocks, we use the "MPS_ORTH" series from the "SVAR Monthly Data" constructed in Bauer and Swanson (2023); for oil shock, we use the monthly structural oil supply shocks constructed by Baumeister and Hamilton (2019); for intermediary net worth shocks, we use "finshock_broad" from Ottonello and Song (2022).

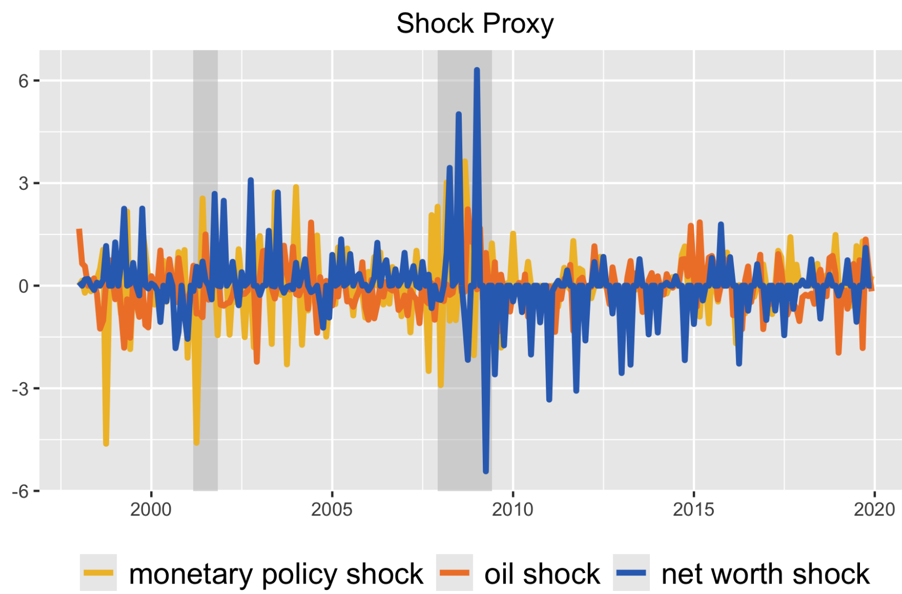

Figure 12: Shock proxies - monetary policy shock from Bauer and Swanson (2023), oil supply shock from Baumeister and Hamilton (2019), financial sector net worth shock from Ottonello and Song (2022); rescaled by standard deviation.

C.1.2 Construction of Instrumental Variables

We describe the joint co-movement of forward rates as a VAR model of order p:

$$z_t = A_0 + A_1 z_{t-1} + \ldots + A_p z_{t-p} + e_t, \quad e_t \sim N(0, \Sigma),$$

where e_t is a vector of normal zero mean i.i.d. shocks with $\Sigma = \mathbb{E}\left[e_t e_t'\right]$. $A_0, \ldots, A_p$ are matrices of appropriate dimensions and z_t is a vector that contains the observable variables: detrended forward rates $\mathbb{E}_t[dr^K_{t+h}]$ and $\mathbb{E}_t[dr^B_{t+h}]$, the log of an index of industrial production, and the three proxies: monetary policy shock proxy from Bauer and Swanson (2023), oil shock proxy from Baumeister and Hamilton (2019), and intermediary net worth shock proxy Ottonello and Song (2022). We set $p = 12$ and use forward rates $\mathbb{E}_t[dr^K_{t+h}]$ and $\mathbb{E}_t[dr^B_{t+h}]$ with $h = 1, 5, 10, 30$ years.

The reduced form residuals can be expressed as linear combination of structural uncorrelated innovations, i.e. $e_t = \Upsilon \eta_t$, where where $\Upsilon\Upsilon' = \Sigma$ and $\mathbb{E}\left[\eta_t \eta_t'\right] = I$.

Our strategy to retrieve the three structural shocks of interest (monetary, oil, and net worth) is to use a timing restriction. We assume that, controlling for all lagged data, each proxy depends on only one structural shock. It does not depend on other structural shocks or lags of the structural shock of interest. With only one proxy this approach would be the same as estimating a VAR with the proxy ordered first and using a recursive identifications scheme. Plagborg-Møller and Wolf (2021) show the equivalence between such an approach and a local projection instrumental variable estimation procedure.

Once we retrieve the three structural shocks, we do a historical decomposition of detrended forward rates. We denote their components driven by the three shocks as $\mathbb{E}_t[\check{dr}^K_{t+h}]$ and $\mathbb{E}_t[\check{dr}^B_{t+h}]$, and refer to them as the *return variations* attributable to these shocks. By construction, these return variations satisfy exclusion restriction: they are linear combinations of the three structural shocks (monetary, oil, and net worth) and thus independent of structural shocks that directly affect the relationship between leverage and returns, υ_t. We use these return variations, $\mathbb{E}_t[\check{dr}^K_{t+h}]$ and $\mathbb{E}_t[\check{dr}^B_{t+h}]$, as our instrumental variables.

To examine whether our instrumental variables satisfy the relevance condition, Table 3 shows the share of forecast error variance for detrended forward rates used in the SVAR model that can be attributed to the three structural shocks. The last row shows R^2 from linear regression of detrended forward rates $\mathbb{E}_t[dr^K_{t+h}]$ and $\mathbb{E}_t[dr^B_{t+h}]$, on their counterparts $\mathbb{E}_t[\check{dr}^K_{t+h}]$ and $\mathbb{E}_t[\check{dr}^B_{t+h}]$, one by one. The three structural shocks explain between 10% and 30% of variation in detrended forward rates.

Table 3: Forecast Error Variance Decomposition of Detrended Forward Rates

		$r^K_{t+1\text{y}}$	$r^K_{t+5\text{y}}$	$r^K_{t+10\text{y}}$	$r^K_{t+30\text{y}}$	$r^B_{t+1\text{y}}$	$r^B_{t+5\text{y}}$	$r^B_{t+10\text{y}}$	$r^B_{t+30\text{y}}$
FEVD	6m	0.06	0.05	0.04	0.08	0.08	0.05	0.07	0.08
	12m	0.22	0.34	0.26	0.15	0.10	0.07	0.11	0.11
	24m	0.22	0.33	0.28	0.21	0.11	0.11	0.12	0.12
R^2		0.23	0.23	0.14	0.11	0.19	0.23	0.22	0.21

C.1.3 Estimation

We estimate $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}$ and γ using the Generalized Method of Moments and the following moment condition:

$$\mathbb{E}\left[\left(d\Theta_t - \sum_{h=1}^{\infty} \gamma^{h-1}\left(\bar{\Theta}_{r^K}\mathbb{E}_t[dr^K_{t+h}] - \bar{\Theta}_{r^B}\mathbb{E}_t[dr^B_{t+h}]\right)\right) \times (1, I_t)^\intercal\right] = 0.$$

In the baseline specification, we have

$$I_t = \left\{\mathbb{E}_t[dr^K_{t+h}], \mathbb{E}_t[dr^B_{t+h}]\right\}_{h\{\in 1,5,10,30\}},$$

and for the IV specifications,

$$I_t = \left\{\mathbb{E}_t[d\check{r}^K_{t+h}], \mathbb{E}_t[d\check{r}^B_{t+h}]\right\}_{h\{\in 1,5,10,30\}}.$$

For the estimation result in Table 1, we use the optimal weighting matrix obtained from an iterative GMM. We use a quadratic spectral kernel to compute the covariance matrix of the vector of sample moment conditions. We use the BFGS algorithm to find the minimum of the objective function. We verify numerically that the objective function is well-behaved. To further alleviate concerns about convergence to a local minimum, we consider 100 different starting points for our estimation procedure and confirm that we obtain numerically similar results.

C.1.4 Robustness

Reliance on Specific Shocks

In order to alleviate concerns that any specific shock proxy might violate the exclusion restriction, we consider three alternative constructions of the return variations,

$$I^{(j)}_t = \left\{\mathbb{E}_t[d\check{r}^{K,(j)}_{t+h}], \mathbb{E}_t[d\check{r}^{B,(j)}_{t+h}]\right\}_{h\{\in 1,5,10,30\}}, \ \forall j = 1,2,3,$$

where for each specification j, we leave out one of the shocks proxies from Bauer and Swanson (2023), Baumeister and Hamilton (2019), and Ottonello and Song (2022) in the construction of $I^{(j)}_t$. We repeat the IV estimation with these alternative specifications and report the results in Table 4. The estimation results from these alternative specifications are similar to those from our main specification in Table 1 and suggest that our result does not rely solely on one particular shock proxy.

Table 4: Estimation with the Exclusion of Specific Shocks

	excl. oil shock	excl. net worth shock	excl. mp shock
size of cross-price, $\bar{\Theta}_{r^K}$	19.72	23.07	20.89
	(9.16)	(12.58)	(12.67)
size of own-price, $\bar{\Theta}_{r^B}$	34.49	23.87	20.43
	(16.75)	(16.87)	(15.70)
forward-looking, γ	0.94	0.95	0.96
	(0.03)	(0.04)	(0.03)
Observations	252	252	252

Note: Estimation uses iterative GMM for optimal weighting matrix. Standard errors use heteroskedastic and autocorrelation consistent estimators. Sample period: January 1999 to December 2019, monthly observation.

State-Dependency

We consider the following generalization of our empirical specification:

$$\mathbb{E}\left[\left(d\Theta_t - \sum_{h=1}^{\infty} \gamma^{h-1}(\bar{\Theta}_{r^K}\mathbb{E}_t[dr^K_{t+h}] - \bar{\Theta}_{r^B}\mathbb{E}_t[dr^B_{t+h}])f(\kappa_1 d\tilde{s}_t) - \kappa_0 d\tilde{s}_t\right) \times (1, I_t)^\intercal\right] = 0,$$

where $f(x) = \frac{2}{1+e^{-2x}}$ is a logistic function with value and slope equal to one at $x = 0$, and maps into the interval $(0, 2)$, and $\tilde{s}_t$ is a proxy for the aggregate state. Parameter κ_0 represents how much Θ_t responds to the aggregate state beyond what is captured by expected returns, and κ_1 captures the level of state-dependency in the responses to expected returns. For our estimation, we use log industrial production normalized by its standard deviation for $\tilde{s}_t$.

We consider two cases:

- *Direct response to $\tilde{s}_t$*: In this case, we set $\kappa_1 = 0$ and focus on estimating κ_0, the direct response of Θ_t to $\tilde{s}_t$. This case entertains the possibility that expected returns miss important information about the aggregate state of the economy.

- *State dependent response to expected returns*: In this case, we estimate κ_0 and κ_1 jointly, analogous to allowing an interaction term between $\tilde{s}_t$ expected returns in linear regression. We parameterize the interaction with the logistic function f to increase numerical stability. Because we have a small sample size, we assume that $f(\kappa_1 d\tilde{s}_t)$ scales responses to all returns in order to limit the number of parameters we need to estimate. A larger estimate of κ_1 (in size) implies that a first-order approximation of $\Theta(\{r^K_{t+h}, r^B_{t+h}\})$ is only useful for a smaller disturbance around the steady state, and our nesting result in Lemma 1 is more limited.

For each of the two cases, we estimate both the baseline specification and the IV specification. Table 5 reports the results.

Table 5: Direct Response to Aggregate State and State Dependency

	direct response		state dependent	
	baseline	IV	baseline	IV
size of cross-price, $\bar{\Theta}_{r^K}$	26.76	27.29	29.35	22.12
	(18.10)	(20.99)	(16.97)	(17.62)
size of own-price, $\bar{\Theta}_{r^B}$	19.84	27.76	20.52	23.07
	(15.99)	(27.04)	(16.40)	(21.70)
forward-looking, γ	0.92	0.94	0.93	0.93
	(0.07)	(0.07)	(0.04)	(0.07)
direct response, α_0	-0.30	-0.55	-0.24	-0.48
	(0.19)	(0.20)	(0.19)	(0.19)
state-dependency, α_1			-0.08	-0.18
			(0.29)	(0.15)
Observations	252	252	252	252

Note: Estimation uses iterative GMM for optimal weighting matrix. Standard errors use heteroskedastic and autocorrelation consistent estimators. Sample period: January 1999 to December 2019, monthly observation.

- The point estimates for $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}, \gamma$ remain similar to the main specification, but the standard errors are large here.
- Estimates for α_0 range from -0.25 to -0.5. This implies that a one-standard-deviation drop in output will lead to an increase in leverage by .25 to 0.5,

keeping all expected returns the same. This is not negligible, which can have two implications: One is that the class of models we study is misspecified, so we need to extend our specification of $\Theta(\cdot)$ to include output directly. Another possibility is that we are mismeasuring leverages and expected returns in the data. For example, our proxy for returns on capital, corporate bond yields, may not contain all relevant information about the banking sector's asset holdings, and this information is picked up when we include industrial production in the estimation.

- Estimates for α_1 are -0.08 to -0.18 for the two specifications. The standard errors are large, and we cannot reject the hypothesis that there is no state dependency. However, the point estimates can still provide useful information through the lens of our framework: a one-standard-deviation drop in output is similar to a $8\% - 18\%$ increase in $\bar{\Theta}_{r^K}$ and $\bar{\Theta}_{r^B}$. To the extent that higher cross-price elasticities weaken the asset market response, these point estimates suggest that the effect of asset purchases will be even weaker during periods with low economic activities.

C.2 Mapping the Model to the Data

C.2.1 Asset Classification and Balance Sheet Overview

This section describes how we consolidate balance sheets in the data to map them to those in the model. Consistent with the definition in Section C.1, we categorize liquid assets to include deposits in checkable, time, savings accounts, money market fund shares, and government liabilities, such as cash, reserve, and Treasury debt. Conceptually, our notion of liquid assets aims to include assets whose values remain relatively unaffected by trade volume or the state of the economy. Due to these attributes, these assets are useful for transactional purposes and command a premium. We do not think trading of illiquid assets necessarily involves a large transaction cost, but simply that they lack certain features we described above.

We obtain the household sector's aggregate balance sheet from the Flow of Funds. Households' liquid asset holdings mostly consist of deposits (72%) and money market funds shares (17%). We adjust the balance sheets of private depository institutions proportionally to equalize their liquid liabilities to the deposit holdings of households.

This adjustment accounts for the fact that around one-third of the banks' liquid liabilities are held by the corporate sector. We apply a similar adjustment to the money market funds, of which half is held by households. In Section C.2.2, we discuss how we can extend the model to account for the liquid assets held by the corporate sector without affecting our analysis.

Table 6: Consolidated Balance Sheets

	assets		liabilities	
households	liquid assets	0.55		
	net illiquid assets	3.43		
			equity	3.97
banks & mmf	liquid assets	0.11		
	capital	0.52		
			liquid liabilities	0.51
			equity	0.14

Note: Consolidated balance sheets of the U.S. economy through the lens of the model. Values are presented as a fraction of the U.S. GDP, averaged over the periods from 1998Q1 to 2019Q4.

Table 6 shows the consolidated balance sheets of the household sector and the corresponding balance sheets of banks and money market funds. Liquidity supplied by the financial sector (liquid liabilities issued by the financial sector minus its liquid assets holdings) amounts to around 40% of GDP and accounts for around 70% of liquid assets held by households. Table 7 paints a picture that is in contrast to a large class of heterogeneous agent models that study monetary and fiscal policies, such as Kaplan et al. (2018). These models emphasize the role of liquid assets in households' consumption-saving behavior, yet mostly abstract away from the financial sector and assume all liquid assets are supplied by the government. Our result in Section 6 shows that the financial sector's response is crucial for understanding aggregate responses to government policies.

C.2.2 Balance Sheet Details

We obtain balance sheet data from the FoF data.

Banks: We obtain the balance sheet of the banking sector following the description in Section C.1.

Money market funds:

- *liquid assets:* Liquid assets held by mmf include: checkable - FL633020000, time and savings deposits - FL633030000, foreign deposits - FL633091003, repo assets - FL632051000, and treasury - FL633061105.

- *imputed net worth*: As the money market funds hold a small part of assets that we categorize as illiquid, we split the total mmf shares (series MMMFFAQ027S from FRED) into liquid liabilities and equity and impute the net worth of mmf by assuming the same effective leverage as the banking sector:

$$\text{mmf net worth} := \frac{\text{total mmf shares - mmf liquid assets}}{\text{effective leverage}}$$

  This imputed split of the mmf balance sheet into liabilities-net worth is consistent with the difference in liquidity among mmf shares implicitly imposed by withdrawal fees for large withdrawals. We categorize mmf net worth as illiquid and compute the liquid component of the mmf shares as the difference between total mmf shares and the imputed mmf net worth.

Households:

- *liquid assets:* We include deposits in checkable (FL193020005), time and saving accounts (FL193030205), the liquid component of the money market fund shares given by $(1 - \frac{\text{mmf net worth}}{\text{total mmf shares}}) \times$household's mmf holdings (FL193034005), and households' holdings of treasury debt, calculated as the total government and municipal securities (FL193061005) net of municipal securities (LM153062005).

- *net illiquid assets*: We calculate households' net illiquid asset holdings as their total assets (FL192000005) net of liquid asset holdings defined above and their liabilities (FL194190005). Moreover, because the illiquid account in our model does not contain holdings of government debt, we further subtract from households' net illiquid asset holdings following items: the unfunded pension claims (FL223073045, FL343073045), the holdings of treasury debt through pension funds, insurance companies, mutual funds, etc.[13]

Accouting for corporate deposits:

[13]Serial numbers of variables we subtract include: LM103061103, LM113061003, LM513061105, LM543061105, LM573061105, LM343061105, LM223061143, LM653061105, LM553061103, LM563061103, LM403061105, FL673061103, LM663061105, LM733061103, and FL503061303

- The size of deposits issued by banks and money market funds exceeds the amount of deposits held by households in the data due to deposits holdings in the corporate sector. When mapping our model to the data, we rescale all balance sheet items of the banking sector and money market funds proportionally such that: (1) liquid liabilities of the money market funds are equal to those held by the households, and (2) liquid liabilities of the banking sector are equal deposits held by households and the money market funds.

- Although our model does not provide a theory of corporate deposit demand, we can extend our model to allow firms to hold the rest of the deposits issued by banks on their balance sheet inside households' illiquid accounts, assuming that firms do not use liquid assets in the production process. This assignment does not affect the consolidated balance sheet of the fund. This is because holding a combination of these deposits in the illiquid account with the corresponding net worth of banks supplying these deposits is equivalent to directly holding capital of the same value. Specifically, consider the following modification to the model: (1) the banking sector has net worth $(1+\chi)n_t$ instead of n_t, (2) the illiquid account passively holds extra deposits χd_t that correspond to the corporate deposits in the data, and (3) capital in the illiquid account is $q_t k_t^F - \chi(n_t + d_t)$ instead of k_t^F

- Let $\tilde{r}_{t+1}^A$ denote returns on illiquid assets associated with these modifications. Direct calculation shows that it is identical to the illiquid returns r_{t+1}^A in Section 2:

$$\begin{aligned}\tilde{r}_{t+1}^A &:= \frac{1}{a_t}(r_{t+1}^K(q_t k_t^F - \chi(n_t + d_t)) + r_{t+1}^N(1+\chi)n_t + r_{t+1}^B \chi d_t) \\ &= \frac{1}{a_t}(r_{t+1}^K(q_t k_t^F - \chi r_t^K q_t k_t^B) + r_{t+1}^N n_t + \chi(r_t^K q_t k_t^B - r_{t+1}^B \chi d_t) + r_{t+1}^B \chi d_t) \\ &= \frac{1}{a_t}(r_{t+1}^K q_t k_t^F + r_{t+1}^N n_t) = r_{t+1}^A.\end{aligned}$$

  Since both the goods market clearing and the liquid asset market clearing conditions are not affected, Lemma 2 implies that aggregate responses with the modifications above are identical to that from the model in Section 2.

Table 7 provides a breakdown of liquid asset positions of the household sector, the banking sector, and money market funds.

Table 7: Liquid asset positions

	liquid assets		liquid liabilities	
households	deposits	0.41		
	mmf shares	0.09		
	treasury	0.05		
banks	cash & reserves	0.03		
	fed funds and repo (net)	0.02		
	treasury	0.01		
			deposits	0.42
mmf	deposits	0.02		
	net repo	0.02		
	treasury	0.01		
			mmf shares	0.09

Note: Liquid asset positions in the U.S. economy through the lens of the model. Values are presented as a fraction of the U.S. GDP, averaged over the periods from 1998Q1 to 2019Q4.

D Nested Benchmark Models

D.1 Examples of Representative Agent Models

This section provides three examples of representative agent models nested in our framework, assuming no idiosyncratic shocks, $z_t^i \equiv 1$, no heterogeneity in preferences, and no restriction on asset holdings $\bar{a} = \bar{b} = -\infty$.

The first example corresponds to the representative agent version of our quantitative model. We show that when calibrated to match the same steady-state aggregate asset holdings and returns, the liquidity demand with respect to returns is significantly more elastic in the representative agent version in comparison to our heterogeneous agent baseline, indicating that the standard heterogeneous agent framework is likely a better starting point for modeling a household sector that is insensitive to changes in returns as observed empirically in the data.

In addition to the first example, we provide two limiting cases where the household sector has the same steady-state asset holdings and returns but features perfectly elastic and inelastic liquidity demand with respect to returns. These examples provide microfoundations for the special cases we use for illustration in Section 4.2.

D.1.1 Comparing Liquidity Demand between RA and HA Households

Consider a representative household solving the following problem:

$$\max_{c_t,a_t,b_t} \sum_{t=0}^{\infty} \beta^t [u(c_t) - \nu(h_t)], \quad \text{s.t.}$$

$$a_t + b_t + c_t + \Phi(a_t, a_{t-1}, r_t^A) = (1 + r_t^B) b_{t-1} + (1 + r_t^A) a_{t-1} + y_t^h,$$

where $y_t^h := \frac{W_t}{P_t} h_t - \mathcal{T}_t(\frac{W_t}{P_t} h_t)$ denote after tax labor income.

Optimality implies:

$$\begin{aligned}
[c_t, b_t]: &\quad u'(c_t) = \beta u'(c_{t+1})(1 + r_{t+1}^B) \\
[c_t, a_t]: &\quad u'(c_t)(1 + \Phi_1(a_t, a_{t-1}, r_t^A)) = \beta u'(c_{t+1})(1 + r_{t+1}^A - \Phi_2(a_{t+1}, a_t, r_{t+1}^A))
\end{aligned}$$

We assume c to denote steady-state consumption and similar a, b, r^A and r^B. Log deviations from the steady state are denoted by $\hat{c}_t, \hat{a}_t, \hat{b}_t$ for quantities, and deviations of returns are denoted by $\hat{r}^A$ and $\hat{r}^B$.

Define $\sigma := -u(c)''c/u'(c)$. First-order approximations of the equilibrium conditions are given by:

$$\begin{aligned}
&\sigma(\hat{c}_{t+1} - \hat{c}_t) = \frac{\hat{r}_{t+1}^B}{1 + r^B}, \\
&\sigma(\hat{c}_{t+1} - \hat{c}_t) + (\zeta_{11} + \zeta_{22})\hat{a}_t + \zeta_{12}\hat{a}_{t-1} + \zeta_{21}\hat{a}_{t+1} = (1 - \zeta_{23})\frac{\hat{r}_{t+1}^A}{1 + r^A} - \zeta_{13}\frac{\hat{r}_t^A}{1 + r^A}, \\
&\bar{a}\hat{a}_t + \bar{b}\hat{b}_t + \bar{c}\hat{c}_t + \Phi_1 a\hat{a}_t + \Phi_2 a\hat{a}_{t-1} = \\
&\qquad\qquad \bar{y}\hat{y}_t + (1 + r^B)\bar{b}\hat{b}_{t-1} + (1 + r^A)a\hat{a}_{t-1} + \bar{b}\hat{r}_t^B + (a - \Phi_3)\hat{r}_t^A,
\end{aligned}$$

where

$$\begin{aligned}
\zeta_{11} &:= \frac{a\Phi_{11}}{1 + \Phi_1}, \ \zeta_{12} := \frac{a\Phi_{12}}{1 + \Phi_1}, \ \zeta_{13} := \frac{(1 + r^A)\Phi_{13}}{1 + \Phi_1}, \\
\zeta_{21} &:= \frac{a\Phi_{21}}{1 + r^A - \Phi_2}, \ \zeta_{22} := \frac{a\Phi_{22}}{1 + r^A - \Phi_2}, \ \zeta_{23} := \frac{(1 + r^A)\Phi_{23}}{1 + r^A - \Phi_2}.
\end{aligned}$$

From the optimality condition for a_t, the path of illiquid asset holdings satisfies the following system:

$$\begin{pmatrix} \hat{a}_{t+1} \\ \hat{a}_t \end{pmatrix} = \begin{pmatrix} -\frac{\zeta_{11}+\zeta_{22}}{\zeta_{21}} & -\frac{\zeta_{12}}{\zeta_{21}} \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \hat{a}_t \\ \hat{a}_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_t^a \\ 0 \end{pmatrix},$$

where $\epsilon_t^a = \frac{1-\zeta_{23}}{\zeta_{21}} \frac{\hat{r}_{t+1}^A}{1+r^A} - \frac{\zeta_{13}}{\zeta_{21}} \frac{\hat{r}_t^A}{1+r^A} - \frac{1}{\zeta_{21}} \frac{\hat{r}_{t+1}^B}{1+r^B}$.

The characteristic polynomial of the matrix is $f(\lambda) = \lambda^2 + \frac{\zeta_{11}+\zeta_{22}}{\zeta_{21}}\lambda + \frac{\zeta_{12}}{\zeta_{21}}$. Suppose that $f(1) < 0$, then $f(0) = \frac{\zeta_{12}}{\zeta_{21}} = \frac{1+r^A-\Phi_2}{1+\Phi_1} = \frac{1}{\beta} > 1$ implies there exists an eigenvalue $\lambda_1 \in (0,1)$ and an eigenvalue $\lambda_2 > 1$. The solution of the system is given by:

$$\hat{a}_t = \lambda_1 \hat{a}_{t-1} - \sum_{s=0}^{\infty} \lambda_2^{-s-1} \epsilon_{t+s}^a.$$

Consider variations in $\{r_s^A\}$. Substituting ϵ_{t+s}^a gives

$$\hat{a}_t = \lambda_1 \hat{a}_{t-1} + \frac{\zeta_{13}\lambda_2^{-1}}{\zeta_{21}} \frac{\hat{r}_t^A}{1+r^A} + \frac{\zeta_{13}\lambda_2^{-1} - (1-\zeta_{23})}{\zeta_{21}} \sum_{u=0}^{\infty} \lambda_2^{-u-1} \frac{\hat{r}_{t+u+1}^A}{1+r^A}$$

Define

$$\vartheta_{r^A} := \frac{\zeta_{13}\lambda_2^{-1} - (1-\zeta_{23})}{\zeta_{21}(1+r^A)}, \quad g_{r^A} := \frac{\zeta_{13}\lambda_2^{-1}}{\zeta_{21}(1+r^A)}, \quad \sigma^a(t) := \frac{1-(\lambda_1\lambda_2^{-1})^t}{1-\lambda_1\lambda_2^{-1}},$$

then the solution for $\hat{a}_t$ can be expressed as

$$\hat{a}_t = \begin{cases} \lambda_2^{t-s} \vartheta_{r^A} \sigma^a(t+1) \hat{r}_s^A, & s > t, \\ \lambda_1^{t-s} \Big(g_{r^A} + \lambda_1 \lambda_2^{-1} \vartheta_{r^A} \sigma^a(s) \Big) \hat{r}_s^A, & s \le t. \end{cases}$$

The budget constraint and the optimality condition for b_t imply

$$\hat{b}_{t+1} - (2+r^B)\hat{b}_t + (1+r^B)\hat{b}_{t-1} = \epsilon_t^b$$

where $\epsilon_t^b = \frac{a-\Phi_3}{b}(\hat{r}_{t+1}^A - \hat{r}_t^A) + \frac{a}{b}(1+\Phi_1)(-\hat{a}_{t+1} + \frac{1+\beta}{\beta}\hat{a}_t - \frac{1}{\beta}\hat{a}_{t-1})$. Therefore,

$$\hat{b}_t = \hat{b}_{t-1} - \sum_{u=0}^{\infty} (1+r^B)^{-u-1} \epsilon_{t+u}^b.$$

We calibrate the representative household sector to compare it with our heterogeneous agent baseline. Given r^A from the data, we calibrate χ_1 so that the steady-state illiquid asset holding a is consistent with the data, we consider the limit where $\beta \to 1$ so that $r^B \to 0$, as in our quantitative model.

Figure 13 compares the responses of liquidity demand with respect to a change in r_s^K, taking into account how the sequence of illiquid returns $\{\hat{r}_s^A\}$ responds to r_s^K as in Figure 2. Liquidity demand responses are of orders of magnitude stronger in the representative agent model than in our heterogeneous agent baseline. This highlights a key feature of the standard heterogeneous agent framework that has not been emphasized in previous work: The inertia and insensitivity to returns among households, consistent with empirical evidence in Gabaix et al. (2024).

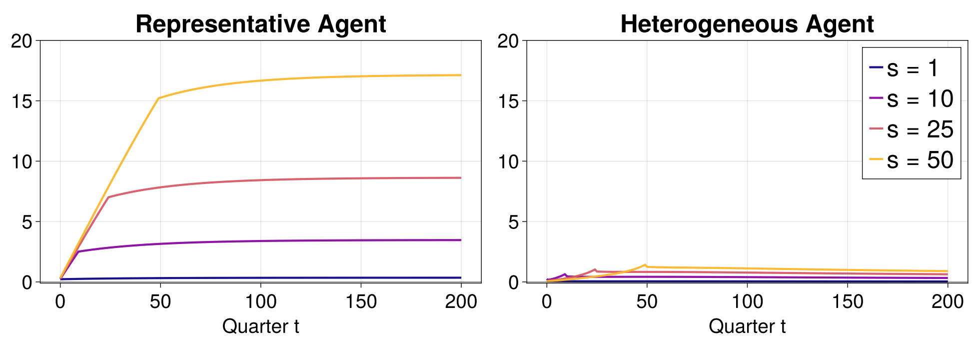

Figure 13: Entries of $-\tilde{\mathbf{B}}_{r^K}$ matrices (see Appendix A.6 for the definition)for the representative agent model with portfolio adjustment cost (left) and for the calibrated two-asset HA model (right). Each line corresponds to a different period s and shows a response of liquidity demand in quarter t with respect to r_s^K

D.1.2 Perfectly Elastic Liquidity Demand

Consider the same representative household as in Section D.1.1, except that the portfolio adjustment cost is given by a function $\tilde{\Phi}(\cdot)$ where the first derivatives are the same as function $\Phi(\cdot)$, i.e., $\tilde{\Phi}_l = \Phi_l$, but the second derivatives are scaled by a scaling parameter κ such that $\tilde{\Phi}_{lk} = \frac{1}{\kappa}\Phi_{lk}$.

Illiquid asset holding is characterized by system similar to that in Section D.1.1, featuring the same transition matrix, as $\frac{\zeta_{11}+\zeta_{22}}{\zeta_{21}}$ and $\frac{\zeta_{12}}{\zeta_{21}}$ are not affected by the scaling parameter κ. As a result, the solution of $\hat{a}_t$ is given by

$$\hat{a}_t = \begin{cases} \lambda_2^{t-s}\tilde{\vartheta}_{r^A}\sigma^a(t+1)\hat{r}_s^A, & s > t, \\ \lambda_1^{t-s}\Big(g_{r^A} + \lambda_1\lambda_2^{-1}\tilde{\vartheta}_{r^A}\sigma^a(s)\Big)\hat{r}_s^A, & s \le t, \end{cases}$$

where $\tilde{\vartheta}_{r^A} := \frac{\zeta_{13}\lambda_2^{-1}-(\kappa-\zeta_{23})}{\zeta_{21}(1+r^A)}$, and $\lambda_1, \lambda_2, \sigma^a(\cdot)$ are identical to that in Section D.1.1.

For κ large, we have $\hat{a}_t \approx \kappa\hat{a}_t^\infty$, where

$$\hat{a}_t^\infty = \begin{cases} \lambda_2^{t-s}\vartheta_{r^A}^\infty\sigma^a(t+1)\hat{r}_s^A, & s > t, \\ \lambda_1^{t-s}\Big(\lambda_1\lambda_2^{-1}\vartheta_{r^A}^\infty\sigma^a(s)\Big)\hat{r}_s^A, & s \le t, \end{cases}$$

where $\vartheta_{r^A}^\infty := \frac{-1}{\zeta_{21}(1+r^A)}$. For liquid assets, we have $\hat{b}_t \approx \kappa\hat{b}_t^\infty$, where

$$\hat{b}_t^\infty = \hat{b}_{t-1}^\infty - \sum_{u=0}^{\infty}(1+r^B)^{-u-1}\epsilon_{t+u}^b,$$

and $\epsilon_t^b = \frac{a}{b}(1+\Phi_1)(-\hat{a}_{t+1}^\infty + \frac{1+\beta}{\beta}\hat{a}_t^\infty - \frac{1}{\beta}\hat{a}_{t-1}^\infty)$. As a result, given any $\boldsymbol{D}_{r^K}$, we have $\boldsymbol{\epsilon}_{r^K}^{-1} = (\boldsymbol{D}_{r^K} - \boldsymbol{B}_{r^K})^{-1} \to \mathbf{0}$ as $\kappa \to \infty$.

D.1.3 Perfectly Inelastic Liquidity Demand

For an example of perfectly inelastic liquidity demand, we consider a representative household with a reduced-form preference over liquid assets. The household solves the following problem:

$$\max_{c_t, a_t, b_t} \sum_{t=0}^{\infty}\beta^t(u(c_t) + v(b_t)), \quad \text{s.t}$$

$$b_t + c_t + a_t = y_t + (1+r_t^B)b_{t-1} + (1+r_t^A)a_{t-1}.$$

Optimality requires:

$$\begin{aligned} [c_t, b_t]: &\quad u'(c_t) - v'(b_t) = \beta u'(c_{t+1})(1+r_{t+1}^B) \\ [c_t, a_t]: &\quad u'(c_t) = \beta u'(c_{t+1})(1+r_{t+1}^A). \end{aligned}$$

Let $\nu = -v''b/v'$ and $\psi = v'/u'$. First-order approximations of the equilibrium conditions are given by:

$$
\begin{aligned}
&(1-\psi)\sigma\hat{c}_{t+1} - \sigma\hat{c}_t + \psi\nu\hat{b}_t = (1-\psi)\frac{\hat{r}^B_{t+1}}{1+r^B}\\
&\sigma\hat{c}_{t+1} - \sigma\hat{c}_t = \frac{\hat{r}^A_{t+1}}{1+r^A}\\
&\bar{b}\hat{b}_t + \bar{c}\hat{c}_t + a\hat{a}_t - (1+r^B)\bar{b}\hat{b}_{t-1} - (1+r^A)a\hat{a}_{t-1} = \bar{y}\hat{y}_t + \bar{b}\hat{r}^B_t + a\hat{r}^A_t
\end{aligned}
$$

From the two optimality conditions:

$$
\hat{b}_t = \sigma\nu^{-1}\hat{c}_t + \epsilon^b_t,
$$

where $\epsilon^b_t = \nu^{-1}\psi^{-1}(1-\psi)\big(\frac{\hat{r}^B_{t+1}}{1+r^B} - \frac{\hat{r}^A_{t+1}}{1+r^A}\big)$.

Define $w\hat{w}_t \coloneqq (1+r^B)\bar{b}\hat{b}_t + (1+r^A)a\hat{a}_t$, where $w \coloneqq (1+r^B)\bar{b} + (1+r^A)a$. The budget constraint becomes:

$$
\bar{b}\hat{b}_t + \bar{c}\hat{c}_t + a\hat{a}_t - w\hat{w}_{t-1} = \bar{y}\hat{y}_t + \bar{b}\hat{r}^B_t + a\hat{r}^A_t.
$$

Use the express for $\hat{b}_t$ and the budget constraint to write $\hat{w}_t$ as:

$$
\begin{aligned}
w\hat{w}_t = {}& (1+r^B)\bar{b}(\sigma\nu^{-1}\hat{c}_t + \epsilon^b_t)\\
&+ (1+r^A)\big(w\hat{w}_{t-1} - (\bar{b}\sigma\nu^{-1} + \bar{c})\hat{c}_t + \bar{b}\sigma\nu^{-1}\epsilon^b_t + \bar{y}\hat{y}_t + \bar{b}\hat{r}^B_t + a\hat{r}^A_t\big).
\end{aligned}
$$

Together with optimality of a_t, we have

$$
\begin{pmatrix}\hat{c}_{t+1}\\ \hat{w}_t\end{pmatrix} = \begin{pmatrix}1 & 0\\ -w^{-1}((1+r^A)\bar{c} + \bar{b}(r^A - r^B)\sigma\nu^{-1}) & 1+r^A\end{pmatrix}\begin{pmatrix}\hat{c}_t\\ \hat{w}_{t-1}\end{pmatrix} + \begin{pmatrix}\epsilon^c_t\\ \epsilon^w_t\end{pmatrix},
$$

where $\epsilon^c_t = \sigma^{-1}\frac{\hat{r}^A_{t+1}}{1+r^A}$ and $\epsilon^w_t = w^{-1}\big((1+r^A)(\bar{y}\hat{y}_t + \bar{b}\hat{r}^B_t + a\hat{r}^A_t) - \bar{b}(r^A - r^B)\epsilon^b_t\big)$.

Consider the limit where $\nu \to \infty$. In this case,

$$
\begin{pmatrix}\hat{c}_{t+1}\\ \hat{w}_t\end{pmatrix} = \begin{pmatrix}1 & 0\\ -w^{-1}(1+r^A)\bar{c} & 1+r^A\end{pmatrix}\begin{pmatrix}\hat{c}_t\\ \hat{w}_{t-1}\end{pmatrix} + \begin{pmatrix}\epsilon^c_t\\ \epsilon^w_t\end{pmatrix},
$$

and $\hat{b}_t = \sigma\nu^{-1}\hat{c}_t + \epsilon^b_t \to 0$ as $\epsilon^b_t = \nu^{-1}\psi^{-1}\big(\frac{\hat{r}^B_{t+1}}{1+r^B} - (1-\psi)\frac{\hat{r}^A_{t+1}}{1+r^A}\big) \to 0$. Therefore, we have $\mathcal{B}_{r^A} \to \mathbf{0}$ as $\nu \to \infty$.

D.2 Alternative Form of Nominal Rigidities

We now show our results generalize to a setup with price rigidities instead of wage rigidities. We keep the labor union structure and set $\kappa_W = 0$: individual households still take labor supply decisions as given, but the unions face no cost in adjusting nominal wages. We modify the production side of the economy by introducing two types of firms: a continuum of intermediate goods producers indexed by $u \in [0, 1]$ and a representative final goods producer.

The representative final goods producer aggregates intermediate goods $\{y_{u,t}\}$ according to

$$y_t = \left(\int y_{u,t}^{\frac{\varepsilon_P - 1}{\varepsilon_P}} du \right)^{\frac{\varepsilon_P}{\varepsilon_P - 1}}.$$

Given the price of the final good $\{P_t\}$ and prices of intermediate goods $\{P_{u,t}\}$, the final good producer chooses $\{y_{u,t}\}$ to maximize profits:

$$\max_{\{y_{u,t}\}} P_t y_t - \int P_{u,t} y_{u,t} d\ell.$$

The solution of the problem implies a standard demand curve $y_{u,t} = \left(\frac{P_{u,t}}{P_t}\right)^{-\varepsilon_P} y_t$, where the price index P_t satisfies $P_t = \left(\int P_{u,t}^{1-\varepsilon_P} du\right)^{\frac{1}{1-\varepsilon_P}}$.

Intermediate goods producer of goods u uses capital, $k_{u,t-1}$, and labor supplied by unions, $h_{u,\ell,t}$, to produce $y_{u,t}$:

$$y_{u,t} = k_{u,t-1}^{\alpha} h_{u,t}^{1-\alpha}, \quad h_{u,t} = \left(\int h_{u,\ell,t}^{\frac{\varepsilon_W - 1}{\varepsilon_W}} d\ell \right)^{\frac{\varepsilon_W}{\varepsilon_W - 1}},$$

where $\varepsilon_W > 1$ is the elasticity of substitution between labor types. Let W_t[14] denote the ideal wage index associated with the cost-minimizing labor mix $h_{u,\ell,t}$ given $h_{u,t}$.

Given the demand for $y_{u,t}$, capital rental rate R_t, nominal wages W_t, initial price $P_{u,-1} = P_{-1}$, each intermediate good producer u rent capital $k_{u,t-1}$, hire labor $h_{u,t}$, and sets nominal price growth $\pi_{P,u,t} := \frac{P_{u,t}}{P_{u,t-1}} - 1$ to maximize real profits, subject to a price adjustment cost:

$$\max_{\pi_{P,u,t}, k_{u,t-1}, h_{u,t}} \sum_{t=0}^{\infty} \frac{\Lambda_{0,t}^u}{P_t} \left[P_{u,t} y_{u,t} - R_t k_{u,t-1} - W_t h_{u,t} - \frac{\kappa_P}{2} \pi_{P,u,t}^2 P_t y_t \right],$$

[14] W_t is the same for all retailers u because they use the same CES aggregator function.

where $\Lambda_{0,t}^{u}$ is some arbitrary discount factor, which, for example, can be a function of $\{r_t^K, r_t^B, y_t\}$. Price adjustment cost is in units of utility and does not affect the resource constraint; $\kappa_P > 0$ parameterizes the level of nominal rigidity. The symmetry between retailers implies they all choose the same price $P_{u,t} = P_t$, produce the same quantity $y_{u,t} = y_t$, and use the same production factors $k_{u,t} = k_t$, $h_{u,t} = h_t$.

Let μ_t denote the markup over marginal cost. In equilibrium, the total real profit from the intermediate goods producers in each period equals to $\left(1 - \mu_t^{-1}\right) y_t$ and factor prices satisfy:

$$\frac{R_t}{P_t} = \frac{\alpha}{\mu_t}\frac{y_t}{k_{t-1}}, \quad \frac{W_t}{P_t} = \frac{1-\alpha}{\mu_t}\frac{y_t}{h_t}.$$

We assume that the fraction α of profits is distributed to capital and the remaining fraction $1-\alpha$ to labor. Given this assumption, holding one unit of capital from period t to $t+1$ earns a return

$$1 + r_{t+1}^K = \max_{\iota_{t+1}} \frac{R_{t+1}/P_{t+1} + \alpha\left(1 - \mu_{t+1}^{-1}\right) y_{t+1}/k_t + q_{t+1}\left(1 + \Gamma\left(\iota_{t+1}\right) - \delta\right) - \iota_{t+1}}{q_t}.$$

Since

$$R_{t+1}/P_{t+1} + \alpha\left(1 - \mu_{t+1}^{-1}\right) y_{t+1}/k_t = \alpha y_{t+1}/k_t,$$

the return on capital above corresponds to Equation 1 in the main text. Together with the law of motion for capital and the first order condition with respect to ι_{t+1}, we obtain the same aggregate investment function $\mathcal{X}$ as in the main text.

We assume the remaining fraction $1 - \alpha$ of profits is distributed to households proportionally to $z_{i,t}h_{i,t}/h_t$. We use $y_{i,t}$ to denote real labor income of the household i plus profits it receives from the intermediate goods producers:

$$y_{i,t} = z_{i,t}\frac{W_t}{P_t}h_{i,t} + \alpha\left(1 - \mu_t^{-1}\right)\frac{y_t}{h_t}z_{i,t}h_{i,t}$$

Because labor unions are identical, $h_{l,t} = h_t$, the labor demand rule implies $h_{i,t} = l(z_{i,t})h_t$, and since $\frac{W_t}{P_t} + \left(1 - \mu_t^{-1}\right)\frac{y_t}{h_t} = (1-\alpha)\, y_t$, we have $y_{i,t} = z_{i,t}l(z_{i,t})(1-\alpha)y_t$. This corresponds to pre-tax labor income in our baseline model. Households in this version of the model face exactly the same problem as in the baseline model, which means that the aggregate functions $\mathcal{A}, \mathcal{B}, \mathcal{C}$ are the same.

D.3 Connection to KMV (2018), ARS (2023)

Kaplan, Moll, Violante (2018)

We describe how our framework nests Kaplan et al. (2018). We focus on the case with no firms' profits and $a_t^G = 0$,[15] In the two-asset HANK model of Kaplan et al. (2018) government debt is the only liquid asset therefore the liquid asset market clearing condition is $\int b_{i,t} di = b_t^G$. There is no liquidity supply of the financial sector $d_t = 0$. All capital is held through illiquid assets, $\int a_{i,t} di = q_t k_t$. The rate of return on illiquid assets equals the rate of return on capital. Because $d_t = 0$, this is consistent with our equation 2.

To ensure that $d_t = 0$ in all periods, it is enough to have $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B} = 0$ and the steady state effective leverage $\bar{\Theta}$ equal to 1. Intuitively, it does not matter whether capital is held directly as k^F or indirectly through banks as k^B, because an extra unit of net worth allows increasing bank capital holdings one-to-one.

In our quantitative study in Section 6 we follow a different strategy. We want to keep the steady state the same for all models to isolate the role of liquidity supply elasticities. This would not be possible with $d_t = 0$. We set the matrices $\mathbf{D}_{r^K}, \mathbf{D}_{r^B}, \mathbf{D}_y$ to be identically zero. This can be done by assuming $\bar{G} = \bar{G}_{\Theta} = \bar{G}_{r^K} = \bar{G}_{r^B} = 0$, and $m_t = m$ and setting $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B} = 0$. These assumptions imply that d_t is constant.

Auclert, Rognlie, Straub (2023)

We show how our work relates to Auclert et al. (2023). First, we demonstrate that our framework with $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B} \to \infty$ implies the same relationship between the rate of return on capital, r_t^K, and the real rate of return on assets as in the model with capital in Section 7.3 of Auclert et al. (2023).

Denote the rate used in the firm's problem in Auclert et al. (2023) (equation 37, on page 35) by r_{t+1}^{IKC}. Assume perfect competition among firms and the law of motion for capital is $k_t = (1 - \delta + \Gamma(\iota_t)) k_{t-1}$, where $\iota_t := x_t / k_{t-1}$. Given these assumptions,[16]

[15] In Kaplan et al. (2018) there is monopolistic competition in the goods market and price rigidities. We abstract from these because our framework features neither of them. The argument remains the same if we enrich our framework with these features.

[16] We make these assumptions to simplify the exposition. The argument remains the same with monopolistic competition and sticky prices (if we modify the firm's problem in our framework) and with alternative capital adjustment costs assumed in Auclert et al. (2023).

the firms' problem is

$$J_t\left(k_{t-1}\right)=\max_{k_t,h_t} F\left(k_{t-1},h_t\right)-\frac{W_t}{P_t}n_t-x_t+\frac{1}{1+r_{t+1}^{IKC}}J_{t+1}\left(\left(1-\delta+\Gamma\left(\frac{i_t}{k_{t-1}}\right)\right)k_{t-1}\right),$$

where $J_t(k_{t-1})$ stands for the value of the firm and $F(k_{t-1},h_t)=k_{t-1}^{\alpha}h_t^{\alpha}$.

The first order condition with respect to x_t and the envelope condition are

$$1=\frac{1}{1+r_{t+1}^{IKC}}J_{t+1}'\left(k_t\right)\Gamma'\left(\iota_t\right),$$
$$J_t'\left(k_{t-1}\right)=F_k\left(k_{t-1},h_t\right)+\frac{1}{1+r_{t+1}^{IKC}}J_{t+1}'\left(k_t\right)\left(-\Gamma'\left(\iota_t\right)\iota_t+\left(1-\delta+\Gamma\left(\iota_t\right)\right)\right).$$

Define $q_t:=\frac{1}{1+r_{t+1}^{IKC}}J_{t+1}'\left(k_t\right)$ and use the first-order condition $1=q_t\Gamma'\left(\iota_t\right)$ to write

$$q_{t-1}\left(1+r_t^{IKC}\right)=F_k\left(k_{t-1},h_t\right)-\iota_t+q_t\left(1-\delta+\Gamma\left(\iota_t\right)\right).$$

After rearranging, we obtain

$$1+r_t^{IKC}=\frac{F_k\left(k_{t-1},h_t\right)-\iota_t+q_t\left(1-\delta+\Gamma\left(\iota_t\right)\right)}{q_{t-1}}.$$

The above formula is exactly the same expression as Equation 1 for r_t^K and shows that r_t^{IKC} corresponds to r_t^K.

In one-account models in Section 4.1 and Section 4.2 of Auclert et al. (2023), the rate of return on assets is equal to r_t^{IKC}. In the two-account model in Section 4.3 the rate of return associated with the illiquid account (denote it by r_t^A, as in our framework) is equal to r_t^{IKC}, and the rate of return on the liquid account (denote it by r_t^B, as in our framework) is given by $(1-\zeta)(1+r_t^{IKC})-1$, where ζ is a constant. Regardless of whether monetary policy controls the rate of return on liquid or illiquid accounts, there is a tight link between r_t^B, the real rate controlled by the central bank (denote it by r_t), and r_t^{IKC}. More specifically, for all $t\geq 0$ we have

$$dr_{t+1}^{IKC}=\frac{1}{1-\zeta}dr_{t+1}^{B}.$$

The relationship between returns is independent of any shifts in excess liquidity. In Proposition 2, we show that relationship results from the limiting case where $\bar{\Theta}_{r^K},\bar{\Theta}_{r^B}\to\infty$ and $\bar{\Theta}_{r^B}/\bar{\Theta}_{r^K}\to 1/(1-\zeta)$.

Next, we show additional conditions, under which aggregate responses to macroeconomic policies are exactly the same in our work and a two-account model of Auclert et al. (2023). For simplicity, we set $a_t^G = 0$ in all periods. Auclert et al. (2023) assume that households have access to two accounts: liquid and illiquid. Both accounts consist of equity and bond holdings. Household i holds a share $\varpi_{i,t}^a$ of illiquid assets and a share $\varpi_{i,t}^b$ of liquid assets in equity. Our framework corresponds to $\varpi_{i,t}^a = 1$ and $\varpi_{i,t}^b = 1 - \frac{b_t^G}{\int b_{i,t} di}$ so that the share of liquid assets invested in equity corresponds to one minus the ratio of government debt sector to total liquidity supply. Households can change their illiquid account position with probability p every period, otherwise $a_{i,t} = (1 + r_t^A) a_{i,t-1}$. We can capture it by having $\Psi_{i,t} = 0$ with probability p and with probability $1-p$: $\Psi_{i,t} = 0$ if $a_{i,t} = (1 + r_t^A) a_{i,t-1}$ and $\Psi_{i,t} = \infty$ if $a_{i,t} \neq (1 + r_t^A) a_{i,t-1}$.

In Auclert et al. (2023):

1. Rates of returns satisfy $1 + r_{t+1}^K = \frac{1}{1-\zeta}(1 + r_{t+1}^B) = 1 + r_{t+1}^A \quad \forall t \geq 0$.
2. Servicing one unit of government debt (in time t goods) issued at time t costs $(1 + r_{t+1}^B)/(1-\zeta)$ units of goods in period $t+1$.
3. The goods market clearing requires $c_t + x_t + g_t + \frac{\zeta}{1-\zeta}(1 + r_t^B) \int b_{i,t-1} di = y_t$.

The first part of the first condition is satisfied for $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B} \to \infty$ and $\bar{\Theta}_{r^B}/\bar{\Theta}_{r^K} \to 1/(1-\zeta)$. Equation 2 states that the second part of the condition cannot hold unless $d_t = 0$ in all periods. This is a key difference between our framework and Auclert et al. (2023). In our framework, *assets* (capital, deposits, government debt) are associated with different returns. The returns received by households on their *accounts* depend on the composition of assets in their liquid and illiquid accounts. In Auclert et al. (2023), all *assets* pay the same return. The returns received by households on their *accounts* differ only because of financial intermediation costs. The following modification of our framework ensures $r_{t+1}^A = r_{t+1}^K$ even with $d_t > 0$. Assume that the passive mutual fund holding capital directly and bank equity has intermediation cost

$$\mu_{t+1} = \left(1 + r_{t+1}^B\right) \frac{\zeta}{1-\zeta} \frac{d_t}{a_t}$$

per unit of illiquid assets a_t. This cost is paid in final goods. Zero profit condition of the fund implies $r_{t+1}^A = r_{t+1}^K$.

The second condition is satisfied if we assume that the government needs to incur extra cost equal to $\mu_t^G = \frac{\zeta}{1-\zeta}(1+r_t^B)$ per unit of debt. The budget constraint of the government becomes

$$b_t^G = g_t + (1+r_t^B)b_{t-1}^G + \mu_t^G b_{t-1}^G - T_t.$$

The sum of intermediation costs in period t is

$$\mu_t^G b_{t-1}^G + \mu_t a_{t-1} = \frac{\zeta}{1-\zeta}\left(d_{t-1} + b_{t-1}^G\right) = \frac{\zeta}{1-\zeta}\int b_{i,t-1} di$$

and this ensures that the goods market condition in our framework is as in Auclert et al. (2023). Because the household and production sides of our economy are exactly the same, and the rates of return satisfy the same restrictions as in Auclert et al. (2023), output responses must be the same.

E Quantitative Appendix

E.1 Government Spending Multiplier: Additional Results

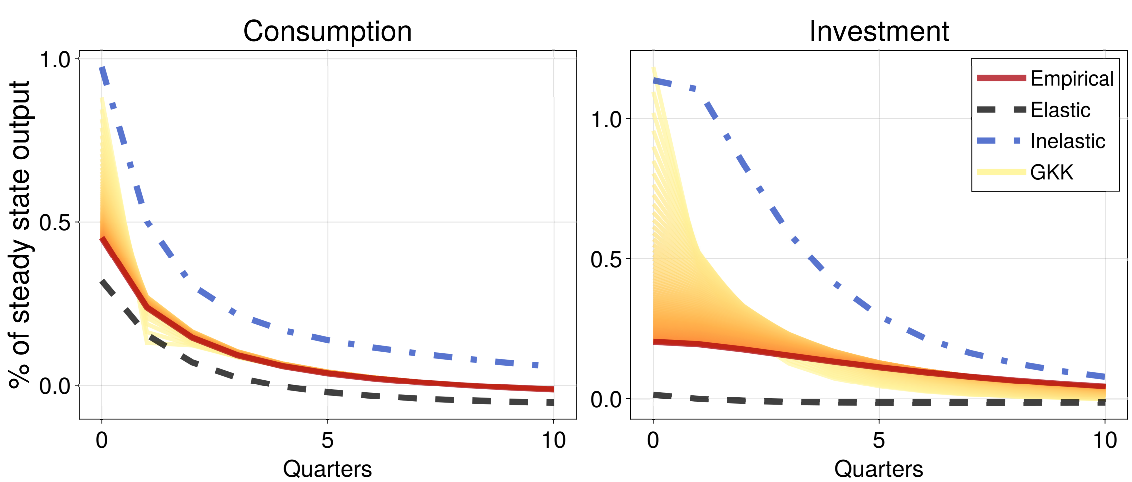

Figure 14: Consumption and investment response to government purchases with $\eta = 0.5, \rho_{b^G} = 0.95$. Red: empirical estimate ($\bar{\Theta}_{r^K} = 25$). Black: perfectly elastic liquidity supply ($\bar{\Theta}_{r^K} \to \infty$). Blue: perfectly inelastic liquidity supply ($\mathcal{D}_{r^K} = 0$). Dark to light yellow: high to low $\bar{\Theta}_{r^K}$ (from $\bar{\Theta}_{r^K} = 25$ to $\bar{\Theta}_{r^K} = 12$ in GKK).

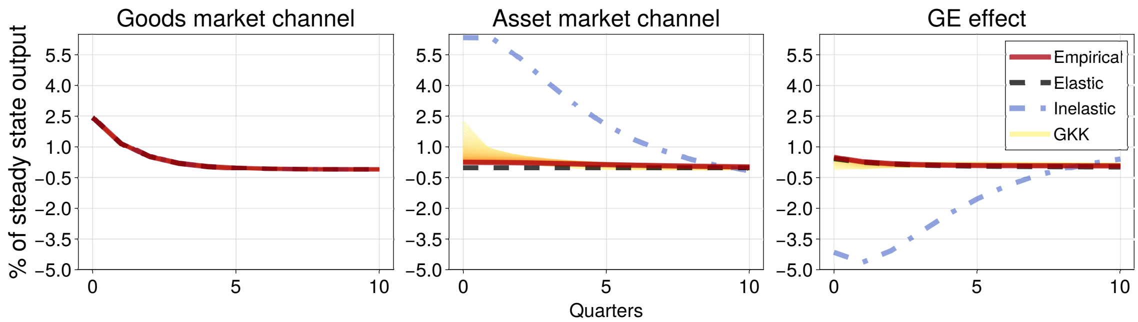

Figure 15: Decomposition of output responses to government purchases with $\eta = 0.5, \rho_{b^G} = 0.95$. Red: empirical estimate ($\bar{\Theta}_{r^K} = 25$). Black: perfectly elastic liquidity supply ($\bar{\Theta}_{r^K} \to \infty$). Blue: perfectly inelastic liquidity supply ($\mathcal{D}_{r^K} = 0$). Dark to light yellow: high to low $\bar{\Theta}_{r^K}$ (from $\bar{\Theta}_{r^K} = 25$ to $\bar{\Theta}_{r^K} = 12$ in GKK).

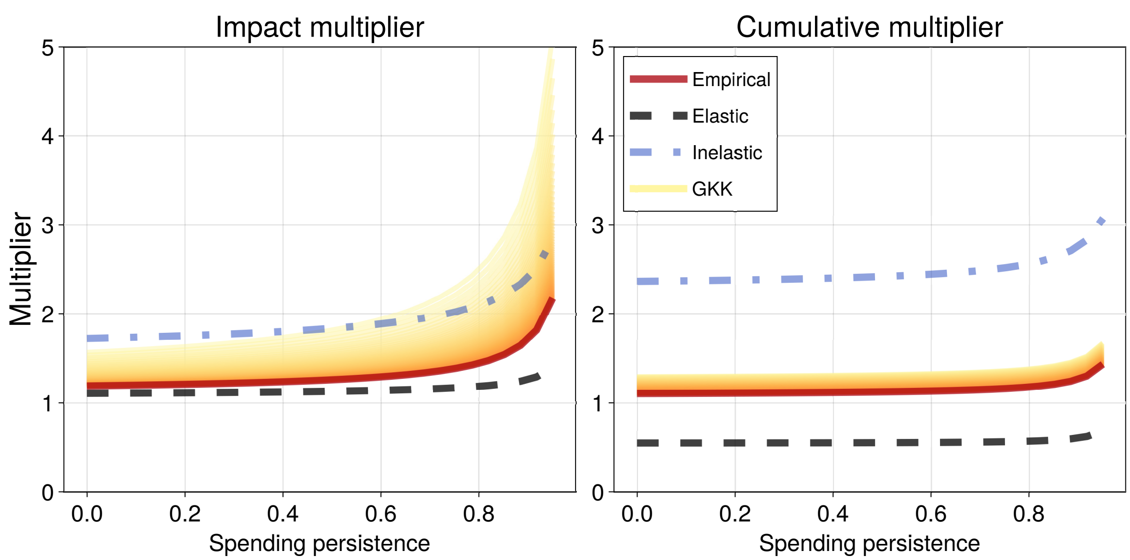

Figure 16: Impact and cumulative government spending multipliers for $\rho_{b^G} = 0.5$ and $\eta \in [0, 0.95]$ Impact multiplier: dy_0/dg_0. Cumulative multiplier: $\sum_{t=0}^{\infty}(1+r^B)^{-t}dy_t/\sum_{t=0}^{\infty}(1+r^B)^{-t}dg_t$. Red: empirical estimate ($\bar{\Theta}_{r^K} = 25$). Black: perfectly elastic liquidity supply ($\bar{\Theta}_{r^K} \to \infty$). Blue: perfectly inelastic liquidity supply ($\mathcal{D}_{r^K} = 0$). Dark to light yellow: high to low $\bar{\Theta}_{r^K}$ (from $\bar{\Theta}_{r^K} = 25$ to $\bar{\Theta}_{r^K} = 12$ in GKK).

E.2 Asset Purchases and Tax Cuts: Additional Results

Figure 17 shows responses of consumption and investment. When the financial sector's liquidity supply has low elasticities with respect to $d\boldsymbol{r}^K$, responses of both consumption and investment are amplified. It demonstrates that differences in the output response seen in Figure 10 are driven mostly by differences in investment. The magnitude of responses of consumption also depends on the cross-price elasticities of liquidity supply but to a much smaller extent.

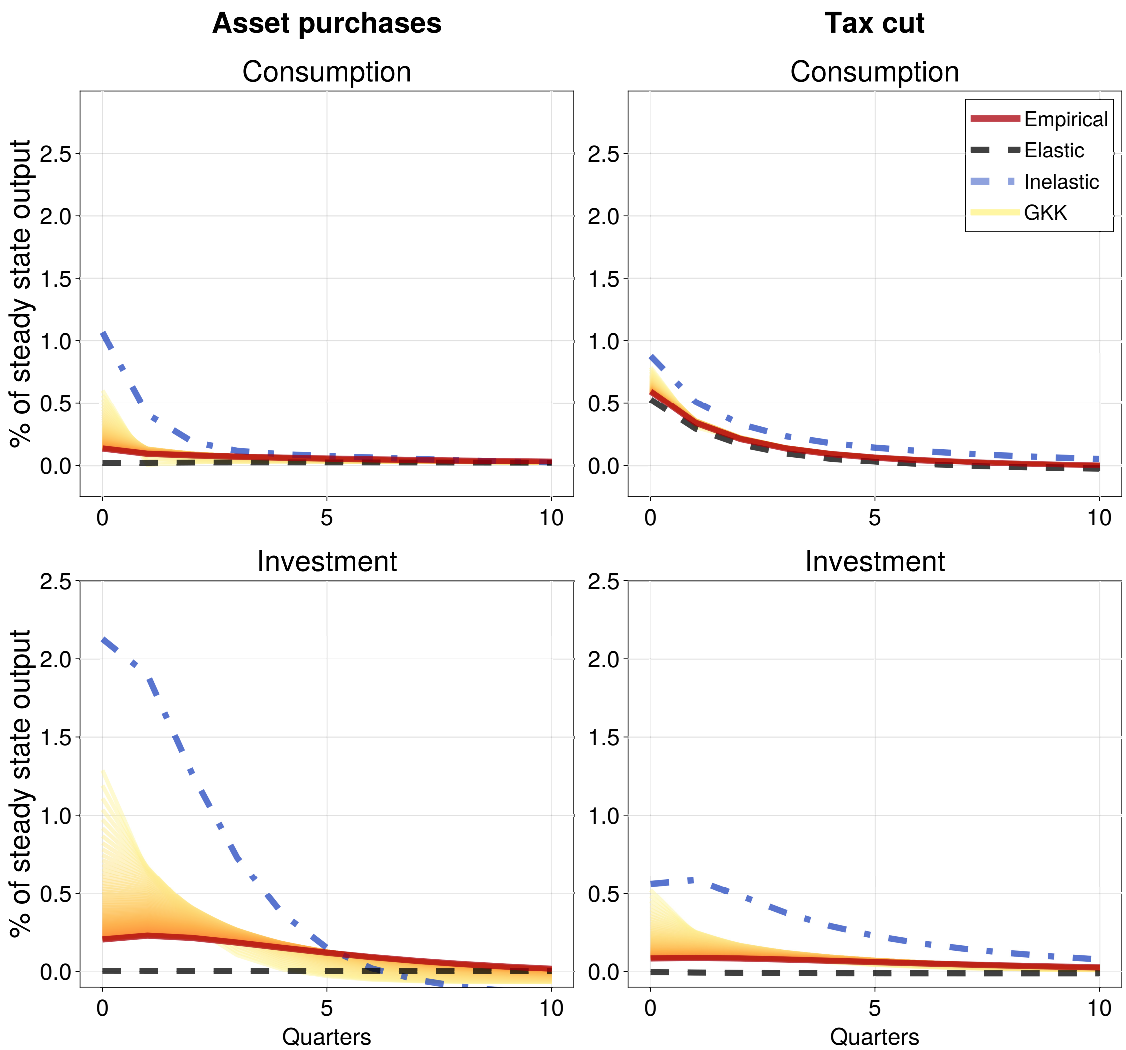

Figure 17: Consumption and investment responses to asset purchases and tax cuts with $\eta = 0.5, \rho_{b^G} = 0.95$. Red: empirical estimate ($\bar{\Theta}_{r^K} = 25$). Black: perfectly elastic liquidity supply ($\bar{\Theta}_{r^K} \to \infty$). Blue: perfectly inelastic liquidity supply ($\mathcal{D}_{r^K} = 0$). Dark to light yellow: high to low $\bar{\Theta}_{r^K}$ (from $\bar{\Theta}_{r^K} = 25$ to $\bar{\Theta}_{r^K} = 12$ in GKK).

The large differences in investment responses are due to firms' responses to capital prices. Figure 18 shows that the range of responses of capital prices is much larger for asset purchases. This is consistent with the important role of the asset market channel.

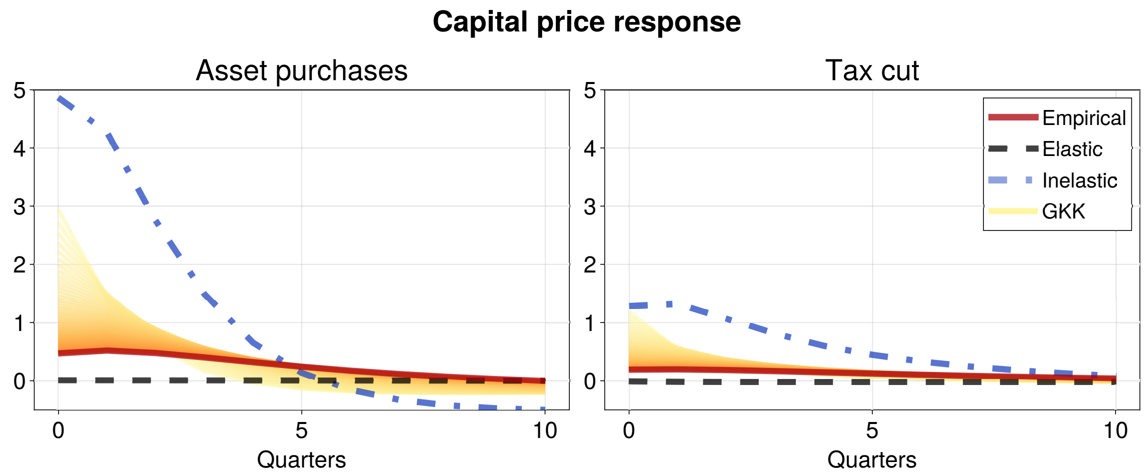

Figure 18: Capital price response to asset purchases and tax cuts with $\eta = 0.5, \rho_{b^G} = 0.95$. Red: empirical estimate ($\bar{\Theta}_{r^K} = 25$). Black: perfectly elastic liquidity supply ($\bar{\Theta}_{r^K} \to \infty$). Blue: perfectly inelastic liquidity supply ($\mathcal{D}_{r^K} = 0$). Dark to light yellow: high to low $\bar{\Theta}_{r^K}$ (from $\bar{\Theta}_{r^K} = 25$ to $\bar{\Theta}_{r^K} = 12$ in GKK).

Figures 19 and 20 below show output responses and decomposition as in Figures 8 and 9 in the main text, but with the y-axis rescaled to show the large magnitude of responses with perfectly inelastic liquidity supply.

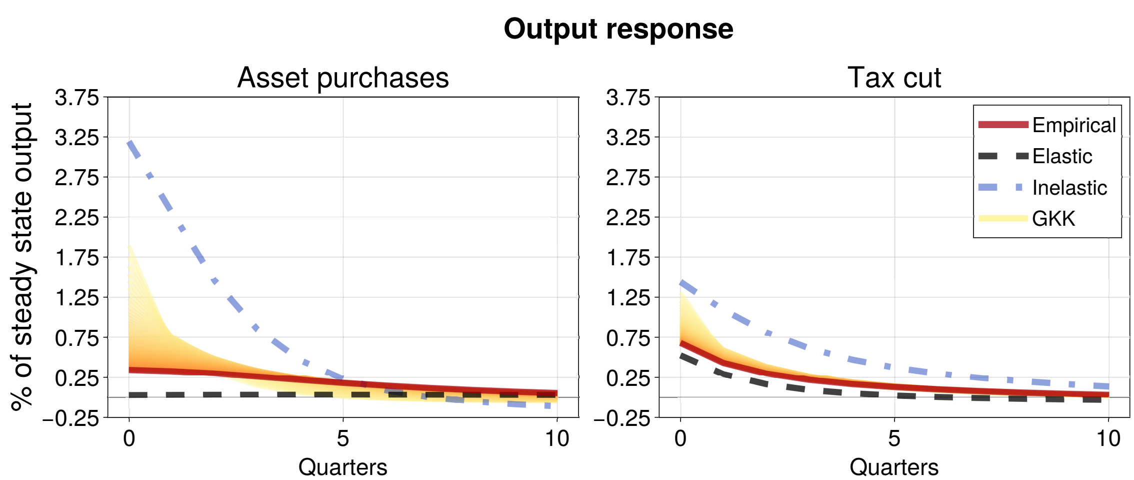

Figure 19: Output response to asset purchases and tax cuts with $\eta = 0.5, \rho_{b^G} = 0.95$. Red: empirical estimate ($\bar{\Theta}_{r^K} = 25$). Black: perfectly elastic liquidity supply ($\bar{\Theta}_{r^K} \to \infty$). Blue: perfectly inelastic liquidity supply ($\mathcal{D}_{r^K} = 0$). Dark to light yellow: high to low $\bar{\Theta}_{r^K}$ (from $\bar{\Theta}_{r^K} = 25$ to $\bar{\Theta}_{r^K} = 12$ in GKK).

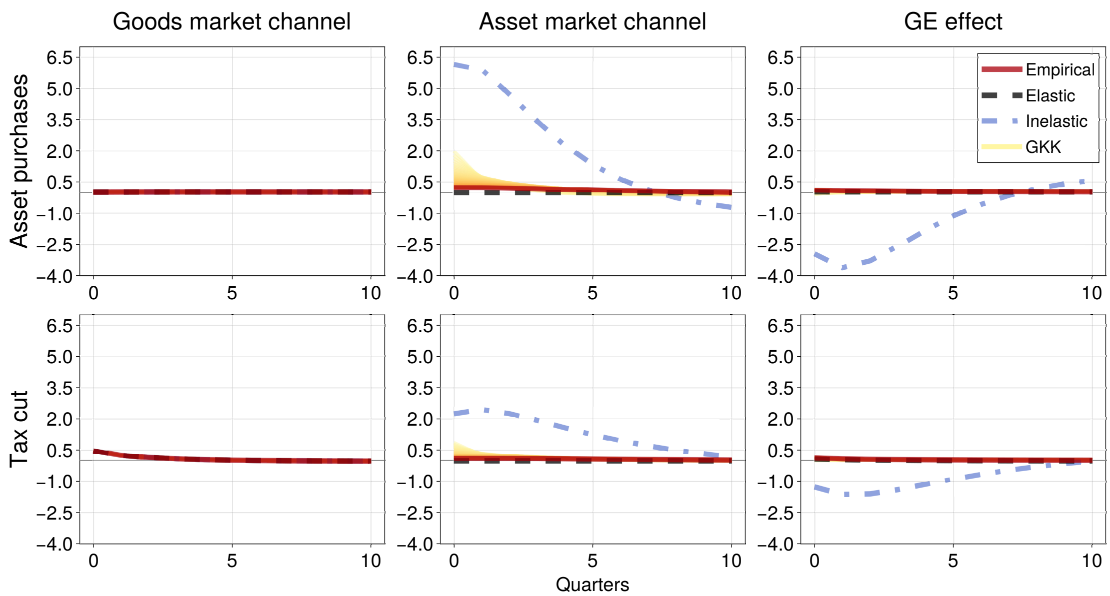

Figure 20: Decomposition of output responses to asset purchases and tax cuts with $\eta = 0.5, \rho_{b^G} = 0.95$. GE effect is defined as the difference between output responses and the sum of goods market channel and asset market channel from Theorem 1. Red: empirical estimate ($\bar{\Theta}_{r^K} = 25$). Black: perfectly elastic liquidity supply ($\bar{\Theta}_{r^K} \to \infty$). Blue: perfectly inelastic liquidity supply ($\mathcal{D}_{r^K} = 0$). Dark to light yellow: high to low $\bar{\Theta}_{r^K}$ (from $\bar{\Theta}_{r^K} = 25$ to $\bar{\Theta}_{r^K} = 12$ in GKK).

E.3 Robustness Check: General Net Worth Process

We study the extent to which our policy conclusions are affected by the parameterization of the net worth accumulation process, function G. Our calibration implies $\bar{G}$, the persistence of net worth, is equal to 0.97. $\bar{G}_{r^K}, \bar{G}_{\Theta}$, sensitivity of net worth to current returns on capital and sensitivity to past leverage, are 3.76 and 0.008. Figure 21 shows impact multipliers and Figure 22 cumulative multipliers. Figure 23 shows the relative effectiveness of asset purchases and tax cuts.

The first panel of Figures 23, 21, and 22 shows results for different values of $\bar{G}$. The red line represents our baseline specification in Figure 10. The gray shades from dark to light represent deviations from our baseline for $\bar{G}$ from 0.1 to 0.97.

We use results from Appendix B.3 to motivate robustness check with respect to $\bar{G}_{r^K}$ and $\bar{G}_{\Theta}$. In Appendix B.3 we show that allowing for endogenous equity issuance decisions is isomorphic to a reparametrization of $\bar{G}_{r^K}$ and $\bar{G}_{\Theta}$. We vary $\bar{G}_{r^K}$ from 1.48 to 3.76 (baseline) and $\bar{G}_{\Theta}$ from 0.008 (baseline) to 0.12. These choices are motivated as follows. Consider the problem of optimal equity issuance described in

B.3, where the intermediary can issue equity e_t subject to a cost function $C(e_t/n_t)$. Assume

$$C\left(\frac{e_t}{n_t}\right) = \frac{\zeta}{2}\left(\frac{e_t}{n_t}\right)^2 - (\bar{\eta} - 1)\frac{e_t}{n_t}$$

with $\zeta > 0$ and where $\bar{\eta}$ is the steady state value of η_t. This cost function is similar to Karadi and Nakov (2021) and Akinci and Queralto (2022), with the difference being the linear term. The presence of the linear term means that endogenous equity issuance (above $\bar{m}$) is zero in the steady state. This change allows us vary the parameter ζ reflect the cost of issuing equity without changing the steady state.

In our robustness checks with respect to $\bar{G}_{r^K}$ and $\bar{G}_{\Theta}$ we set $\bar{G}_{r^K}$ and $\bar{G}_{\Theta}$ to values corresponding to the parameter ζ from ∞ (baseline, in which it is impossible to issue equity) to 3, a number much below 28 used in Karadi and Nakov (2021) and Akinci and Queralto (2022). The second and the third panel in Figures 23, 21, and 22 show the effect of changing $\bar{G}_{r^K}, \bar{G}_{\Theta}$ separately. The fourth panel varies them together. Again, the red line corresponds to the baseline and the gray shades from light to dark represent outcomes for reparametrizations with values closer (dark) or farther to the baseline (light).

The results of the robustness check indicate that, given our estimates of $\bar{\Theta}_{r^K}$ and γ, the role of the function G is limited, and our conclusions about the multipliers and the effectiveness of asset purchases vs. tax cuts in Section 6 remain unchanged.

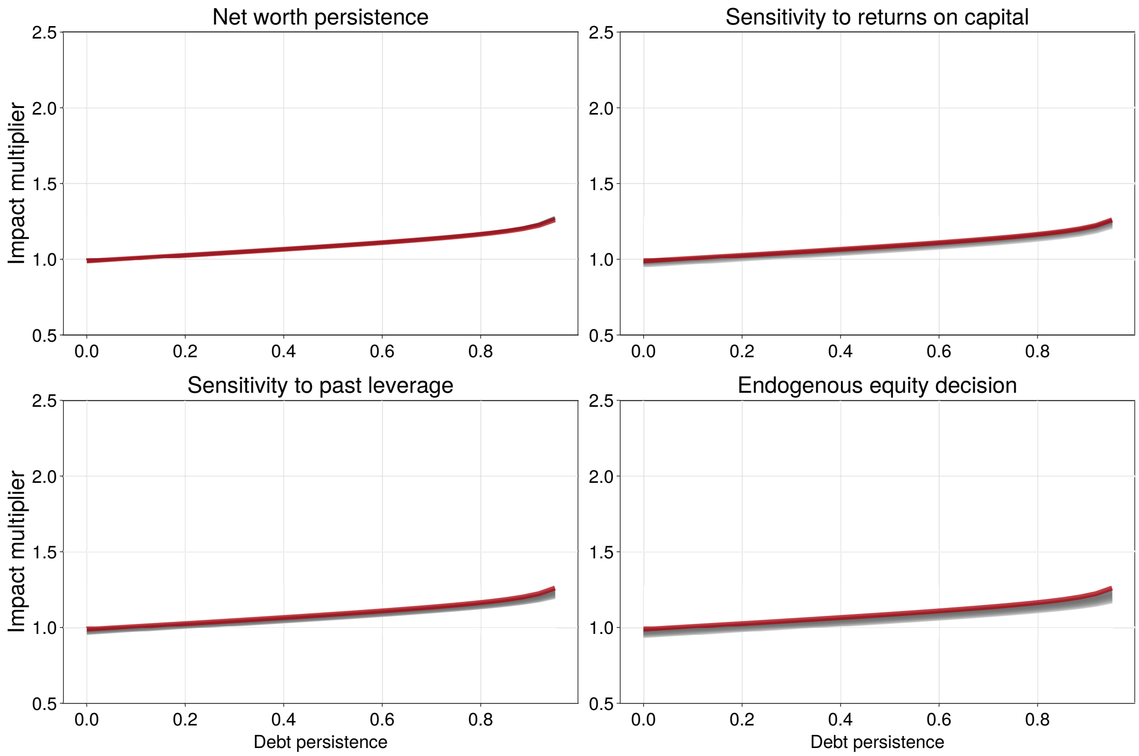

Figure 21: Impact multipliers. Red: baseline. Dark to light gray: models closer and farther to the baseline – see text for the description.

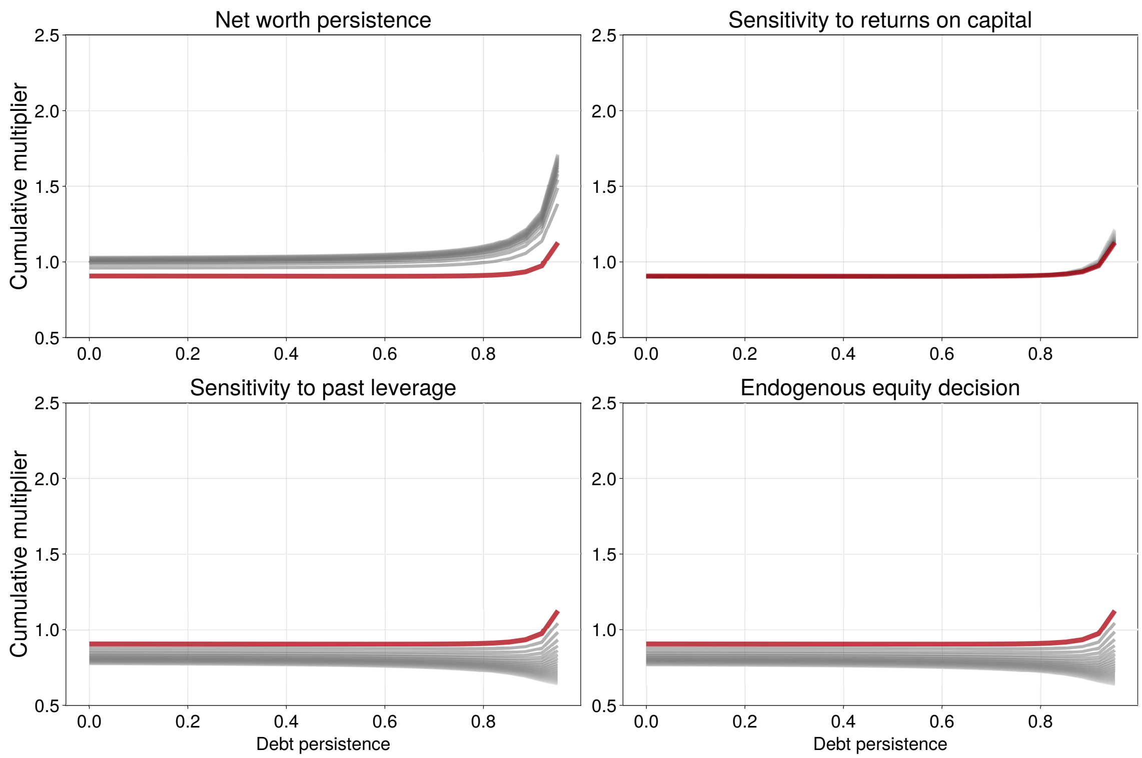

Figure 22: Cumulative multipliers. Red: baseline. Dark to light gray: models closer and farther to the baseline – see text for the description.

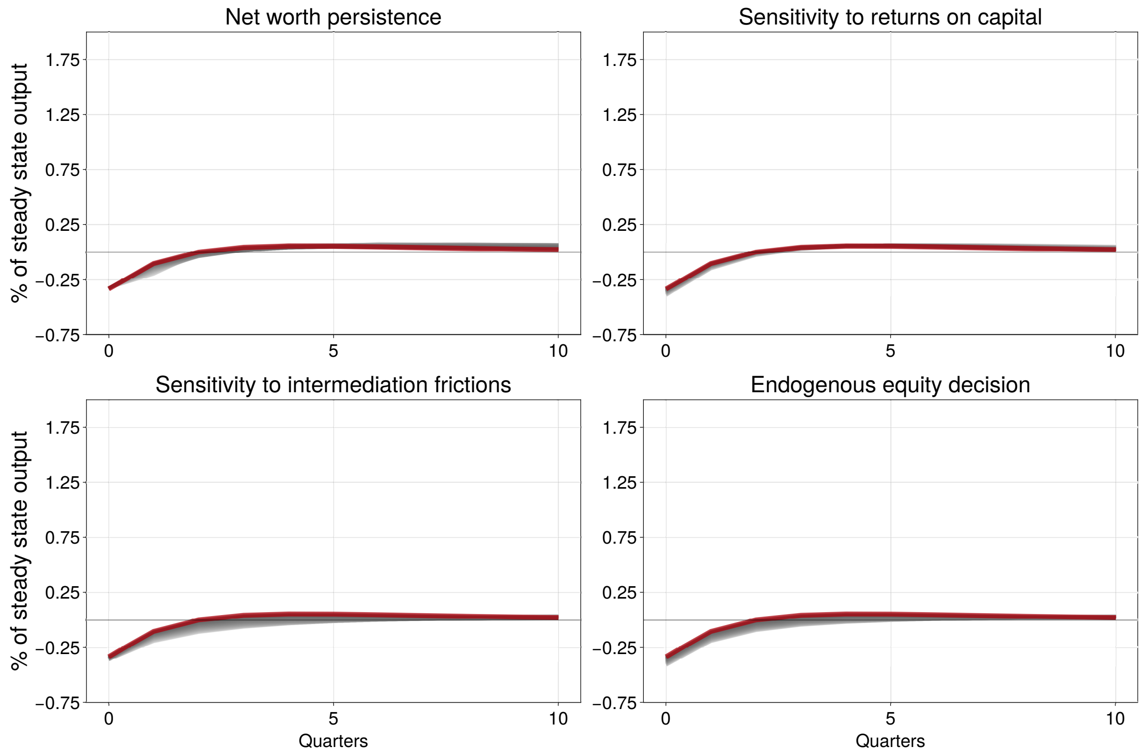

Figure 23: Difference between output response to asset purchases and tax cuts. Red: baseline. Dark to light gray: models closer and farther to the baseline – see text for the description.