Financial Intermediation and Aggregate Demand: A Sufficient Statistics Approach

Yu-Ting Chiang[†] Piotr Zoch[‡]

June 18, 2024

Click here for the most recent version.

Abstract

We derive sufficient statistics to describe how the financial sector affects aggregate responses to macroeconomic policies. Our framework nests financial intermediation with various microfoundations and features heterogeneous and illiquid households. Relevant features of the financial sector are summarized by its liquidity supply elasticities. These elasticities can be linked directly to data and are quantitatively important for debates over the effectiveness of policies targeting the financial sector vs. households. Among workhorse models, output responses differ by orders of magnitude due to implicit assumptions about these elasticities. Our estimates for the U.S. imply a stronger effect targeting households than the financial sector.

Keywords: financial frictions, HANK, monetary and fiscal policy, liquidity

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 Anmol Bhandari, Harris Dellas, Bill Dupor, Mikhail Golosov, Lars Hansen, Greg Kaplan, Kurt Mitman, B. Ravikumar, Ricardo Reis, Wenting Song, Ludwig Straub, Mikayel Sukiasyan, Harald Uhlig, Christian Wolf, and Mark Wright for valuable suggestions. Piotr 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. Previously circulated as "Asset Supply and Liquidity Transformation in HANK."

1 Introduction

Macroeconomic policies affect aggregate outcomes through various channels. Some policies, such as fiscal transfers and tax cuts, target the household sector. Other policies, such as monetary policy or government asset purchases, aim to influence outcomes through the asset markets. Understanding the transmission mechanism is useful because it helps us answer controversial policy questions, such as whether allocating resources to the financial sector is effective compared to policies that target households.

Much effort has focused on identifying and measuring features that summarize households' responses to policies (e.g., the intertemporal marginal propensity to consume, as in Auclert et al. 2023). However, few attempts have been made to identify features that matter for asset market transmission. The challenge stems partly from the complexity of existing financial intermediation models, which often involve microfoundations with hard-to-measure parameters like asset diversion rates or monitoring costs. Consequently, it is difficult to isolate features of the financial sector that are most relevant for aggregate outcomes and to quantify their impact. This paper addresses this problem.

Our main idea is a simple observation that financial intermediaries are, effectively, suppliers of assets: They take one form of asset, such as loans, and transform it into a different type of asset, such as deposits. And for many policies that take effect through changes in price and quantities of assets, the exact microfoundations are irrelevant as long as we know the shape of the supply system. The goal of this paper is to build on this observation and derive a set of sufficient statistics to summarize the role of financial intermediation and show how it interacts with other blocks of the economy to affect aggregate outcomes.

Our framework is general, nesting models of financial intermediation with various microfoundations and allowing for household heterogeneity and illiquidity. Households consume and save in different forms of assets. Some assets are preferable to others due to either reduced-form preference or the existence of portfolio adjustment costs. We refer to the preferable assets as *liquid assets*. The financial sector can transform one form of asset into another: By issuing liquid assets and holding illiquid capital, the financial sector supplies liquidity to the economy. The supply of liquidity is subject to intermediation frictions: The level of intermediation in any period is described by a function of expected future returns, representing frictions commonly used in the literature. We assume production is subject to nominal rigidities, and therefore, the government can influence real returns on liquid assets directly. The government sets policies that take place in both the real sector, such as spending, taxes, and transfers, and the asset markets, including interest rate policies, issuance of government debt, and asset purchases. We study how these policies affect the aggregate outcomes of the economy.

We show that intermediation frictions in a large class of models can be reduced to a simple structure, including those with frictions originating from asset diversion in Gertler and Karadi (2011), costly state verification in Bernanke et al. (1999), reduced-form leverage cost in Cúrdia and Woodford (2016), and collateral constraints similar to Kiyotaki and Moore (1997). Features of the underlying frictions are summarized by how the financial sector's leverage responds to expected returns over different horizons, with two parameters governing the size of responses and one parameter that describes how forward-looking the responses are. This unifying structure can be linked to data directly and allows us to provide an empirical summary of the intermediation frictions without taking a strong stance on the exact microfoundation. Using the structure, we derive a set of *liquidity supply elasticities* for the financial sector, which describes its response to expected and realized returns over an infinite time horizon. In particular, the cross-price elasticities with respect to returns on capital are central to an asset market channel: If cross-price elasticities are low, the financial sector finds liquid assets and illiquid capital less substitutable; other things equal, an increase in excess liquidity (e.g., government debt) leads to larger increases in capital prices and raises aggregate demand through investment and consumption.

We study how the transmission of policies depends on the financial sector in a demand-and-supply system of goods and assets. Up to first-order approximation, the financial sector's liquidity supply elasticities are sufficient statistics that contain all relevant information about its response to policies and aggregate variables. Through the demand-and-supply system, policies affect aggregate output through three channels: (1) A goods market channel through which policies directly shift aggregate demand, such as tax cuts affecting consumption. (2) The asset market channel through which policies such as government debt issuance shift excess liquidity and affect aggregate demand through asset markets. (3) A modified Keynesian cross through which the feedback between aggregate demand and income is altered by asset market responses. Our sufficient statistics are crucial for policies working through the latter two channels, as asset market responses depend on the financial sector. In particular, asset market responses are mostly determined by the financial sector's liquidity supply if households' liquidity demands are inelastic, as will be in the quantitatively relevant case. Our characterization is closely connected to works by Auclert et al. (2023), Wolf (2021b), Dávila and Schaab (2023), McKay and Wolf (2023), and Wolf (2021a), who study aggregate responses to policies in the sequence space. While these works abstract away from financial intermediation, we show the exact channels through which financial frictions interact with other parts of the economy to affect aggregate outcomes.

We summarize relevant features of intermediation frictions by estimating how the banking sector's leverage responds to expected returns and calculate the implied liquidity supply elasticities for the U.S. economy. Our estimation faces an important threat to identification: Any unobserved shocks that directly affect the banking sector's leverage decision will affect expected returns in general equilibrium and lead to omitted variable bias. To overcome the problem, we construct a set of instrumental variables, which are return variations attributed to identified shocks from three common shock proxies for monetary policy from Bauer and Swanson (2023), oil supply from Baumeister and Hamilton (2019), and intermediary net worth from Ottonello and Song (2022). With these instruments, we estimate how leverage responds to expected returns. Our estimation result implies that the liquidity supply elasticities of the U.S. banking sector are around twice as large as those implied by a typical calibration in financial intermediation models. Empirical estimates of the elasticities are useful because implicit assumptions about these elasticities range from zero to infinity among workhorse macro models. We show these differences lead to diverging conclusions about policies.

We demonstrate that our sufficient statistics for the financial sector have quantitatively important policy implications. The policy question we study is motivated by the "Wall Street vs. Main Street" debate: Can injecting resources into the financial sector stimulate the economy more effectively than transferring them to households? We calibrate the household sector to a standard two-asset HANK model, matching households' asset holdings and consumption responses in the data. We compare two policies where the government issues debt of 1% steady-state annual GDP, with one policy directing resources to purchase illiquid assets and the other paying out the proceeds as tax cuts. Across common assumptions about the financial sector, the effects of each policy differ by orders of magnitude. The difference in output responses between asset purchases and tax cuts ranges from 1.75% to −0.5% on impact. Calibration in standard financial intermediation models predicts a stronger effect targeting the financial sector (asset purchases) with a difference of 0.6%. By contrast, liquidity supply from our empirical estimates indicates weakened asset market responses and a stronger effect targeting households (tax cuts) with a gap of −0.33%.

Our quantitative result complements recent works that integrate financial intermediation models into heterogeneous agent frameworks featuring realistic consumption and saving behaviors; for example, Lee et al. (2020), Fernández-Villaverde et al. (2020), Lee (2021), Mendicino et al. (2021), and Ferrante and Gornemann (2022). These papers are quantitative in nature. Our sufficient statistics approach isolates theoretical objects that are most relevant to the interaction between the two paradigms. Our empirical measures of the key elasticities provide useful target moments for quantitative models focusing on such interaction.

2 Model

2.1 Households

Time is discrete, $t \in \{0, \ldots, \infty\}$. Households are indexed by $i \in [0,1]$ and have time separable preferences. Households derive utility from final goods consumption $c_{i,t}$ and potentially their holdings of two types of assets: illiquid assets $a_{i,t}$ and liquid assets $b_{i,t}$, 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:

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

subject to budget constraints

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

and borrowing constraints $a_{i,t} \geq \underline{a}$, $b_{i,t} \geq \underline{b}$, where $y_{i,t} = z_{i,t}\frac{W_t}{P_t}h_{i,t}$ denotes their real labor income. 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. 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 $\mathcal{T}_t(y_{i,t})$. There is no aggregate uncertainty, and households form expectations over idiosyncratic shocks $z_{i,t}$.

Our formulation of households is general enough to encompass standard representative agent models ($z_{i,t} \equiv 1$ without 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 specialized benchmarks useful for illustration, heterogeneous agent models generate consumption and asset allocation behaviors that are crucial for understanding aggregate responses to policies.

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_ℓ, supplied by unions indexed by $\ell \in [0,1]$:

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

and $\varepsilon_W > 1$ 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:

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

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

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

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:

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

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:

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

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 rigidity. The symmetry between unions implies nominal wages sum to $z_{i,t} W_t h_{i,t}$ for each household.

2.3 The Financial Sector

The financial sector consists of a financial intermediary and a passive mutual fund. The intermediary can transform part of the aggregate capital stock into liquid assets, thereby supplying liquidity to the economy. The mutual fund is held by households as illiquid assets, which comprise the intermediary's net worth and the rest of the capital stock.

Intermediary: The intermediary issues liquid assets $\tilde{d}_t$ to finance holdings of capital and government debt, k_t^B and b_t^B. We assume liquid assets issued by the intermediary are perfect substitutes for the government debt and thus pay the same real rate of return r_t^B. The intermediary's liquidity supply $d_t := \tilde{d}_t - b_t^B$ is given by its liquid asset issuance net of government debt holdings. Through this process, the intermediary transforms illiquid capital into liquid assets that households prefer. For example, the process can represent the intermediary's superior ability to manage loans. Without the intermediary, households would need to perform the task themselves and incur cost $\Phi(\cdot)$.

However, the intermediary's ability to transform assets is constrained by frictions.

The intermediary has net worth n_t at time t. With each unit of net worth, the intermediary is only able to intermediate capital of value Θ_t, which represents the underlying frictions. The intermediation frictions and net worth are given by:

$$\Theta_t = \Theta(\{r^K_{s+1}, r^B_{s+1}\}_{s\geq t}), \quad n_t = G(\Theta_{t-1}, r^K_t, r^B_t)n_{t-1} + m. \tag{2}$$

Intermediation frictions $\Theta(\cdot)$ depends on the entire path of future returns $\{r^K_{s+1}, r^B_{s+1}\}_{s\geq t}$, which respectively represent the future investment opportunity and funding cost for the intermediary. This specification of the intermediation frictions allows us to nest a class of frictional financial intermediation models as special cases, along with a few useful extensions of these models. We discuss this nesting result in Section 3.1. Net worth n_t evolves as a fraction of the past net worth $G(\cdot)n_{t-1}$ plus an exogenous net worth inflow m. The total of the two represents the net allocation of resources from the passive mutual fund. Function G depends on endogenous variables predetermined at time t: Θ_{t-1} and realized returns r^K_t, r^B_t. This formulation contains the common exogenous rules for net worth, as in Gertler and Kiyotaki (2010). Moreover, in Appendix B.3, we show that, for the purpose of understanding aggregate responses, our formulation is equivalent to a class of models with endogenous dividend and equity issuance decisions.

Given intermediation friction Θ_t and net worth n_t, the capital holding and liquidity supply of the intermediary satisfy

$$q_t k^B_t = \Theta_t n_t, \quad d_t = (\Theta_t - 1)n_t.$$

Liquidity supply follows directly from the balance sheet of the intermediary, $q_t k^B_t = n_t + d_t$. The balance sheet also implies that holding one unit of intermediary net worth from period t to $t+1$ gives a leveraged return:

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

Illiquid assets holdings: Illiquid assets in the economy are supplied by a passive mutual fund. The mutual fund holds intermediary net worth n_t and adjusts its capital holding k^F_t to satisfy illiquid asset demand. Let a_t denote the value of the mutual fund: $a_t = n_t + q_t k^F_t$. The return on illiquid assets is given by the weighted average of returns on capital and intermediary net worth

$$r_{t+1}^{A}=\frac{1}{a_t}(r_{t+1}^{N}n_t+r_{t+1}^{K}q_tk_t^{F}). \tag{3}$$

2.4 Government

The government sets a sequence of government purchases g_t, government debt b_t^G, liquid rate target r_t^B, total tax revenue T_t, and illiquid assets holdings a_t^G. The government debt is real debt, and monetary policy adjusts the nominal interest rate to keep the (real) liquid rate target for all $t>0$; the liquid rate in period 0, r_0^B, is predetermined. The government collects tax revenue through a 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:

$$b_t^G-(1+r_t^B)b_{t-1}^G=a_t^G-(1+r_t^A)a_{t-1}^G+g_t-T_t. \tag{4}$$

2.5 Definition of Equilibrium

Given $\left\{g_t,b_t^G,r_t^B,T_t\right\}$, an equilibrium consists of prices $\left\{q_t,P_t,R_t,W_{\ell,t},r_t^A,r_t^K\right\}$ and allocations $\{y_t,c_{i,t},x_t,h_t,h_{i,\ell,t},k_t,k_t^F,k_t^B,a_t,a_t^G,a_{i,t},b_{i,t},n_t,d_t\}$ 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:

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

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 liquid assets 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 asset holdings of the fund. The capital market clears when total capital held by the intermediary and the fund equals the aggregate stock of capital, $k_t^F + k_t^B = k_t$.

3 Intermediation and Liquidity Supply

We first provide a general characterization of financial intermediation. We show the formulation in Section 2 contains a large class of financial intermediation models with various objective functions and different constraints. Despite the generality, we show these models share a simple structure that describes how the liquidity supply of the financial sector responds to changes in returns. This simple structure of liquidity supply not only clarifies the distinct strengths and restrictions of each nested model but also provides us with a concise summary of the entire class of models. The purpose of our nesting result is not to discriminate between these models. Instead, we seek this summary so that we can systematically study in Section 4 how frictions in the financial sector interact with the real sector to determine aggregate outcomes.

3.1 Nesting Models of Financial Intermediation

We now provide examples of models nested in our framework and lay out details of these models 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.[1]

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

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

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

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

These models are nested by the framework described in Section 2.3, as stated in the following lemma:

Lemma 1 *Suppose that $\{q_t k_t^B, d_t\}$ is a solution to the problem in model $j \in \{1, \dots, 4\}$. There exists a function $\Theta(\{r_{s+1}^K, r_{s+1}^B\}_{s\geq t})$ such that $q_t k_t^B = \Theta_t n_t$ and $d_t = (\Theta_t - 1)n_t$. Moreover, when evaluated at the stationary equilibrium,*

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

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 sensitivity parameters that govern how strongly Θ_t responds to returns, and γ is a forward-looking component that controls how Θ_t responds to returns in horizon $s - t$.

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

By contrast, costly state verification and costly leverage models feature no forward-looking component, $\gamma = 0$, and intermediation does not respond to expected returns

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

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 sensitivity parameters, $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}$. The 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^K_{t+1}. However, $\bar{\Theta}_{r^K}$ and $\bar{\Theta}_{r^B}$ are pinned down by steady-state leverage and returns, similar to asset diversion models.

Besides nesting models commonly used 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 $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}$ and a forward-looking component $\gamma > 0$. We discuss these generalizations in Appendix B.2.

3.2 Liquidity Supply Elasticities

Intermediation frictions $\Theta(\{r^K_{t+s}, r^B_{t+s}\})$ and net worth process $G(\Theta_{t-1}, r^K_t, r^B_t)$ of the financial intermediaries together determine the financial sector's ability to intermediate assets, as represented by its ability to supply liquidity d_t. By studying how liquidity supply responds to returns, we can understand how aggregate conditions affect the financial sector's ability to intermediate assets. To study this mapping, we define the liquidity supply function, $\mathcal{D}_t(\{r^K_s, r^B_s\}^\infty_{s=0})$, as the level of liquidity supply d_t given returns $\{r^K_s, r^B_s\}^\infty_{s=0}$. The responses of liquidity supply to returns are described by two sets of semi-elasticities:

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

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

where $\Sigma(s) := \bar{G}_\Theta \frac{1-(\gamma\bar{G})^s}{1-\gamma\bar{G}}$, 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 $\frac{\partial \mathcal{D}_t / \partial r^K_s}{\mathcal{D}_t}$ 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. □

These elasticities describe how features of the financial sector determine its liquidity supply in response to returns. For example, the cross-price elasticities, $\frac{\partial \mathcal{D}_t / \partial r_s^K}{\mathcal{D}_t}$, are positive and increasing in $\bar{\Theta}_{r^K}$. If cross-price elasticities are high, the financial sector is willing to provide more liquidity in response to an increase in r_s^K. On the contrary, only a small decrease in r_s^K will be necessary for the financial sector to increase its holdings of government debt (therefore reducing its net liquidity supply). In other words, capital and liquid assets are more substitutable for the financial sector if cross-price elasticities are high.

The simple structure in Proposition 1 describes how liquidity supply in time t responds to changes in returns at time s. If $s > t$, an increase in r_s^K directly increases liquidity supply through relaxing intermediation friction Θ_t with sensitivity $\bar{\Theta}_{r^K}$. Moreover, it relaxes frictions in all periods before 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_s^K with $s \leq t$ has no direct effect on Θ_t and affects liquidity supply only through the propagation of net worth: Net worth at time s increases directly by $\bar{G}_{r^K}$ and through accumulation from all periods before, $\Sigma(s)$. Both propagate to period t with rate $\bar{G}$.

In the next section, we show that these elasticities of liquidity supply are key objects that summarize relevant features of the financial sector for aggregate responses. In principle, these elasticities could be estimated nonparametrically with proper instruments. Yet, Proposition 1 is useful because it shows that, for a large class of models, these infinite-dimensional elasticities reduce to a simple structure with a few parameters. The structure allows us to provide an empirical summary of these models by estimating key parameters. Moreover, these elasticities are policy invariant to the extent that the policy instruments affect the financial sector through changes in returns. As a result, the structure allows us to systematically study how financial frictions affect macro policies without taking a stance on the exact microfoundation of the underlying frictions.

4 Aggregate Responses to Policies

We now study how the financial sector interacts with the real sector to determine aggregate responses to policies in general equilibrium.

4.1 A Demand-and-Supply Representation

We recast the aggregate behavior of agents as a demand-and-supply system of goods and assets. Given government policies and key aggregate variables, we solve the optimization problem for each type of 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 changes in $\{r_s^K, r_s^B\}$. 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, $\mathcal{B}_t$.[3] Similarly, we obtain an investment function, $\mathcal{X}_t$, from the *production block*. Lemma 2 represents the equilibrium as the solution to the 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:*

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

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

and

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

where function $\mathcal{R}_t^A$ is derived from the accounting identity in Equation 3, and government asset holdings $\left\{a_t^G\right\}$ satisfy the government budget constraint in Equation 4. Moreover, given the steady state, functions $\mathcal{C}_t, \mathcal{B}_t$, and $\mathcal{X}_t$ do not depend on Θ and G.

Proof. See Appendix A.2. □

[3]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, $\Phi_t(a_{i,t}, a_{i,t-1})$.

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 $\{y_t, r_t^K\}_{t=0}^{\infty}$ such that (1) demand for final goods equals output produced, and (2) liquidity demand equals liquidity supply.

The 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 summarized by $\mathcal{D}_t$. The liquidity supply elasticities characterized in Proposition 1 are, therefore, sufficient statistics to understand how frictions in the financial sector affect aggregate responses for all models nested in our framework, up to first-order approximation. The exact microfoundations of the financial frictions do not matter as long as they generate the same liquidity supply elasticities. Moreover, as $\mathcal{D}_t$ can be defined independently of the household sector, the result is compatible with all household specifications within our framework, including the standard representative agent frameworks commonly used in macro-finance models, as well as models that emphasize the importance of household consumption responses, such as 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 highlights an important feature of households besides their consumption responses: their liquidity demand $\mathcal{B}_t$ in the liquid asset market clearing condition.[4] 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, household liquidity demand interacts with the liquidity supply of the financial sector to determine how government policies affect aggregate outcomes $\{y_t, r_t^K\}$ through asset markets.

[4] Aguiar et al. (2021) emphasize an asset demand function, but all assets are perfect substitutes in their economy. Auclert et al. (2023) study a two-asset economy, but, as we show in Proposition 2, their setup liquidity demand does not affect aggregate outcomes due to their assumptions about the financial sector.

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 focus on 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 government policies and aggregate output $d\boldsymbol{y}$. We then use the solution for $d\boldsymbol{r}^K$ as a function of government policies and output 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 *excess liquidity* as the excess supply of liquid assets given policies and output:

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

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 3. We use $\boldsymbol{\epsilon}_{r^K}$ to denote its derivatives 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 how returns on capital respond to shifts in excess liquidity due to policies and aggregate output in order to clear the liquid asset market.

Proposition 2 *In equilibrium, returns on capital satisfy*

$$d\boldsymbol{r}^K = (-\boldsymbol{\epsilon}_{r^K})^{-1}[d\boldsymbol{b}^G + \boldsymbol{\epsilon}_T d\boldsymbol{T} + \boldsymbol{\epsilon}_{r^B} d\boldsymbol{r}^B + \boldsymbol{\epsilon}_y d\boldsymbol{y}]. \tag{5}$$

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 $\{r^A_s\}$. In this case, $\boldsymbol{\epsilon}_{r^K}$ is determined fully by liquidity supply $\mathcal{D}_t$. Furthermore, suppose the financial sector features no forward-looking component in their intermediation frictions, $\gamma = 0$, and maintains a constant net worth $n_t = m$. In this case, Proposition 1 implies $(-\epsilon_{r^K})^{-1} = (-\bar{\Theta}_{r^K} m)^{-1}\boldsymbol{I}$, and Proposition 2 implies an increase in db^G_t leads to a decrease in dr^K_{t+1} by $(-\bar{\Theta}_{r^K} m)^{-1} db^G_t$. Intuitively, because household liquidity demand is inelastic, an increase in public liquidity, b^G_t, will need to be absorbed by the financial sector through holding more government debt and decreasing their net liquidity supply. This requires a decrease in expected returns dr^K_{t+1}, accompanied by an increase in capital price q_t. When $\bar{\Theta}_{r^K}$ is low (inelastic liquidity supply), capital and liquid assets are less substitutable for the financial sector. The same increase in b^G_t, therefore, leads to a larger decrease in expected return and a large increase in capital price.

Two polar assumptions about the financial sector provide important benchmarks. On 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. As a result, the financial sector accommodates shifts in excess liquidity without any changes in asset prices, and r^K_{t+1} is fully determined by r^B_{t+1}: 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, $\bar{\Theta}_{r^K} = \bar{\Theta}_{r^B} = G = 0$, the 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.2, 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 7.

Two polar cases of liquidity demand also provide useful intuition. 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. As we discuss in Section 6, a household sector that features inelastic liquidity demand is consistent with existing empirical evidence that suggests households are insensitive and inert to changes in asset returns. Moreover, we show that inelastic liquidity demand is a core quantitative implication of standard two-asset HANK models: when household consumption-saving behaviors are calibrated to match evidence from microdata, the implied household liquidity demand elasticities are orders of magnitude smaller than our measures of liquidity supply elasticities for the financial sector. As a result, asset market responses are mostly determined by features of the financial sector, represented by the structure in Proposition 1.

Aggregate Output Responses

We combine the 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 where $\boldsymbol{\Psi}_{r^K}(t,s)$ 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:*

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

where $\boldsymbol{\Omega} \coloneqq \boldsymbol{\Psi}_{r^K}(-\boldsymbol{\epsilon}_{r^K})^{-1}$.

Proof. See Appendix A.5. □

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

Matrix $\mathbf{\Omega}$ captures the key mechanism through which the asset market responses transmit into aggregate demand. Matrix $\mathbf{\Omega}$ consists of two components. First, $(-\boldsymbol{\epsilon}_{r^K})^{-1}$ describes by how much returns on capital $\boldsymbol{r}^K$ adjust in response to excess liquidity, representing the asset market responses in Proposition 2. Second, given changes in $\boldsymbol{r}^K$, matrix $\mathbf{\Psi}_{r^K}$ describes how aggregate demand responds. For example, a negative entry of $\mathbf{\Psi}_{r^K}(t,s)$ 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 q_t). This mechanism manifests itself in the asset market channel (channel 2): shifts in excess liquidity due to $d\boldsymbol{b}^G$, $d\boldsymbol{T}$, and $d\boldsymbol{r}^B$ affects aggregate demand through matrix $\mathbf{\Omega}$, which depends crucially on the financial sector through its cross-price elasticities of liquidity supply $\mathcal{D}_t$.

The same mechanism also appears in the third channel, where asset market responses, through matrix $\mathbf{\Omega}\boldsymbol{\epsilon}_y$, modify the traditional Keynesian cross logic represented by $\mathbf{\Psi}_y$. An increase in aggregate income affects excess liquidity through $\boldsymbol{\epsilon}_y$. For example, an increase in income can increase liquidity demand and reduce excess liquidity. A reduction in excess liquidity, through a positive entry of $\mathbf{\Omega}$, lowers aggregate demand. Therefore, the same force that increases aggregate demand through the asset market channel can, in fact, dampen the Keynesian cross feedback. The stronger the asset market channel, the stronger the dampening effect.

4.3 Policy Comparison: Asset Purchases v.s. Tax Cuts

Theorem 1 shows how policies rely on different channels to affect aggregate outcomes. It allows us to isolate key features of the economy to study what government policies should target to effectively influence aggregate outcomes. This question is central to the Wall Street v.s. Main Street debate: Instead of injecting resources into the financial sector, can the government more effectively stimulate the economy by transferring resources to the people?

Assume that steady state government illiquid assets $a^G = 0$ and consider two alternative policies. The first policy features a transitory government asset purchase program in which the government sets a sequence of illiquid asset positions $d\tilde{a}_t^G$, accompanied by the issuance of government debt db_t^G together with paths of goods purchase dg_t and liquid rate dr_t^B. Given $d\tilde{a}_t^G$, db_t^G, dg_t and dr_t^B, the government adjusts tax revenue $d\tilde{T}_t$ to satisfy the government budget constraint. The policy implies a sequence of net asset purchases:

$$d\Delta_t \coloneqq d\tilde{a}_t^G - (1 + r^A) d\tilde{a}_{t-1}^G,$$

where r^A is evaluated at the steady state.

For the second policy, the government keeps illiquid asset positions constant, $d\check{a}_t^G = 0$. Instead of purchasing assets, the government pays out $d\Delta_t$ as tax cuts:

$$d\check{T}_t \coloneqq d\tilde{T}_t - d\Delta_t,$$

maintaining the same paths for db_t^G, dg_t, dr_t^B. Since the two policy experiments differ only between asset purchases and tax cuts: $d\tilde{a}_t^G, d\tilde{T}_t$ versus $d\check{a}_t^G, d\check{T}_t$, we effectively control for any differences due to other policy variables, for example, the stance of monetary policy dr_t^B or a debt ceiling for db_t^G.

To compare aggregate outcomes, we define $d\boldsymbol{y}^{\text{asset}}$ and $d\boldsymbol{y}^{\text{tax cut}}$ to be the output responses to the two policies and $\widehat{d\boldsymbol{y}} \coloneqq d\boldsymbol{y}^{\text{asset}} - d\boldsymbol{y}^{\text{tax cut}}$ to be the difference in output responses. Theorem 1 immediately implies:

Corollary 1 *Given any $\{d\boldsymbol{g}_t, d\boldsymbol{r}_t^B, d\boldsymbol{b}_t^G\}$, for $a^G = 0$, the difference between output responses to asset purchases and tax cuts is given by*

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

Corollary 1 isolates key features of the economy that determine the relative effectiveness of the two policies. The two policies feature the same Keynesian cross but operate through distinct channels in goods and asset markets. Through the goods market channel, tax cuts reduce the tax burden by $d\boldsymbol{\Delta}$, which directly affects aggregate demand $\boldsymbol{\Psi}_T d\boldsymbol{\Delta}$. Since only consumption directly responds to taxes, $\boldsymbol{\Psi}_T = \boldsymbol{C}_T$. As a result, how much tax cuts affect output through the goods market is fully determined by households' consumption responses.

On the other hand, through the asset market channel, both policies feature the same increase in public liquidity $d\boldsymbol{b}^G$. Yet, asset purchases generate larger shifts in excess liquidity than tax cuts, with the difference given by $\boldsymbol{\epsilon}_T d\boldsymbol{\Delta}$. The term $\boldsymbol{\epsilon}_T = -\boldsymbol{B}_T$ captures households' excess savings due to tax cuts, which absorbs the increase in excess liquidity. Shifts in excess liquidity affect aggregate demand through matrix $\boldsymbol{\Omega}$. If household liquidity demand is insensitive to returns, asset market responses are determined by the financial sector: the asset market channel is strong if liquidity supply is inelastic and vice versa. As a result, which policy can more effectively stimulate output boils down to the relative strength of the goods and asset market channels. This is a quantitative question determined by households' marginal propensity to consume and the financial sector's liquidity supply elasticities.

5 Empirical Summary of Financial Intermediation

We estimate key parameters of financial intermediation $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}, \gamma$ with identified structural shocks. Our estimates provide an empirical summary of the intermediation frictions within the class of models nested in our framework while remaining agnostic about the underlying microfoundation.

5.1 Estimating Intermediation Frictions

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

$$d\Theta_t = \sum_{h=1}^{\infty} \gamma^{h-1} \Big(\bar{\Theta}_{r^K} \mathbb{E}_t[dr^K_{t+h}] - \bar{\Theta}_{r^B} \mathbb{E}_t[dr^B_{t+h}] \Big) + \upsilon_t, \tag{6}$$

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: 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 banking sector's market value, n_t, and net supply of liquid assets, d_t (liquid liabilities minus liquid assets holdings). Liquid assets and liabilities 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

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

Returns: We use the yield curves on U.S. Treasury bonds to construct liquid rates over different horizons, $\mathbb{E}_t[dr^B_{t+h}]$. For expected returns on capital, $\mathbb{E}_t[dr^K_{t+h}]$, we have data 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 6 may contain exogenous "leverage shocks" that affect the banking sector's ability to sustain their leverage for a given level of 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, given the expected returns. 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 changes in the level of intermediation that originate from the financial sector affect expected returns in general equilibrium, these shocks 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 business strategy in the banking sector, or a change in the banking sector's fundamentals 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 the 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

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 uses $I_t \in \left\{\mathbb{E}_t[dr^K_{t+h}],\ \mathbb{E}_t[dr^B_{t+h}]\right\}$ in our estimation. Had we had a linear model, this corresponds to the ordinary least square 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: $I_t \in \left\{\mathbb{E}_t[d\check{r}^K_{t+h}],\ \mathbb{E}_t[d\check{r}^B_{t+h}]\right\}$. 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

Note: We use monthly data from January 1999 to December 2019. Estimation uses iterative GMM for optimal weighting matrix. Standard errors use heteroskedastic and autocorrelation consistent (HAC) estimators and are displayed in parentheses.

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 in the quarterly spread between the two returns. Effective leverage responds to future returns with γ around 0.94, which implies a "half-life" of around three years: the response to a spread increase three 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. However, using the point estimates for state dependency, a back-of-envelope calculation suggests that Θ_{r^K} and Θ_{r^B} are within around 30% of our estimates as long as the aggregate state is within 2 standard deviations over the business cycle.

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.2, 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).

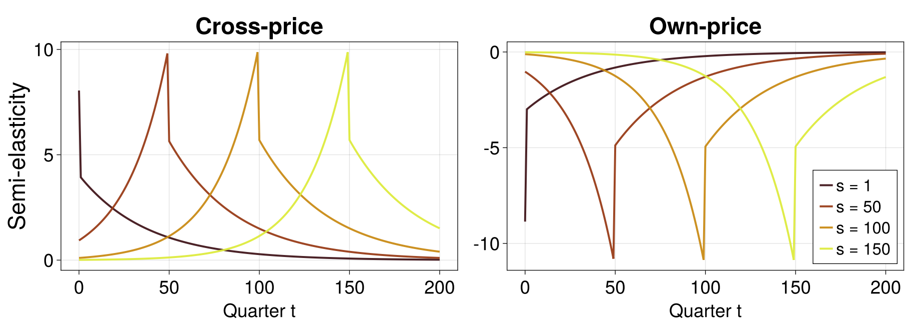

Figure 1: Semi-elasticities of liquidity supply. Each line corresponds to a different period s and shows semi-elasticity of liquidity supply in quarter t with respect to r_s^K and r_s^B.

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 prior to period s, reflecting the forward-looking component of intermediation frictions. After period s, liquidity supply drops sharply but remains elevated due to propagation through net worth. The size of the initial response $\frac{\partial \mathcal{D}_0/\partial r_1^K}{\mathcal{D}_0}$ is 8.05, implying an 8.05% increase in liquidity supply in response to a one-percentage increase in r_1^K. The own-price elasticities have a similar pattern with the opposite sign.

Benchmarks: Common Assumptions in Macro Models

We compare our estimates to three common assumptions on liquidity supply in workhorse macro models and contrast their implications in Section 7.

The first two benchmarks are models with perfectly inelastic and elastic liquidity supply. As we discussed in Section 4.2, these polar cases respectively correspond to the assumptions in Kaplan et al. (2018) and Auclert et al. (2023). These are important benchmarks because they are the current backbone for quantitative analysis of monetary and fiscal policy.

We use a Gertler-Karadi-Kiyotaki (asset diversion) model as a benchmark to compare our empirical estimates to financial intermediation models nested in our framework. This model is important because variants of GKK constitute the majority of macro models with financial intermediation. Intermediation frictions in GKK feature a forward-looking component, consistent with our empirical estimates. However, as we discuss in Section 3.1, GKK imposes a tight restriction on the liquidity supply elasticities, linking them to the steady-state returns and leverage of the intermediary. The intermediation friction parameters imposed by a GKK model are:[5]

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

The values of these parameters are, respectively, 11.9, 12.0, and 0.998, given the steady-state returns and leverage in our sample. 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. As we show in Section 7, these differences lead to diverging conclusions about policies.

[5] The formula depends on the discount factor used by the intermediary. These formula uses $\{r_{t+1}^K\}$ as discount rates. Formulas for other alternatives are provided in Appendix B.1.

6 Calibration

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 data on the banking sector balance sheet, the market value of banks, 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.

6.1 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 equals δ and the price of capital equals 1. We set $\kappa_I = 0.5$ so that the elasticity of investment to capital price is 2. Unions allocate labor uniformly among households: $l(z_{i,t}) \equiv 1/\int z_{i,t} di$. Given that monetary policy targets the real liquid rate, 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 the degree of nominal wage rigidities, 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. Liquid assets supplied by the government is 21% of steady-state output, consistent with aggregate liquid asset positions in the data shown in Appendix C. We assume the government holds no illiquid assets in the steady state, which implies that government purchases are 14.8% of GDP.

Households

Preferences: There are two types of households, indexed by s. Their population shares are 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 for $s = 1$ and to 2 for $s = 2$.[6] 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):

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

We set χ_0 to 0.1 and assume that households cannot hold negative asset positions, $\underline{a} = \underline{b} = 0$.

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 empirical 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% respectively), and the first quarter marginal propensity to consume out of \$500 transfer (MPC) (20%). Table 1 shows the model replicates the five targeted moments and reports calibrated parameter values.

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 standard two-asset HANK model calibrated to match standard targets. Households' consumption responses to an increase in liquid assets

[6] Aguiar et al. (2020) argue that allowing for heterogeneity in discount factors and intertemporal elasticities of substitution is important for matching consumption behavior observed in the data.

(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, when the portfolio adjustment cost is calibrated to match empirically plausible MPC, the consumption responses also inform us about households' liquidity demand with respect to illiquid returns.

Implied Liquidity Demand

Figure 2 shows household liquidity demand in comparison to our estimates of the financial sector's liquidity supply and excess liquidity supply of the economy. Each line represents responses to an increase in r_s^K, taking into account its effect on the sequence of illiquid returns $\{r_s^A\}$. Liquidity demand responses are an order of magnitude smaller than liquidity supply responses. This highlights a central implication of the standard two-asset heterogeneous agent models: the same frictions that generate large consumption responses also imply an inelastic liquidity demand in response to changes in illiquid returns. 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. As we discussed in Section 4.2, the large differences between the elasticities of liquidity demand and supply have an important implication on aggregate outcomes: 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. Despite its importance, workhorse macro models feature a wide variety of assumptions about the financial sector. These assumptions lead to crucial differences in their policy implications.

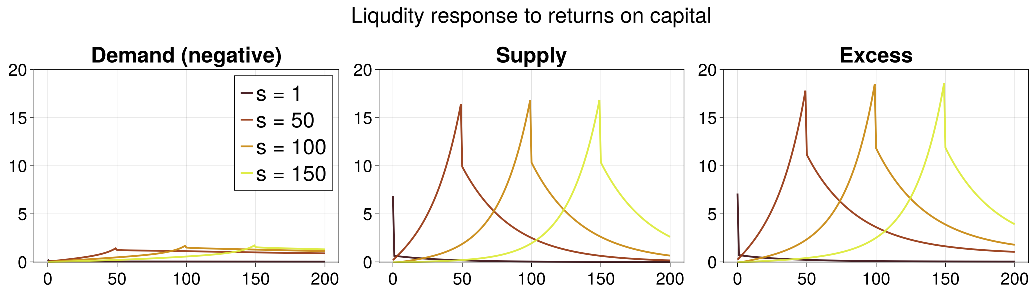

Figure 2: Entries of $-\tilde{\mathbf{B}}_{r^K}$, $\mathbf{D}_{r^K}$, and $\boldsymbol{\epsilon}_{r^K}$ matrices (see Appendix A.6 for their definitions). Each line corresponds to a different period s and shows a response of liquidity demand, liquidity supply, or excess liquidity in quarter t with respect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

7 Policy Comparison

We demonstrate that our sufficient statistics about the financial sector have quantitatively important policy implications. We consider the policy question raised in Section 4.3: Are asset purchases more effective at stimulating aggregate output than tax cuts? Through our sufficient statistics, we systematically compare the policy implications of existing models due to their implicit assumptions about the financial sector and contrast these assumptions to our empirical baseline. We show these assumptions lead to qualitatively distinct policy conclusions with quantitatively large differences.

7.1 Asset Purchases v.s. Tax Cuts

We now parameterize the two policy alternatives described in Section 4.3. For asset purchases, the government's illiquid asset positions $d\tilde{a}^G_t$ are financed by government debt of the same amount: $db^G_t = d\tilde{a}^G_t$, and goods purchases and liquid rates are kept constant: $dg_t = dr^B_t = 0$. Given these assumptions, the government budget constraint implies tax revenue: $d\tilde{T}_t = (r^B - r^A)d\tilde{a}^G_{t-1}$. We assume the government's illiquid asset positions follow:

$$d\tilde{a}^G_t = \rho\, d\tilde{a}^G_{t-1} + s_t, \quad s_t = s_0\eta^t,$$

where $\rho = 0.95$ specifies the persistence of the position and $\eta = 0.5$ implies that the government increases illiquid asset holdings for four quarters before it starts to sell them back. We set s_0 so that the peak response of debt is 1% of annual GDP. Net asset purchases are given by $d\Delta_t = d\tilde{a}^G_t - (1 + r^A)d\tilde{a}^G_{t-1}$.

For tax cuts, the government keeps the same dg_t, dr_t^B, db_t^G, but sets $d\check{a}_t^G = 0$ and reduces tax revenue to $d\check{T}_t = (r^B - r^A)d\tilde{a}_{t-1}^G - d\Delta_t$. The path of tax cuts follows a corresponding path to asset purchases: transfers are received by households mostly within four quarters, after which the government increases taxes. The paths of government debt, net asset purchases, and tax revenue implied by the two policies are compared in Figure 3.

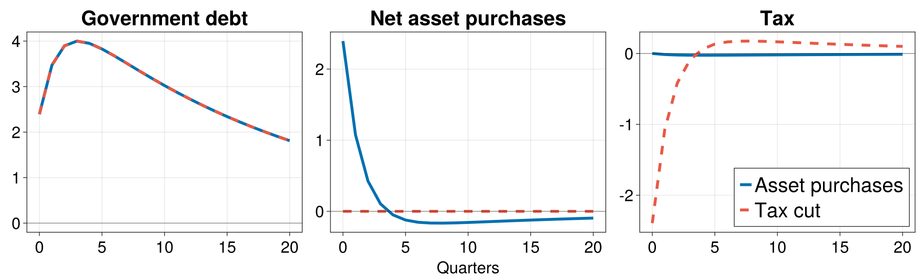

Figure 3: Government debt, net asset purchases, and taxes; x-axis: quarters, y-axis: % of steady-state GDP.

7.2 Output Responses Across Models

Liquidity supply elasticities of the financial sector have crucial implications on which policy can more effectively stimulate output. Figure 4 compares output responses to asset purchases (left panel) and tax cuts (right panel) under different assumptions about the liquidity supply elasticities. The red line represents output responses with liquidity supply elasticities implied by our empirical estimates, $\bar{\Theta}_{r^K} = 24.2$. Yellow shades from dark to light represent models with decreasing values for $\bar{\Theta}_{r^K}$ from the empirical estimates to that implied by the GKK specification, $\bar{\Theta}_{r^K} = 11.9$.[7] The blue and black lines indicate responses with perfectly inelastic and elastic liquidity supply.

The two policies increase output in all models. Yet, the magnitude of output responses differs significantly across models. Between models with inelastic and elastic liquidity supply, asset purchases lead to output responses ranging from 3.2% to 0.3% on impact. Similarly, output responses to tax cuts range from 1.4% to 0.7% between models.[8] Despite assumptions on liquidity supply leading to clear differences for both policies,

[7]We keep the forward-looking component γ at its empirical estimate of 0.957 for this exercise.

[8]Figures 10 and 11 in Appendix E.1 shows that these differences are driven mostly by investment responses to changes in capital prices.

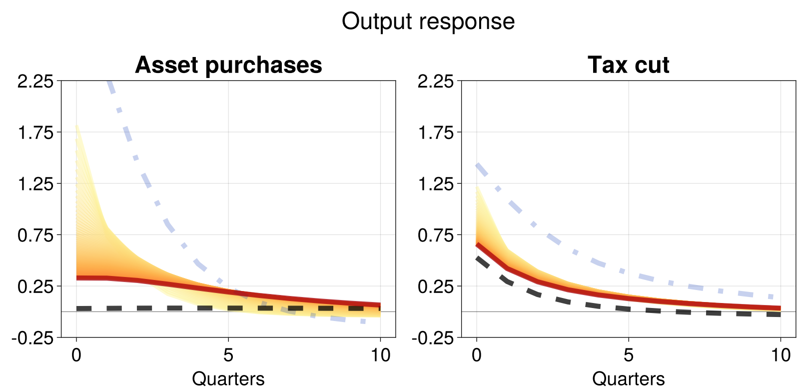

Figure 4: Output responses to asset purchases and tax cuts. y-axis: % of steady-state output. Red: empirical elasticities. Dark to light yellow: high to low elasticities, from empirical estimates to GKK. Blue: perfectly inelastic. Black: perfectly elastic. Figure 12 in Appendix E.1 shows a version with the full range of responses.

the effects of asset purchases are noticeably more sensitive to assumptions on the liquidity supply elasticities. To understand the contrast between the two policies, we decompose output responses into the three channels in Theorem 1.

7.3 Decomposition of Output Responses

We decompose output responses into the goods market channel, the asset market channel, and the modified Keynesian cross as in Theorem 1:

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

Figure 5 shows the decomposition for asset purchases (upper panels) and tax cuts (lower panels), with each line corresponding to the same model as in Figure 4. For both policies, the goods market channel is identical for all financial sector specifications. Asset purchases have little effect through this channel as their impact on tax revenue $d\tilde{T}_t$ is limited. Tax cuts, on the contrary, generate a non-negligible effect due to the high MPCs of the households, $\boldsymbol{\Psi}_T = \boldsymbol{C}_T$: the initial 2.5% reduction of the tax burden generates an increase in aggregate demand on impact around 0.5%, corresponding to the 20% MPC in our calibration.

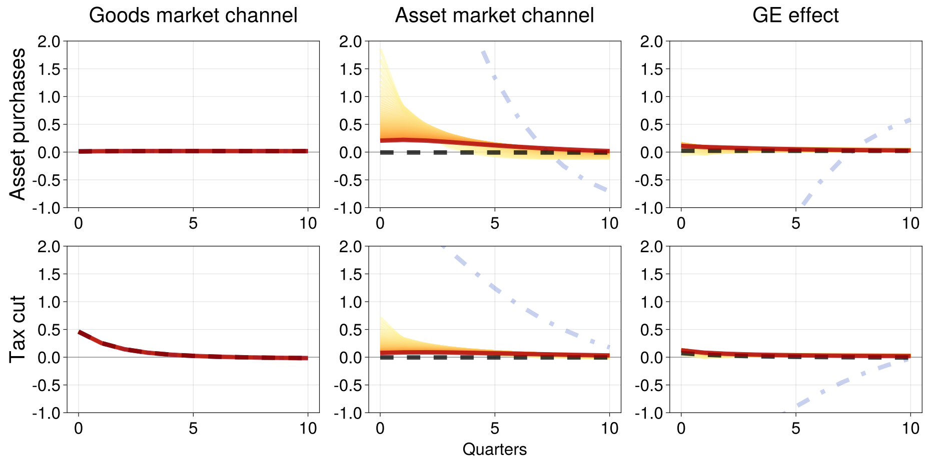

Figure 5: Decomposition of output responses using Theorem 1. The y-axis: % of steady-state GDP. Red: empirical elasticities. Dark to light yellow: high to low elasticities, from empirical estimates to GKK. Blue: perfectly inelastic. Black: perfectly elastic. Figure 13 in Appendix E.1 shows a version with the full range of responses.

The asset market channel, on the other hand, demonstrates how assumptions about the financial sector can significantly alter policy predictions. Between liquidity supply elasticities from our empirical estimates and those imposed by a GKK model, aggregate demand responses differ by an order of magnitude through this channel for both policies. The contrast is even more drastic between the perfectly elastic and inelastic benchmarks. Assumptions about the financial sector generate a wide range of responses through matrix $\mathbf{\Omega}$: excess liquidity induces changes in expected returns and leads to consumption and investment responses — the key mechanism discussed in Proposition 2 and Theorem 1. The fact that the financial sector plays a crucial role in the asset market channel is a joint result of our quantitative inference on household liquidity demand: Because household liquidity demand is insensitive to returns, asset market responses mostly reflect features of the financial sector.

In contrast to the goods market channel through which tax cuts have stronger effects, asset purchases generate a larger shift in excess liquidity and have stronger effects through the asset markets. As asset purchases rely more on the asset market channel, the policy is also more sensitive to the assumption about the financial sector: If liquidity supply is inelastic, asset purchases have a much stronger effect; if it is perfectly elastic, both policies have zero effect through this channel. As a result, different assumptions about liquidity supply lead to a wider range of responses for asset purchases than tax cuts in Figure 4 through this channel.

Finally, we present the modified Keynesian cross as the general equilibrium (GE) effect in Figure 5, which shows the difference between total output responses and the sum of the first two channels. A prominent aspect of the modified Keynesian cross is a strong dampening response when liquidity supply is perfectly inelastic (blue line): While the policies generate strong aggregate demand through the asset market channel, increases in aggregate income shift out liquidity demand, reduce excess liquidity, and significantly dampens the total output responses. The dampening force reduces the impact of the asset market channel by up to 40%. The same dampening force is also present in other specifications, although the effect is less pronounced because changes in excess liquidity have a smaller impact when liquidity supply is elastic. In fact, if liquidity supply is elastic enough, the dampening effect vanishes, and the modified Keynesian cross starts to resemble a standard Keynesian cross.

7.4 Policy Implications of Liquidity Supply Elasticities

We calculate the difference in output responses between policies:

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

These differences demonstrate the policy conclusion of each model: a positive value indicates that asset purchases have a stronger effect on output and vice versa. Moreover, by focusing on these differences, we effectively control for responses due to other policy variables: The policy conclusions do not hinge on the specifications of dg_t, dr_t^B, db_t^G as the difference in output responses depends only on $d\boldsymbol{\Delta}$.

Each line in Figure 6 represents a model with liquidity supply elasticities corresponding to those in Figure 4. At one extreme, models with perfectly inelastic liquidity supply (blue line) predict asset purchases have a stronger effect on output than tax cuts: the difference in output response amounts to 1.8% of steady-state output on impact. Liquidity supply implied by financial intermediation models of the Gertler-Karadi-Kiyotaki type (light yellow) gives a qualitatively similar prediction: Asset purchases are more effective in stimulating output. At the other extreme, models with perfectly elastic liquidity supply (black line) assume away the asset market channel and conclude that tax cuts have a stronger effect. In comparison to these benchmarks, liquidity supply elasticities from our empirical estimates generate a non-negligible response to asset purchases but predict that tax cuts targeting households have a relatively strong effect on output.

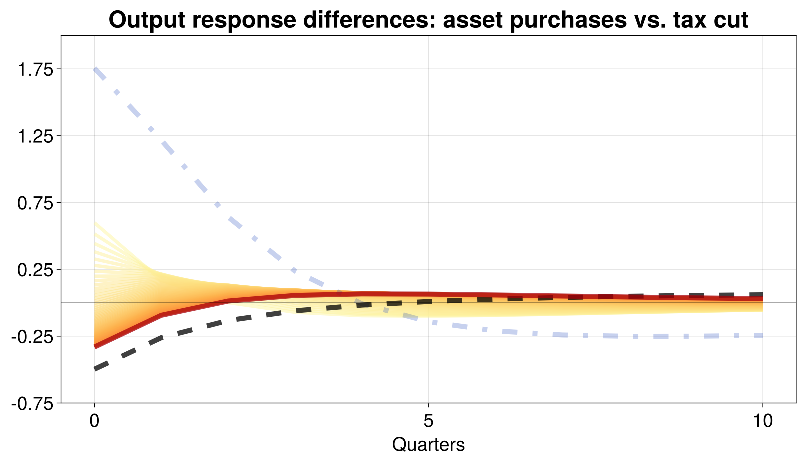

Figure 6: Difference between output response to asset purchases and tax cuts. The y-axis: % of steady-state GDP. Red: empirical elasticities. Light to dark yellow: low to high elasticities starting from GKK. Blue: perfectly inelastic. Black: perfectly elastic.

Our policy conclusion reflects several features of the economy. First, households feature strong consumption responses due to heterogeneity and illiquidity, which break Ricardian equivalence and generate large effects from deficit-financed tax cuts. Yet, the portfolio adjustment cost that generates household illiquidity has an equally important implication on their liquidity demand. Inelastic liquidity demand implies that asset market responses are primarily determined by the financial sector. Relevant features of the financial sector are summarized by our sufficient statistics — the liquidity supply elasticities. Our estimates of these elasticities are higher than the implicit assumptions of standard financial intermediation models. High elasticities imply weak asset market responses. As a result, we conclude that policies relying less on the asset market channel, such as tax cuts, can more effectively stimulate output responses than policies that target the financial sector, such as asset purchases.

8 Conclusion

We show that the liquidity supply elasticities of the financial sector are sufficient statistics to understand how it affect aggregate responses to policies. In a framework that allows for realistic household consumption-saving behaviors and nests financial intermediation with various microfoundations, these elasticities summarize all relevant frictions in the financial sector. In general equilibrium, asset market responses are almost entirely determined by these elasticities, as households are insensitive to changes in returns. Assumptions on liquidity supply vary widely among workhorse models and lead to opposite policy conclusions. Our sufficient statistics provide an empirical summary of the relevant frictions and inform which policies can most effectively affect aggregate outcomes.

The importance of these elasticities implores a comprehensive measurement beyond the scope of this paper. These elasticities are useful guides for understanding how various macroprudential regulations imposed on the financial sector affect aggregate outcomes. Liquidity demand from the production sector and the international market are absent in our analysis, but they are essential to understand how the financial sector affects the production process in a global economy. 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).

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).

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.

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 supply 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:

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

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:

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

Production

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

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

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

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

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 3. 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 3 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

$$\mathcal{R}_{t+1}^A\left(\left\{r_s^K, r_s^B, y_s; \mathcal{D}_t\right\}_{s=0}^{\infty}\right) := \frac{\left(1+r_{t+1}^K\right) - \left(1+r_{t+1}^B\right)\mathcal{L}_t\left(\left\{y_s, r_s^K; \mathcal{D}_t\right\}_{s=0}^{\infty}\right)}{1-\mathcal{L}_t\left(\left\{y_s, r_s^K; \mathcal{D}_t\right\}_{s=0}^{\infty}\right)} - 1 \quad (7)$$

Market Clearing

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

$$\mathcal{C}_t(\{y_s, r_s^A; r_s^B, T_s\}_{s=0}^{\infty}) + \mathcal{X}_t(\{y_s, r_s^K\}_{s=0}^{\infty}) + g_t = y_t, \quad (8)$$

$$\mathcal{B}_t\left(\left\{y_s, r_s^A; r_s^B, T_s\right\}_{s=0}^{\infty}\right) = \mathcal{D}_t\left(\left\{r_s^K, r_s^B\right\}_{s=0}^{\infty}\right) + b_t^G. \quad (9)$$

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 7, 9, and 9. We solve a_s^G from Equation 4, 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 5.

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

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

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

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

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

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 $\{d\boldsymbol{y}, d\boldsymbol{r}^K, d\boldsymbol{r}^B, d\boldsymbol{b}^G\}$, the limit of the liquid asset market clearing condition is

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

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

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

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

Let $\boldsymbol{\Omega} := \boldsymbol{\Psi}_{r^K}(-\boldsymbol{\epsilon}_{r^K})^{-1}$, 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 $\{dr^B_{s+1}\}_{s=0}^{\infty}$ 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

$$dr_0^K = \frac{\alpha}{\bar{k}}\mathbf{e}_1^{\intercal}d\boldsymbol{y} + (1 - \delta)\left(\mathbf{q}_y^{\intercal}d\boldsymbol{y} + \mathbf{q}_{r^K}^{\intercal}d\boldsymbol{r}^K\right), \tag{10}$$

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 10 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 7 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 10 to define

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

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:

$$\begin{aligned} \tilde{\mathbf{C}}_y :=& \mathbf{C}_y + \mathbf{C}_{r^A}\mathbf{R}_y^A + \tilde{\mathbf{C}}_{r_0^A,y}, \qquad & \tilde{\mathbf{C}}_{r^K} :=& \mathbf{C}_{r^A}\mathbf{R}_{r^K}^A + \tilde{\mathbf{C}}_{r_0^A,r^K}, \\ \tilde{\mathbf{C}}_{r^B} :=& \mathbf{C}_{r^B} + \mathbf{C}_{r^A}\mathbf{R}_{r^B}^A, \qquad & \tilde{\mathbf{C}}_T :=& \mathbf{C}_T. \end{aligned}$$

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

$$\begin{aligned} \tilde{\mathbf{B}}_y :=& \mathbf{B}_y + \mathbf{B}_{r^A}\mathbf{R}_y^A + \tilde{\mathbf{B}}_{r_0^A,y}, \qquad & \tilde{\mathbf{B}}_{r^K} :=& \mathbf{B}_{r^A}\mathbf{R}_{r^K}^A + \tilde{\mathbf{B}}_{r_0^A,r^K}, \\ \tilde{\mathbf{B}}_{r^B} :=& \mathbf{B}_{r^B} + \mathbf{B}_{r^A}\mathbf{R}_{r^B}^A, \qquad & \tilde{\mathbf{B}}_T :=& \mathbf{B}_T. \end{aligned}$$

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

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

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 $\Theta\left(\left\{r^B_{s+1}, r^K_{s+1}\right\}_{s\geq t}\right)$. We also demonstrate that in all these models financial frictions result in $\Theta_t\left(\cdot\right)$ 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. We can write the value function in a recursive form:

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

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

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

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

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

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

If the incentive compatibility constraint is binding, we have

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

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

$$\Theta_{j,t} = \frac{\Lambda_{t,t+1}\left(f\theta + (1-f)\,\Theta_{j,t+1}\right)\left(1 + r^B_{t+1}\right)}{1 - \Lambda_{t,t+1}\left(f\theta + (1-f)\,\Theta_{j,t+1}\right)\left(r^K_{t+1} - r^B_{t+1}\right)}. \tag{11}$$

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 11 forward and write $\Theta_t = \Theta\left(\left\{r^B_{s+1}, r^K_{s+1}\right\}_{s\geq t}\right)$. Aggregating individual banks $\int_0^1 l^B_{j,t} dj = q_t k^B_t$ and $\int_0^1 n_{j,t} dj = n^B_t$ we obtain

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

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 11 with respect to returns and evaluating the resulting expression at the steady state. Depending on assumptions on bankers' discount rates, we have:

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

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

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

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

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 $\psi(1) = 1$.[9] 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

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

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 2 remains the same. More specifically, assume that 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 $\Upsilon_t\left(\frac{q_t k_t^B}{n_t}\right) n_t = \bar{\Upsilon}_t$, $r_{t+1}^N n_t$ 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

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

By using the balance sheet, we can rewrite it as

$$q_t k_t^B \leq \frac{1 + r_{t+1}^B}{1 + r_{t+1}^B - \vartheta\left(1 + r_{t+1}^K\right)} n_t. \tag{12}$$

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 12 in that $1+r^K_{t+1}$ 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 $\{r^K_s\}$ and output $\{y_s\}$. Yet, we expect the two collateral constraints to generate similar dynamics when most of the changes in $1+r^K_{t+1}$ 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: $d_t \leq \vartheta q_t k^B_t$. Using $d_t = q_t k^B_t - n_t$, we have

$$q_t k^B_t \leq \frac{1}{1-\vartheta} n_t.$$

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:

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

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

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:

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

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

$$\eta_{j,t} = \max_{\delta_{j,t}} \left(\delta_{j,t}\right)^{1-\gamma} + \beta \eta_{j,t+1} \Big(\left(1 + r^B_{t+1}\right) + \left(r^K_{t+1} - r^B_{t+1}\right)\vartheta(\delta_{j,t}) - \delta_{j,t} \Big)^{1-\gamma}$$

Optimality requires:

$$\delta^{-\gamma}_{j,t} + \beta \eta_{j,t+1} \Big(\left(1 + r^B_{t+1}\right) + \left(r^K_{t+1} - r^B_{t+1}\right)\vartheta(\delta_{j,t}) - \delta_{j,t} \Big)^{-\gamma} \left(\vartheta'(\delta_{j,t}) - 1\right) = 0.$$

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

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

Define $\Theta_t = \vartheta(\varrho(\eta_{j,t+1}, r^K_{t+1}, r^B_{t+1}))$. Since $\eta_{j,t}$ follows a first-order difference equation with a terminal condition at $t \to \infty$, we can write $\Theta_t = \Theta(\{r^K_s, r^B_s\}_{s>t})$ 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:

$$v_{j,t}(n_{j,t}) = \max_{\Theta_{j,t}} \varsigma(\Theta_{j,t}, r^K_{t+1}, r^B_{t+1}) \times (\zeta(n_{j,t}) + v_{j,t}(n_{j,t})) + \beta v_{j,t+1}(n_{j,t+1}),$$

$$q_t k^B_{j,t} = \Theta_{j,t} n_{j,t}, \quad n_{j,t+1} = \Gamma(\Theta_{j,t}, r^K_{t+1}, r^B_{t+1}) n_{j,t},$$

where $\zeta(x) = \frac{1}{1-\gamma} x^{1-\gamma}$. The solution $\Theta_{j,t}$ of these problems has the structure in Lemma 1, $\Theta_{j,t} = \Theta(\{r^K_s, r^B_s\})$. 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 $n_{t+1} = G(\Theta_{t-1}, r^K_t, r^B_t) n_t + m$ consistent with the framework given by Equation 2.

B.3 Endogenous Equity and Dividend decision

We study liquidity supply in an important class of models with endogenous equity and dividend decisions. These models are not nested by our formulation in Equation 2, 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 2 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:

$$v_{j,t}(n_t) = \max_{\xi_t} \Lambda_{t,t+1} \left(f \tilde{n}_{t+1} + (1-f) \left[\mathbb{E}_t \left[v_{t+1} \left(\tilde{n}_{t+1} + \xi_t n_t \right) \right] - C(\xi_t) n_t \right] \right)$$

$$\tilde{n}_{t+1} = \left(1 + r^K_{t+1}\right) l_t - \left(1 + r^B_{t+1}\right) d_t,$$

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 2. 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_s^K, dr_s^B in some period s, and let $dr_t^K = dr_t^B = 0$, $\forall t \neq s$. 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

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

From the proof of Proposition 1, we have

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

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

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

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

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

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

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

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

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:

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

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 7 (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 7 (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 7 (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.

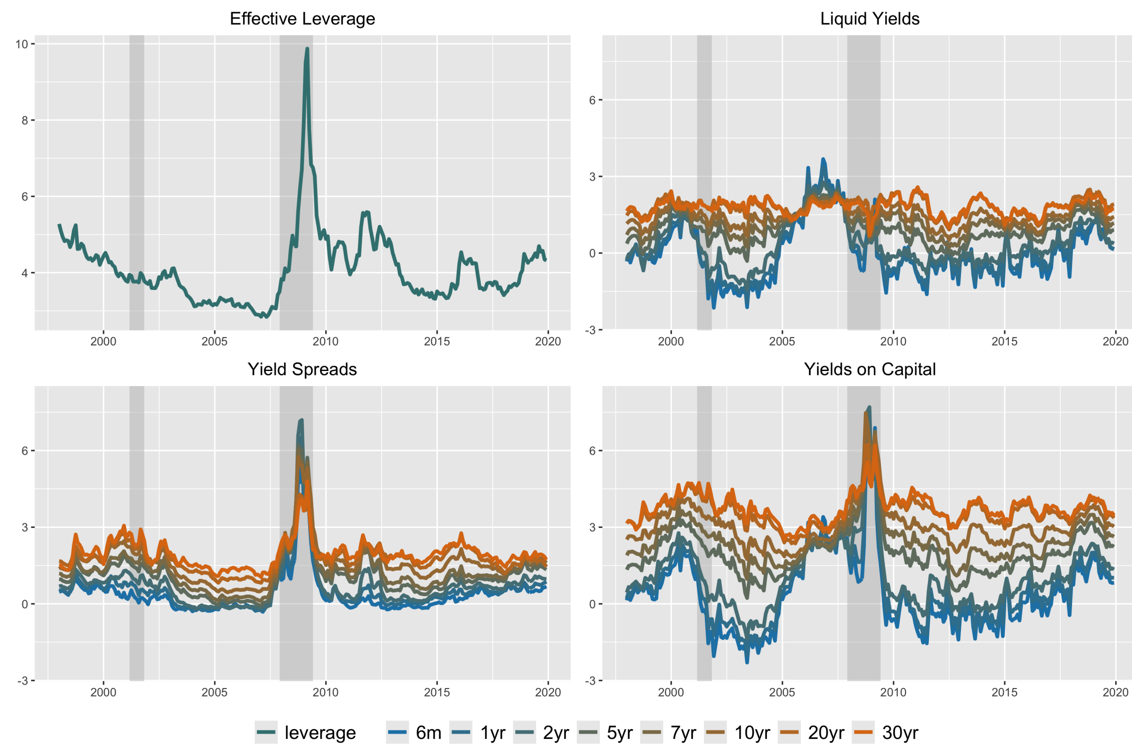

Figure 7: Top-left: Effective leverage calculated as 1+(*liquid liabilities*−*liquid assets*)/*market value of bank equity*; top-right: real yields on Treasuries, calculated as nominal yields net of inflation expectations from the Cleveland Fed; bottom-left: the spreads between High-Quality Market (HQM) 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).

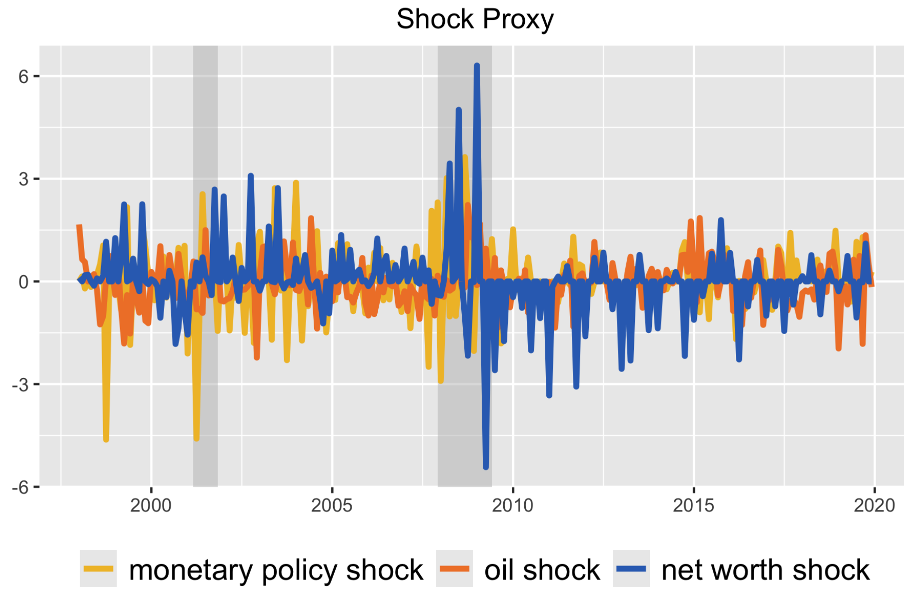

Figure 8: Shock proxies - monetary policy shock from Bauer and Swanson (2023), oil supply shock from Baumeister and Hamilton (2019), financial sector net worth shock from Ottonello and Song (2022); rescaled by standard deviation.

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, v_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

		$r^K_{t+1\text{y}}$	$r^K_{t+5\text{y}}$	$r^K_{t+10\text{y}}$	$r^K_{t+30\text{y}}$	$r^B_{t+1\text{y}}$	$r^B_{t+5\text{y}}$	$r^B_{t+10\text{y}}$	$r^B_{t+30\text{y}}$
	6m	0.06	0.05	0.04	0.08	0.08	0.05	0.07	0.08
FEVD	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:

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

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: Monthly data from January 1999 to December 2019. Estimation uses iterative GMM for optimal weighting matrix. Standard errors use heteroskedastic and autocorrelation consistent (HAC) estimators and are displayed in parentheses.

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 large 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 very small disturbance around the steady state, and our nesting result in Lemma 1 has limited application.

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: Monthly data from January 1999 to December 2019. Estimation uses iterative GMM for optimal weighting matrix. Standard errors use heteroskedastic and autocorrelation consistent (HAC) estimators and are displayed in parentheses.

- 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 7 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.[10]

Accouting for corporate deposits:

[10]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 $\tilde{r}^A_{t+1}$ denote returns on illiquid assets associated with these modifications. Direct calculation shows that it is identical to the illiquid returns r^A_{t+1} in Section 2:

$$
\begin{aligned}
\tilde{r}^A_{t+1} &:= \frac{1}{a_t}(r^K_{t+1}(q_t k_t^F - \chi(n_t + d_t)) + r^N_{t+1}(1+\chi)n_t + r^B_{t+1}\chi d_t) \\
&= \frac{1}{a_t}(r^K_{t+1}(q_t k_t^F - \chi r^K_t q_t k_t^B) + r^N_{t+1} n_t + \chi(r^K_t q_t k^B_t - r^B_{t+1}\chi d_t) + r^B_{t+1}\chi d_t) \\
&= \frac{1}{a_t}(r^K_{t+1} q_t k_t^F + r^N_{t+1} n_t) = r^A_{t+1}.
\end{aligned}
$$

  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 $y_t^h \coloneqq \frac{W_t}{P_t}h_t - \mathcal{T}_t(\frac{W_t}{P_t}h_t)$ 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 $\sigma \coloneqq -u(c)''c/u'(c)$. 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

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

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 9 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).

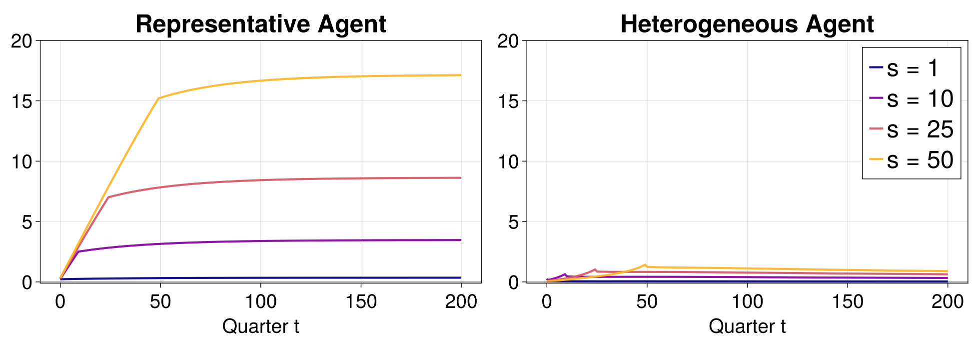

Figure 9: Entries of $-\tilde{\mathbf{B}}_{r^K}$ matrices (see Appendix A.6 for the definition)for the representative agent model with portfolio adjustment cost (left) and for the calibrated two-asset HA model (right). Each line corresponds to a different period s and shows a response of liquidity demand 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 $\boldsymbol{\epsilon}_{r^K}^{-1} = (\boldsymbol{D}_{r^K} - \boldsymbol{B}_{r^K})^{-1} \to \boldsymbol{0}$ 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 $w\hat{w}_t:=(1+r^B)\bar{b}\hat{b}_t+(1+r^A)a\hat{a}_t$, where $w:=(1+r^B)\bar{b}+(1+r^A)a$. 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 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 $a_t^G = 0$,[11] 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 3.

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 7 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,[12]

[11] 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.

[12] 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

$$\begin{aligned}
1&=\frac{1}{1+r_{t+1}^{IKC}}J'_{t+1}\left(k_t\right)\Gamma'\left(\iota_t\right),\\
J'_t\left(k_{t-1}\right)&=F_k\left(k_{t-1},h_t\right)+\frac{1}{1+r_{t+1}^{IKC}}J'_{t+1}\left(k_t\right)\left(-\Gamma'\left(\iota_t\right)\iota_t+\left(1-\delta+\Gamma\left(\iota_t\right)\right)\right).
\end{aligned}$$

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 3 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 Additional results

Figure 10 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 6 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.

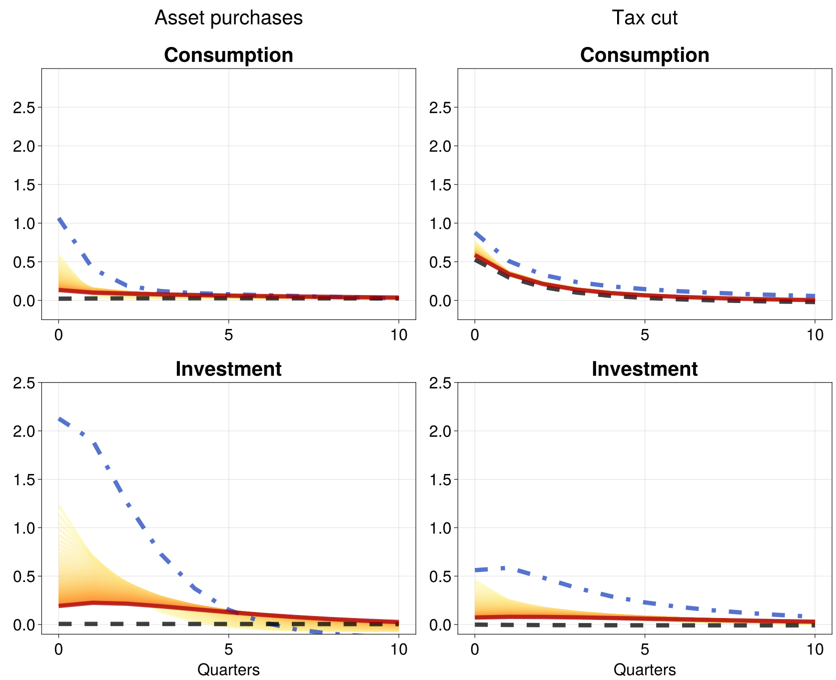

Figure 10: Consumption and investment response to asset purchases and tax cuts. The y-axis: % of steady-state GDP. Red: empirical elasticities. Light to dark yellow: low to high elasticities starting from GKK. Blue: perfectly inelastic. Black: perfectly elastic.

The large differences in investment responses are are due to firms' responses to capital prices. Figure 11 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.

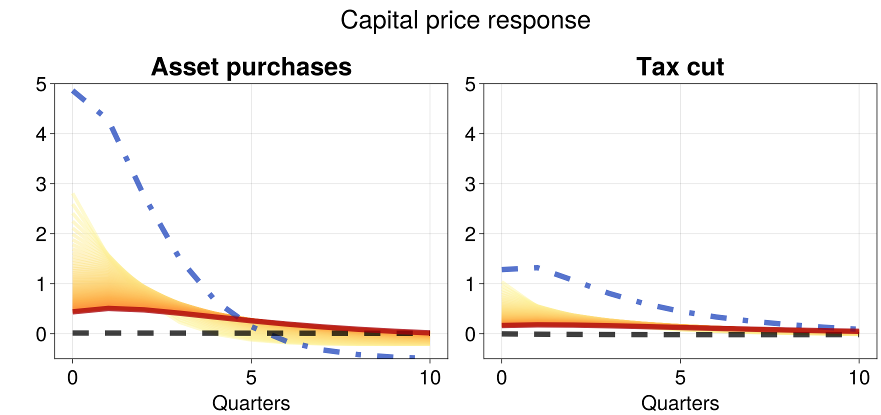

Figure 11: Capital price response to asset purchases and tax cuts. The y-axis: % of steady-state GDP. Red: empirical elasticities. Light to dark yellow: low to high elasticities starting from GKK. Blue: perfectly inelastic. Black: perfectly elastic.

Figures 12 and 13 below show output responses and decomposition as in Figures 4 and 5 in the main text, but with the y-axis rescaled to show the large magnitude of responses with perfectly inelastic liquidity supply.

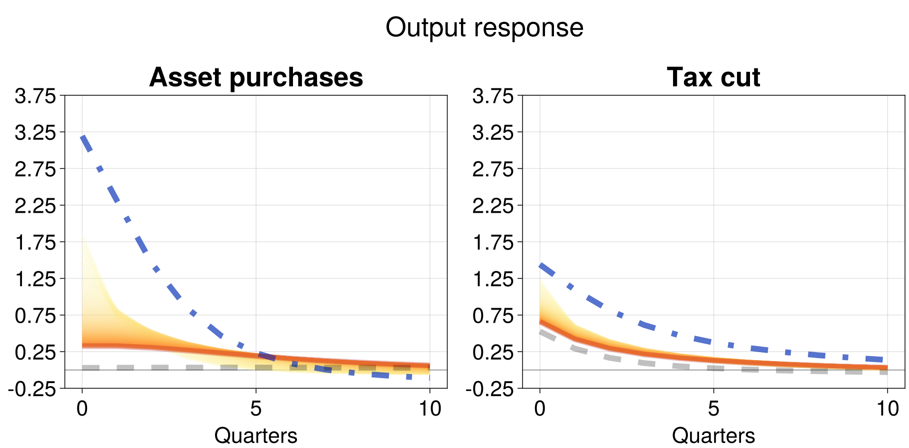

Figure 12: Output response. The y-axis: % of steady-state GDP. Red: empirical elasticities. Light to dark yellow: low to high elasticities starting from GKK. Blue: perfectly inelastic. Black: perfectly elastic.

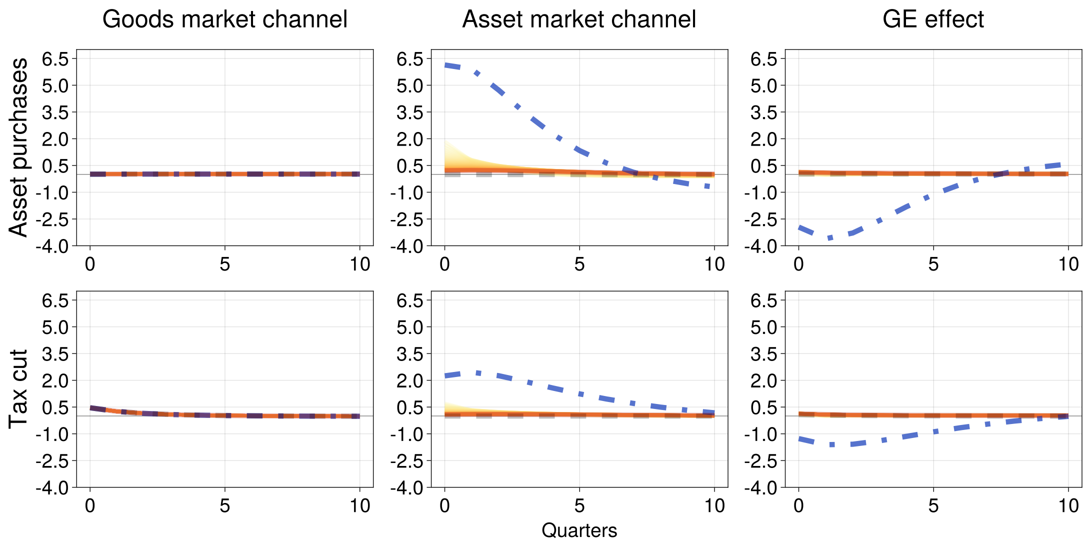

Figure 13: Decomposition of output responses. The y-axis: % of steady-state GDP. Red: empirical elasticities. Light to dark yellow: low to high elasticities starting from GKK. Blue: perfectly inelastic. Black: perfectly elastic.

E.2 Robustness Check with respect to the net worth accumulation process

We study the extent to which our policy conclusions about the relative effectiveness of policies 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 intermediation frictions, are 3.76 and 0.008.

The first panel of Figure 14 shows the relative effectiveness of asset purchases and tax cuts for different values of $\bar{G}$. The red line represents our baseline specification in Figure 6. 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 $C\left(e_t/n_t\right)$.

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 Figure 6 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 dark to light represent outcomes for reparametrizations with values farther (dark) or closer to the baseline (light).

The results of robustness check indicate that, given our estimates of $\bar{\Theta}_{r^K}$ and γ, the role of the function G is limited. This is in sharp contrast to the difference in output responses resulting from $\bar{\Theta}_{r^K}$ as shown in Figure 6.

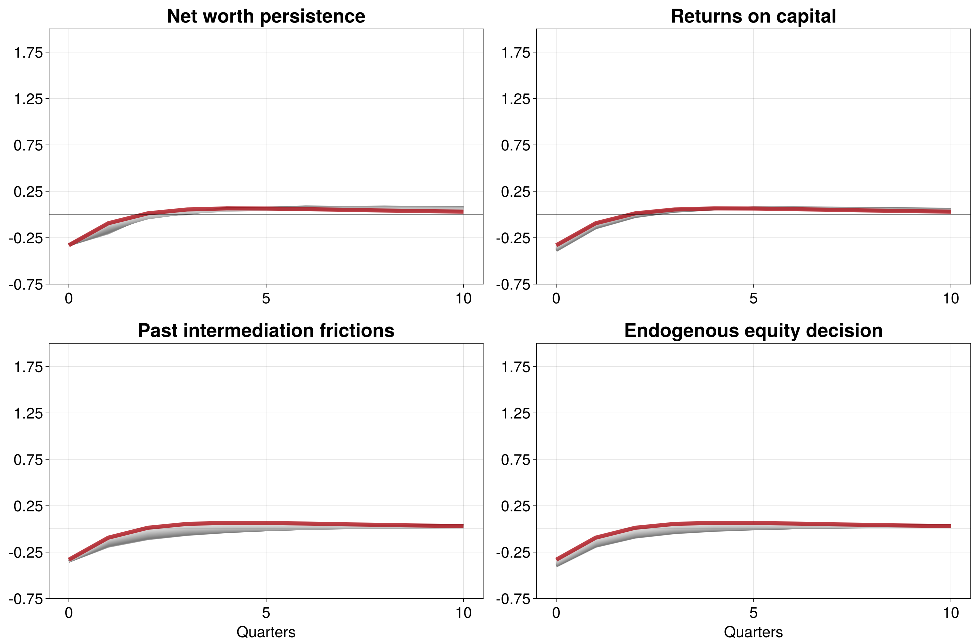

Figure 14: Difference between output response to asset purchases and tax cuts. The y-axis: % of steady-state GDP. Red: baseline. Light to dark gray: models farther and closer to the baseline – see text for the description.