Financial Intermediation and Aggregate Demand: A Sufficient Statistics Approach

Yu-Ting Chiang[†] Piotr Zoch[‡]

August 28, 2023

Click here for the most recent version.

Abstract

We derive sufficient statistics that describe how the financial sector transmits macroeconomic policies to aggregate demand. Our framework nests models of financial intermediation with various microfoundations and generates crucial aspects of aggregate demand with household heterogeneity. The financial sector supplies liquidity by issuing liquid assets to finance illiquid capital. The elasticities of liquidity supply with respect to returns are sufficient statistics that describe how the financial sector affects aggregate responses to policies. We measure the elasticities with data on price and quantity, sidestepping the difficulty of measuring microfoundations of financial frictions. Quantitatively, these elasticities are central to the debate over the effectiveness of policies targeting the financial sector versus households. In common models analyzing these policies, output responses differ by orders of magnitude due to implicit assumptions about these elasticities. Estimates of the elasticities for the U.S. imply a stronger effect of targeting households than the financial sector.

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

JEL code: E2, E6, H3, H6

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

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

We thank Anmol Bhandari, Harris Dellas, Bill Dupor, Mikhail Golosov, Lars Hansen, Greg Kaplan, Kurt Mitman, B. Ravikumar, Ricardo Reis, Ludwig Straub, Mikayel Sukiasyan, Harald Uhlig, 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

This paper derives sufficient statistics that describe how the financial sector transmits macroeconomic policies to aggregate demand. The interaction between the financial sector and the real sector is central to policies such as monetary and fiscal policies and government asset purchase programs. Yet, existing models with aggregate demand driven by realistic household consumption-saving behavior usually abstract away from the financial sector, and models of financial intermediation lack the key features on the household side. Neglecting either element can lead to incorrect conclusions about the effectiveness of policies and errors in policy decisions. Moreover, merely combining these models is not enough. To make quantitative statements about policies, we need to identify and match features of the financial sector most relevant to the interaction.

We provide a framework that nests a large class of financial intermediation models with various microfoundations and allows for rich household heterogeneity that generates crucial aspects of aggregate demand. We show that features of the financial sector relevant for aggregate responses are encapsulated by a set of sufficient statistics. These statistics can be linked directly to data on prices and quantities of assets, allowing us to sidestep the need to measure deep parameters of financial frictions specific to particular microfoundations. We show that these sufficient statistics are quantitatively important for macroeconomic policy choices. Workhorse models predict aggregate output responses that differ by orders of magnitude due to implicit assumptions about the financial sector. These assumptions manifest in differences in the sufficient statistics they imply. We measure these statistics empirically and demonstrate how they alter policy conclusions in the debate over the effectiveness of stimulating the financial sector versus households.

In our framework, households consume and save using different forms of assets, among which some are liquid and, hence, preferable to others. Households are heterogeneous due to idiosyncratic income risks and their preferences for liquidity, and they face the standard incomplete markets. Production is subject to nominal rigidities, which allows policies to affect aggregate demand. The financial sector supplies liquidity by issuing liquid assets (deposits) and holds a portfolio of illiquid capital and liquid assets (e.g., government debt). Through this process, the financial sector supplies liquid assets (deposits net of liquid asset holdings) to the economy, subject to financial frictions. We introduce a simple formulation of the financial sector that nests existing financial intermediation models with frictions resulting from asset diversion, costly state verification, costly leverage, or collateral constraints. The formulation allows us to derive our results for the class of nested models with generality. The government sets policies that take place in both the real sector, such as spending on goods, taxes, and transfers, and the asset markets, including interest rate policies, issuance of government debt, and asset purchases.

We show that all relevant details of the financial sector are summarized by a liquidity supply function with expected returns as inputs. The elasticities of liquidity supply with respect to returns are sufficient statistics that describe how the financial sector transmits policies to aggregate responses up to first-order approximation. In particular, the *cross-price elasticities* of liquidity supply with respect to returns on capital are central to the interaction between the financial sector and aggregate demand: Intuitively, these elasticities capture how much the financial sector is willing to substitute between holding liquid assets and illiquid capital. If cross-price elasticities are low, the two assets are not good substitutes. Excess liquidity (e.g., due to government debt issuance) leads to a large increase in the relative price of capital and liquid assets. Holding rates on liquid assets constant, the same excess liquidity generates higher capital prices and raises aggregate demand through investment and consumption. Through this mechanism, the financial sector influences how excess liquidity affects aggregate demand.

We measure the liquidity supply elasticities for the U.S. economy. Estimating these intertemporal elasticities non-parametrically is challenging because these are infinite-dimensional objects. Yet, for the models nested in our framework, we show these elasticities have a special structure controlled by a few parameters and observable steady-state variables. We exploit this structure to calculate semi-structural estimates of the elasticities, using data on bank balance sheets, market valuations of bank equity, and yield curves. Our estimate of the liquidity supply semi-elasticity with respect to return on capital one quarter ahead is around 8. The estimate is twice as large as implied by standard models of financial intermediation and far from the two polar cases of perfectly elastic and inelastic supply, which are commonly assumed in workhorse heterogeneous-agent models.

Differences in the liquidity supply elasticities have quantitatively important policy implications. These elasticities are central to the "Wall Street vs. Main Street" debate: whether transferring resources to the financial sector is more effective in stimulating the economy than transferring resources to households. Consider two alternative policies: both involve issuing government debt of 1% steady-state annual GDP, but one policy directs resources to buy illiquid assets while the other uses the proceeds for tax cuts, holding other policy instruments fixed in both cases. Cumulative first year output responses relative to steady-state annual GDP range from 1.3% to 0.02% for asset purchases and from 0.7% to 0.3% for tax cuts among workhorse models heterogeneous-agent models, featuring liquidity supply from perfectly inelastic to perfectly elastic. These differences are quantitatively significant and lead to qualitatively different predictions: The relative effectiveness of policies, as measured by the gap between output response to asset purchases and tax cuts, ranges from 0.6% to −0.25%, with standard models of financial intermediation generating a gap of 0.2% due to implicit assumptions about the liquidity supply elasticities. Our estimates for the U.S. economy imply a gap of −0.05%, indicating a stronger effect of targeting the household sector than the financial sector.

Finally, to isolate how the financial sector affects the transmission of policies, we decompose output responses into three channels: (1) A goods market channel that captures the direct effect of policies on aggregate demand, such as consumption responding to tax cuts. (2) An asset market channel: as policies shift liquidity demand and supply, the financial sector transmits excess liquidity into capital prices and aggregate demand. (3) A Keynesian cross in which feedback between aggregate income and aggregate demand is modified by the financial sector. The strength of these channels determines the relative effectiveness of policies as policy instruments affect the goods and assets markets differently. We show that policy predictions diverge among the range of models considered almost entirely due to the asset market channel.

Literature

Our work is related to an extensive literature that emphasizes the importance of household heterogeneity in understanding the effects of macroeconomic policies (e.g., Gornemann et al. (2012), McKay et al. (2016), Kaplan et al. (2018)). Our approach is closest to a strand of this literature that provides theoretical characterization of aggregate responses to policies in the sequence space, e.g., Auclert et al. (2023), Auclert et al. (2021) and Wolf (2021b), Dávila and Schaab (2023), McKay and Wolf (2023), Koby and Wolf (2020) and Wolf (2021a). These papers abstract from financial intermediation. We provide a framework to study how to aggregate responses in these models depend on the financial sector, derive sufficient statistics to summarize its role, and demonstrate that the financial sector is crucial for policy analysis qualitatively and quantitatively.

Models of financial intermediation nested in our framework include 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), among other numerous variations of these models. We summarize intermediation frictions in these models with sufficient statistics and demonstrate that they are quantitatively important for policy choices. Our emphasis on liquidity and the feedback between goods and assets markets is similar in spirit to Kiyotaki and Moore (2019), and our estimation of liquidity supply elasticity is related to the study of aggregate demand for Treasury debt by Krishnamurthy and Vissing-Jorgensen (2012).

Our paper complements recent works that incorporate frictional financial intermediation into heterogeneous-agent models (e.g., Lee et al. (2020), Fernández-Villaverde et al. (2020), Lee (2021), Mendicino et al. (2021), Schroth (2021), Ferrante and Gornemann (2022)). These papers are quantitative in nature. Our theoretical approach identifies the liquidity supply elasticities as key objects that govern the interaction between the financial sector and the real sectors. We provide empirical measures of these elasticities, which are 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 $i \in [0, 1]$. Preferences are time separable, and the future is discounted with factor $\beta_i \in (0, 1)$. Household i derives utility from final good consumption $c_{i,t}$, disutility from labor $h_{i,t}$. Households can save in liquid and illiquid assets, $b_{i,t}$ and $a_{i,t}$, which pay real returns r_t^B and r_t^A respectively. Trading of illiquid asset $a_{i,t}$ incurs portfolio adjustment costs, captured by a function $\Phi_t\left(a_{i,t}, a_{i,t-1}\right)$. Each household solves the following maximization problem:

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

subject to budget constraints

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

where $y_{i,t} = \frac{W_t}{P_t} z_{i,t} h_{i,t}$ denotes the 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 $z_{i,t}$.

2.2 Production

Final goods and Capital

A representative firm produces final good y_t with capital k_{t-1} and differentiated types of labor $h_{\ell,t}, \ell \in [0,1]$:

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

where $h_{\ell,t}$ is supplied by labor union ℓ, and $\varepsilon_W > 1$ is the elasticity of substitution between labor types. The firm maximizes profit, taking wages $\{W_{\ell,t}\}$ and rental rate of capital R_t as given:

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

Capital is held by a mutual fund and a bank, $k_t = k_t^F + k_t^B$. Over time, capital evolves according to

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

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

$$1 + r^K_{t+1} = \max_{\hat{\iota}_{t+1}} \frac{R_{t+1}/P_t + q_{t+1}\left(1 + \Gamma\left(\hat{\iota}_{t+1}\right) - \delta\right) - \hat{\iota}_{t+1}}{q_t}. \tag{1}$$

Labor supply

There is a continuum of labor unions indexed by $\ell \in [0, 1]$. Every household i provides $h_{i,\ell,t}$ units of labor to the unions: $h_{i,t} = \int h_{i,\ell,t} d\ell$. Each union aggregates labor from households into union-specific labor services: $h_{\ell,t} = \int z_{i,t} h_{i,\ell,t} di$.

Labor unions are monopolistically competitive and set nominal wages $\{W_{\ell,t}\}$ with growth rate $\pi_{W,\ell,t} := \frac{W_{\ell,t}}{W_{\ell,t-1}} - 1$, subject to a quadratic adjustment cost to maximize utilitarian welfare of the households:

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

The level of nominal rigidity is parameterized by $\kappa_W > 0$. Wage adjustment cost is borne as disutility by the labor union and does not enter the resource constraint. Given labor demand, income of household i is given by: $W_t z_{i,t} h_{i,t} = \int W_{\ell,t} z_{i,t} h_{i,\ell,t} d\ell$, where W_t is the ideal wage index.

2.3 The Financial Sector

A representative bank issues deposits to finance liquid assets and illiquid capital holdings. At the time t, given net worth n_t, the bank issues deposits $\tilde{d}_t$, and chooses capital and liquid asset holdings, k^B_t and b^B_t. We assume deposits and other liquid assets (government debt) are perfect substitutes and pay the same real rate of return r^B_t. The bank's liquidity supply d_t is defined as the difference between its liquid asset issuance and holdings:

$$d_t := \tilde{d}_t - b^B_t.$$

The bank chooses capital holdings k_t^B and liquidity supply d_t to maximize its flow return r_{t+1}^N, solving the following problem

$$r_{t+1}^N n_t = \max_{k_t^B,\ d_t} r_{t+1}^K q_t k_t^B - r_{t+1}^B d_t \qquad (\mathcal{P})$$

subject to its balance sheet and a financial constraint:

$$q_t k_t^B = d_t + n_t, \quad q_t k_t^B \leq \Theta\left(\left\{r_{s+1}^K, r_{s+1}^B\right\}_{s\geq t}\right) n_t.$$

The bank allows households to finance capital without incurring portfolio adjustment costs Φ_t when households need to liquidate assets quickly. This captures how banks perform *liquidity transformation* in the economy. Bank's ability to fund capital by issuing liquid assets is limited by the financial constraint. The degree of financial friction potentially depends on the entire path of future returns r_s^B and r_s^K, which reflects the future funding cost and investment opportunities in the economy. This specification of the financial constraint allows us to nest a class of frictional financial intermediation models as special cases. We discuss this nesting property in Section 3.1 and Appendix C.1.

We assume the bank follows an exogenous rule that pays out a fraction f of the accumulated net worth as dividends and receives a constant equity injection m from the fund. The net worth of the banking sector evolves according to

$$n_{t+1} = (1 - f)n_t(1 + r_{t+1}^N) + m. \tag{2}$$

We generalize the net worth process in Appendix E.1 to allow for various forms of state-dependent equity injection as in Gertler and Kiyotaki (2010) and Karadi and Nakov (2021).

Illiquid assets holdings

The illiquid assets are held as a passive mutual fund, a_t. The fund consists of the net worth of the bank n_t and capital of value $q_t k_t^F$. The balance sheet of the fund is given by

$$a_t = q_t k_t^F + n_t,$$

and the rate of return on illiquid assets is

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

2.4 Government

Government policies are described by government purchases g_t, tax rate τ_t, liquid government debt b_t^G, and illiquid assets holdings a_t^G, and liquid rate targets $r_t^B = r_t$ for all $t > 0$. We assume that the government debt is real and that the government sets the nominal interest rate i_t^B to keep the real rate r_t^B at its target, following Woodford (2011). The liquid rate in period 0 is predetermined and equals $\bar{r}^B$. The tax revenue collected by the government is $T_t = \int \mathcal{T}(y_{i,t})di$. 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 Equilibrium definition

Given $\{g_t, \tau_t, b_t^G, r_t\}$, an equilibrium consists of prices $\{q_t, P_t, R_t, W_{\ell,t}, r_t^A, r_t^B, r_t^K\}$ 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 budget constraints; (2) firms maximize profit and investment rate maximizes the return on capital, (3) nominal wages maximize payoff of the labor unions; (4) the bank maximizes return on net worth subject to its financial constraint and balance sheet, and net worth follows its law of motion; (5) the illiquid return r^A is given by the balance sheet of the mutual fund; (6) the government budget constraint holds, and (7) markets clear:

$$\int c_{i,t} + \Phi_t\left(a_{i,t}, a_{i,t-1}\right) di + x_t + g_t = y_t, \quad \int b_{i,t} di = d_t + b_t^G, \quad \int a_{i,t} di + a_t^G = q_t k_t - d_t,$$

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 bank and the government; and (iii) in the illiquid asset market, the fund net worth is equal to the total of household and government's holdings of illiquid assets. The capital market clears when capital holdings of the bank and the fund equal the aggregate stock of capital, $k_t^F + k_t^B = k_t$. We focus on an equilibrium in which the financial constraint of the bank is always binding. Figure 1 summarizes the balance sheets of agents in the economy.

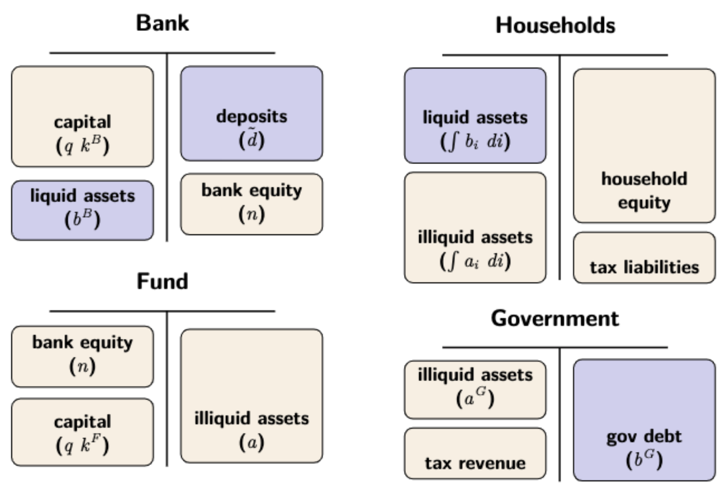

Figure 1: Balance sheets of agents in the economy. Banks supply net liquid assets $d_t = \tilde{d}_t - b_t^B$. Together with liquid government debt b_t^G, they add up to household liquid asset holdings $\int b_{i,t}\ di$.

3 Liquidity Supply

3.1 Nesting models of financial intermediation

The financial sector in our framework issues liquid assets to finance illiquid capital, subject to a financial constraint. We now show that the financial sector in Section 2.3 nests a large class of models that feature financial intermediaries with various objective functions and face different constraints. These models share a special structure that allows us to characterize the elasticities of the financial sector's liquidity supply, which are sufficient for understanding the first-order approximation of aggregate responses. We provide an overview of the models nested in our framework below and layout details of these models in Appendix C.1.

Model 1, asset diversion (Gertler and Kiyotaki (2010), Gertler and Karadi (2011)): Bankers in these models can divert a fraction $1/\theta$ of assets. If this happens, depositors force a bank into bankruptcy. In order to ensure that a banker is better off continuing instead of diverting assets, the funding a bank can receive from depositors depends on its continuation value $v_t(n_t) = \eta_t n_t$:

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

where $\Lambda_{t,t+1}$ denotes a banker's discount factor.[1]

Model 2, costly state verification (Bernanke et al. (1999)): Banks receive idiosyncratic returns on their assets, which the lenders can only observe by incurring a monitoring cost. The bank's capital holdings are linked to its net worth and expected returns:

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

where ψ^{BGG} is an increasing function determined by the distribution of idiosyncratic returns and the monitoring cost.

Model 3, costly leverage (Uribe and Yue (2006), Eggertsson et al. (2019), Chi et al. (2021) and Cúrdia and Woodford (2011)): Banks need to incur a convex cost $\Upsilon\left(\frac{q_t k_t^B}{n_t}\right) n_t$ that depends on the level of financial intermediation. The optimal leverage is linked to the spread between returns on capital and deposits:

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

Model 4, collateral constraint (similar to Kiyotaki and Moore (1997), Bianchi and Mendoza (2018), Ottonello et al. (2022)): Liquidity supplied by the bank is limited by the value of collateral backing it. For example, if the value of collateral includes the market price of capital next period plus the rental rate net of user cost:[2]

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

The models described above are nested by the bank's Problem $\mathcal{P}$ in Section 2.3, as stated in the following lemma:

[1] We allow the discount rate to be $(1 + r_{t+1}^B)^{-1}$ or $(1 + r_{t+1}^K)^{-1}$. In fact, our analysis holds for any function of the two returns.

[2] Among models with collateral constraints, the exact form of constraints differs due to assumptions about what can be pledged as collateral. For example, in the original version of Kiyotaki and Moore (1997), the value of collateral contains only the market price of capital. We discuss different variations in Appendix C.1.

Lemma 1 *Suppose that* $\{d_t^j, n_t^j\}$ *solves the bank's problem in Model* $j \in \{1, \dots, 4\}$*. There exists a function* $\Theta_t := \Theta(\{r_{s+1}^K, r_{s+1}^B\}_{s \geq t})$ *such that* $\{d_t^j, n_t^j\}$ *is the solution to Problem* $\mathcal{P}$*. 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 determined by parameters of Model* j *and steady-state variables.*

Proof. See Appendix C.1. □

The microfoundations of each model map into different $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}, \gamma$. For models that feature asset diversion, $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}, \gamma > 0$ capture sensitivity of a banker's continuation value to the two returns at various horizon. The sensitivity depends on the assumption about what a banker can do with the diverted asset. In this class of models, $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}, \gamma$ are determined by the steady-state levels of leverage and returns, and there is no extra parameter in microfounded models to govern them. These models impose a tight connection between the steady-state leverage and sensitivity: $\bar{\Theta}_{r^K}$, $\bar{\Theta}_{r^B}$, and γ are strictly increasing in steady-state leverage.

In models that feature costly state verification and costly leverage, $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B} > 0$ and $\gamma = 0$. In costly state verification models, $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}$ are linked to the distribution of idiosyncratic returns in the steady state and the monitoring cost; whereas, in costly leverage models, $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}$ are determined by the curvature of the leverage cost function at the steady state. In these models, there are extra parameters that govern the sensitivities, $\bar{\Theta}_{r^K}$ and $\bar{\Theta}_{r^B}$, separately from the steady-state leverage and returns. However, financial constraints do not respond to expected rates more than one period ahead: Θ_t does not respond to r_{s+1}^K and r_{s+1}^B for $s > t$.

Finally, for collateral constraints nested in our framework, $\gamma = 0$ because changes in the value of the collateral are captured by changes in r_{t+1}^K. Depending on whether the constraints involve only the current value of assets or also their next period returns, we have $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B} = 0$ or $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B} > 0$.

3.2 Liquidity Supply Elasticities

Frictions in the financial sector determine how liquidity supply responds to returns. To summarize this mapping, we define the liquidity supply function, $\mathcal{D}_t(\{r_s^K; r_s^B\}_{s=0}^{\infty})$, as the solution d_t of the bank's Problem $\mathcal{P}$ given $\{r_s^K; r_s^B\}_{s=0}^{\infty}$. The response of liquidity supply to changes in returns is described by two sets of semi-elasticities: the *own-price* and *cross-price* semi-elasticities of liquidity supply.

Lemma 2 *The own- and cross-price semi-elasticities of liquidity supply at the stationary equilibrium are given by:*

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

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

where $G := (1-f)\left[\left(\bar{r}^K - \bar{r}^B\right)\bar{\Theta} + \left(1+\bar{r}^B\right)\right], \quad \Sigma(s) := (1-f)(\bar{r}^K - \bar{r}^B)\frac{1-(\gamma G)^s}{1-\gamma G}.$

Proof. See Appendix A.1. □

These intertemporal elasticities are infinite-dimensional objects where each (t, s) pair captures the response of liquidity in time t to changes in returns at time s. Depending on the relative timing of t and s, the formulas for the cross-price semi-elasticities are split into two cases: (1) For $s \le t$, changes in returns have no direct effect on Θ_t. Liquidity supply in period t is affected only through net worth accumulation in the past. An increase in r_s^K increases net worth in period s and relaxes the constraints $\Theta(\cdot)$ in all periods before period s, as captured by the function $\Sigma(s)$. These effects propagate forward from period s to period t through net worth, which declines at the rate G due to dividend payout f. (2) For $s > t$, an increase in r_s^K directly affects the constraint Θ_t. Moreover, it relaxes all financial constraints before period t, which further increases liquidity supply in period t through net worth accumulation, as captured by the same function $\Sigma(t)$. The intuition is similar for the own-price elasticities, $\frac{\partial \mathcal{D}_t / \partial r_s^B}{\mathcal{D}_t}$. In Appendix E.1, we show that the liquidity supply elasticities take a similar form when the bank's net worth process features state-dependent equity injections.

The cross-price and own-price semi-elasticities of liquidity supply are, respectively, positive and negative. Larger values of $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}$ and γ correspond to larger semi-elasticities (in absolute values): the cross-price elasticities are increasing in $\bar{\Theta}_{r^K}$ and γ, whereas the own-price elasticities are decreasing in $\bar{\Theta}_{r^B}$ and γ. Within the class of models we study, these three parameters control the infinite-dimensional intertemporal elasticities. This simple structure allows us to perform comparative statics and systematically compare how different features of financial intermediation affect aggregate responses to policies.

4 Aggregate Responses to Policies

4.1 A Demand-and-Supply Representation

We recast the aggregate behavior of agents as the equilibrium of a demand-and-supply system.[3] Given prices and government policies, 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, $\mathcal{D}_t$. The same logic applies to the *household block* of the model: Given a sequence of output, taxes, returns on assets, and the initial asset distribution, we can solve the households' consumption-saving problem to obtain an aggregate consumption function, $\mathcal{C}_t$, and an aggregate liquid asset demand function $\mathcal{B}_t$.[4] Similarly, we obtain an aggregate investment function, $\mathcal{X}_t$, from the *production block*. Lemma 3 represent the equilibrium as the solution to a demand-and-supply system of these aggregate functions.

Lemma 3 *Given government policies* $\left\{g_s, T_s, r_s^B, b_s^G\right\}_{s=0}^{\infty}$, *there exist functions* $\mathcal{C}_t, \mathcal{B}_t$, *and* $\mathcal{X}_t$, *such that the equilibrium output and returns on capital* $\left\{y_s, r_s^K\right\}_{s=0}^{\infty}$ *solve the*

[3] Aguiar et al. (2021), Auclert et al. (2023), Auclert et al. (2021), and Wolf (2021a) use a similar representation.

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

following system:

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

and

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

where returns on iliiquid assets, r_t^A, are given by the function $\mathcal{R}_t^A$ derived from the accounting identity in Equation 3, and the government illiquid asset holdings $\{a_t^G\}$ satisfy the government budget constraint in Equation 4. Moreover, functions $\mathcal{C}_t, \mathcal{B}_t$, and $\mathcal{X}_t$ do not depend on specifications of the financial sector, such as the financial frictions represented by the function $\Theta(\cdot)$.

Proof. See Appendix A.2. □

The two main equations in Lemma 3 correspond to the goods market and the liquid asset market clearing conditions.[5] Given government policies, an equilibrium is described by sequences $\{y_t, r_t^K\}_{t=0}^{\infty}$ such that (1) aggregate demand for final goods equals output produced, and (2) aggregate liquid asset demand that equals the supply of liquid assets.

The demand-and-supply formulation allows us to separate different blocks of the model: On one hand, functions $\mathcal{C}_t, \mathcal{B}_t$, and $\mathcal{X}_t$ contain all relevant information about household heterogeneity and the production sector; these functions do not depend on the characteristics of the financial sector. On the other hand, the financial sector enters the system only through its liquidity supply, $\mathcal{D}_t$. *All* relevant properties of the financial sector are summarized by the function $\mathcal{D}_t$: As far as aggregate dynamics are concerned, details of the microfoundations matter only insofar as they imply different a liquidity supply function. Since all relevant information about the financial sector is contained in $\mathcal{D}_t$, the elasticities of liquidity supply are sufficient statistics that summarize how the financial sector affects the transmission of policies to aggregate outcomes, up to first-order approximation. In the class of models nested

[5]We can reduce the system to the market clearing conditions of the goods market and the liquid asset market because the illiquid asset market clearing condition is redundant by Walras' law. In principle, one can reformulate Lemma 3 with any two of the three markets.

in our framework, these sufficient statistics take a simple structure controlled by parameters $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}, \gamma$, which respectively govern the sizes and forward-lookingness of the elasticities. This simple structure reduces the infinite-dimensional liquidity supply elasticities to several parameters, which allows us to systematically compare how financial intermediation affects aggregate responses to policies.

Equilibrium Approximation

We consider perturbations of government policies around the steady state, such that policy variables $\{dg_t, dT_t, dr_t^B, db_t^G, da_t^G\}_{t=0}^{\infty}$ satisfy government budget constraints and converge to zero as $t \to \infty$. We focus on the equilibrium for which first-order deviations of all endogenous variables converge to zero. To simplify notation, we use a column vector $\boldsymbol{y}$ to represent $\{y_t\}_{t=0}^{\infty}$, the sequence of output, and use $d\boldsymbol{y}$ for its first-order deviation. Notation for $\boldsymbol{T}, \boldsymbol{b}^G$, $\boldsymbol{g}$ is similar. We use $\boldsymbol{r}^K$ to represent $\{r_{t+1}^K\}_{t=0}^{\infty}$, the sequence of rates of return on capital, and use $d\boldsymbol{r}^K$ for its first-order deviation; notation for liquid rates $\boldsymbol{r}^B$ follows the same convention.[6]

Useful Notation

We define *excess liquidity* for the liquid asset market as

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

where $r^A(\boldsymbol{y}, \boldsymbol{r}^K, \boldsymbol{r}^B)$ expresses the sequence of illiquid returns as a function of output, returns on capital, and liquid rates, using the accounting identity from Equation 3, as detailed in Appendix B.1. We use $\boldsymbol{\epsilon}$'s to denote the derivatives of excess liquidity with respect to its arguments: $\boldsymbol{\epsilon}_{r^K}$ is a matrix with each row corresponding to a different period t: $\boldsymbol{\epsilon}_{r^K}(t, \cdot) \coloneqq \frac{\partial}{\partial \boldsymbol{r}^K} \mathcal{E}_t(\cdot)$. Derivatives $\boldsymbol{\epsilon}_{r^K}, \boldsymbol{\epsilon}_{r^B}$ describe how excess liquidity responds to returns. They are directly linked to the cross- and own-price elasticities of liquidity supply we characterized in Section 3.2.

Similarly, we use Ψ_t to represent the *aggregate demand* for the goods market

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

[6] The sequences for returns start from period 1 because the initial liquid rate, r_0^B, is predetermined, and we can solve the initial period realized return on capital as a function of output and expected returns, $r_0^K(\boldsymbol{y}, \boldsymbol{r}^K)$.

Derivatives such as $\boldsymbol{\Psi}_{r^K}$ capture how aggregate demand responds to aggregate income $d\boldsymbol{y}$, returns on capital $d\boldsymbol{r}^K$, and government policies. For example, the row of $\boldsymbol{\Psi}_{r^K}$ corresponding to time t is given by $\boldsymbol{\Psi}_{r^K}(t,\cdot) \coloneqq \frac{\partial}{\partial \boldsymbol{r}^K}\Psi_t(\cdot)$.

4.2 Aggregate Responses

We characterize the equilibrium in two steps. First, we study how returns on capital $d\boldsymbol{r}^K$ must adjust to clear the liquid asset market given government policies and aggregate output $d\boldsymbol{y}$. We then use the solution for $d\boldsymbol{r}^K$ as a function of $d\boldsymbol{y}$ and government policies $d\boldsymbol{b}^G$, $d\boldsymbol{T}$, and $d\boldsymbol{r}^B$ to find the path of aggregate output that satisfies the goods market clearing condition.

Excess Liquidity and Asset Markets Responses

An equilibrium in the liquid asset market is reached when the liquid asset demand from the households equals liquid assets supplied by the financial sector and the government. Proposition 1 shows how returns on capital respond to shifts in excess liquidity due to exogenous policies.

Proposition 1 *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, for $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B} \to \infty$ *with* $\bar{\Theta}_{r^B}/\bar{\Theta}_{r^K} \to \varsigma$, *we have* $d\boldsymbol{r}^K = \varsigma d\boldsymbol{r}^B$.

Proof. See Appendix A.3. □

Intuitively, an increase in excess liquidity (e.g., due to an increase in $d\boldsymbol{b}^G$) pushes up the relative price between capital and liquid assets, reflected as a decrease in spread between $d\boldsymbol{r}^K$ and $d\boldsymbol{r}^B$. Given $d\boldsymbol{r}^B$ targeted by monetary policy, an increase in the relative price between capital and liquid assets leads to an increase in the price of capital, q_t. The magnitude of the increase in the price of capital depends on the cross-price elasticity of liquidity supply through $\boldsymbol{\epsilon}_{r^K}$. Intuitively, if the financial sector's liquidity supply is inelastic in period t, the two assets are not good substitutes, and a large decrease in expected returns on capital r^K_{t+1} is required for banks to increase their liquid asset holdings and decrease their liquidity supply.

In the limiting case with perfectly elastic liquidity supply ($\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B} \to \infty$), 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}. Because assets are perfect substitutes, the asset markets are no longer segmented. The perfect link between asset markets allows monetary policy to directly control the returns on capital with liquid rates. As we discuss in Appendix E.2, this limiting case corresponds closely to Auclert et al. (2023).

Aggregate Output Responses

Aggregate output responses to government policies depend on the financial sector through the liquid asset market. We totally differentiate the demand and supply functions in the goods market clearing condition and use the expression for returns on capital, $d\boldsymbol{r}^K$, from Proposition 1 to characterize the equilibrium aggregate output response, $d\boldsymbol{y}$.

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

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

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

Proof. See Appendix A.4. □

Aggregate output responds to government policies through three channels: (1) The goods market channel shows how government purchase, tax, and liquid rate directly affect aggregate demand in the goods market. (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 that captures the feedback between aggregate income and aggregate demand through the goods and asset markets.

The asset market channel (channel 2) shows how the asset market logic described in Proposition 1 translates into aggregate output responses, and it is captured by the two components of matrix $\boldsymbol{\Omega}$. Consider an increase in liquid government debt $d\boldsymbol{b}^G$. Proposition 1 shows that if an entry in matrix $(-\boldsymbol{\epsilon}_{r^K})^{-1}$ is negative, an increase in excess liquidity leads to a decrease in expected returns on capital $d\boldsymbol{r}^K$. The lower cross-price elasticities (smaller $\bar{\Theta}_{r^K}$), the stronger the response of rates of return on capital. On the other hand, matrix $\boldsymbol{\Psi}_{r^K}$ describes how changes in returns on capital affects aggregate demand. For example, if a decrease in expected return on capital in period s leads to higher capital prices and increases investment and consumption in period t, then the corresponding entry of $\boldsymbol{\Psi}_{r^K}$ is negative. In this case, an increase in excess liquidity generates higher aggregate demand through lower expected returns on capital and higher capital prices.

The same mechanism works in channel (3), although it serves as a force that modifies the traditional Keynesian cross logic. When aggregate income increases, households demand more liquid assets which decreases excess liquidity. The same logic as in channel (2) implies that an increase in liquid asset demand leads to higher expected returns on capital and lower capital price, which decreases aggregate demand through investment and consumption. Therefore, a positive entry in $\boldsymbol{\Omega}$ is associated with a dampening force to the Keynesian cross logic, and the dampening force is more substantial with lower cross-price elasticities.

Policy Comparison: Asset Purchases v.s. Tax Cuts

We apply Theorem 1 to study policies targeting different sectors of the economy. Consider two alternative policies in which the government issues the same amount of debt $\{b_t^G\}$, monetary policy targets the same path of liquid rate $\{r_t^B\}$, and government purchase follows the same path $\{g_t\}$. One policy targets the financial market, and the government specifies a path $\{\Delta_t\}$ of net illiquid asset purchases (or sales): $\Delta_t = a_t^G - (1+r_t^A)a_{t-1}^G$, and collects tax revenue T_t to balance the budget. The other policy targets the household sector, and the government pays the same amount, $\{\Delta_t\}$, to households as (gross) tax cuts instead of purchasing assets.

Let $\widehat{d\boldsymbol{y}} := d\boldsymbol{y}^{\text{asset}} - d\boldsymbol{y}^{\text{tax cut}}$ be differences in output responses between the two policies. Theorem 1 immediately implies

Corollary 1 *Given any $\{d\boldsymbol{b}_t^G, d\boldsymbol{r}_t^B, d\boldsymbol{g}_t\}$, the differences between aggregate output responses to government asset purchases and tax cuts are given by*

$$\widehat{d\boldsymbol{y}} = \left(\mathbf{I} - \boldsymbol{\Psi}_y - \boldsymbol{\Omega}\ \boldsymbol{\epsilon}_y\right)^{-1} \times \left(\boldsymbol{\Psi}_T d\boldsymbol{\Delta} + \boldsymbol{\Omega}\ \boldsymbol{\epsilon}_T d\boldsymbol{\Delta}\right).$$

Tax cuts affect aggregate demand directly through the goods market (channel 1), while asset purchases do not. This difference in the goods market is captured by $\boldsymbol{\Psi}_T d\boldsymbol{\Delta}$. Yet, asset purchases create more excess liquidity and lead to a stronger response through the asset market channel. The difference in excess liquidity is given by $\boldsymbol{\epsilon}_T d\boldsymbol{\Delta}$. The relative strength between the two channels depends on the elasticities of liquidity supply: the goods market channel is fully determined by household consumption response to tax cuts and does not depend on the elasticities of liquidity supply.[7] On the other hand, if liquidity supply is elastic, the asset market channel is weakened, and excess liquidity has little effect on aggregate demand. As the two policies operate through different channels, assumptions about the financial sector lead to different conclusions about which one is more effective.

5 Taking the Model to the Data

In this section, we take the model to the data to prepare for a quantitative assessment of how the financial sector affects aggregate responses to policies. We consolidate household balance sheets into holdings of liquid and illiquid assets and develop a mapping between the liquid asset positions in our model and those of the U.S. economy. Next, we estimate the three parameters that govern the financial sector's liquidity supply elasticities, using information about the banking sector balance sheet, the market value of banks, and yield curves on Treasury and corporate bonds. Finally, we discuss our calibration for the rest of the model.

5.1 Asset Classification and Balance Sheets

We categorize liquid assets to encompass deposits (such as checkable, time, savings account, and money market fund shares) and government debt (including 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. All assets fall on a spectrum in terms of liquidity. When mapping the model to the data, we must draw a line to classify some assets as liquid and label all other assets as illiquid. We do not think trading of illiquid assets neces-

[7] More generally, the goods market channel in Theorem 1 will also be independent of assumptions about the financial sector if $d\boldsymbol{r}^B = 0$.

sarily involves a large transaction cost, but simply that they lack certain features we described above.

We obtain the household sector's balance sheet from the Flow of Funds data. Households' liquid asset holdings mostly consist of deposits (72%) and money market funds shares (17%). To measure the balance sheets of the banking sector, we use the Call Report data filed by depository institutions, which we link to the CRSP data to obtain the market value of the net worth of the banking sector. We adjust banks' balance sheets 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 Appendix D, we discuss the details of the mappings between the model and the data, including how we can extend the model to account for the liquid assets held by the corporate sector without affecting the equilibrium of the model.

Table 1: Consolidated Balance Sheets

	assets		liabilities	
households	liquid assets	0.58		
	net illiquid assets	3.35		
			equity	3.93
banks & mmf	liquid assets	0.14		
	capital	0.52		
			liquid liabilities	0.53
			equity	0.13

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 2000Q2 to 2020Q2.

Table 1 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 39% of GDP and accounts for 67% of liquid assets held by households. Table 3 paints a picture that is in contrast to many workhorse heterogeneous-agent models that study monetary and fiscal policies. Most heterogeneous-agent models emphasize the role of liquid assets in households' consumption-saving behavior, yet many of them abstract away from the financial sector and assume all liquid assets are supplied by the government (e.g. Kaplan et al. (2018)). Since the financial sector is an important supplier of liquid assets, it is natural to suspect that its response will be a quantitatively important factor for aggregate responses to excess liquidity created by government policies.

5.2 Elasticities of Liquidity Supply

The own- and cross-price elasticities of liquidity supply are sufficient statistics that summarize all relevant features of the financial sector. In principle, these elasticities can be estimated nonparametrically, but the implementation is challenging because these elasticities are infinite-dimensional objects. We take a semi-structural approach instead: We measure these elasticities using the special structure derived in Section 3 for the class of models nested in our framework. Our main estimation aims to recover $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}$, and γ from the data. These three parameters determine the relationship between returns and effective leverage (net supply of liquid assets divided over bank net worth), and they are directly connected to various canonical theoretical microfoundations. These parameters are semi-structural in the sense that they allow us to summarize relevant features of the financial sector without taking a stance on the exact microfoundation. Yet, they are invariant to the set of macroeconomic policies that we study, as these policies take effect through quantities and prices in the intertemporal demand-supply system. Estimates of these parameters provide a concise empirical summary of the relevant features of the financial sector, and they are useful empirical moments for guiding modeling choices when building models of financial intermediation with detailed microfoundations.

The empirical counterpart of the relationship between the effective leverage and returns implied by Lemma 1 is

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

The assumption required for identification of $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}, \gamma$ is that υ_t satisfies

$$\mathbb{E}[\upsilon_t] = \mathbb{E}\left[\upsilon_t \times \mathbb{E}_t\left[dr^K_{t+h}\right]\right] = \mathbb{E}\left[\upsilon_t \times \mathbb{E}_t\left[dr^B_{t+h}\right]\right] = 0, \quad \forall t, h. \tag{6}$$

These moment conditions form the basis of our estimation strategy. The identification assumption underlying our empirical strategy is that the effective leverage of the financial sector Θ_t responds to aggregate shocks only through its response to changes in returns. We assume that leverage is a purely endogenous choice, and there are no exogenous shocks to leverage choices. We thus attribute the difference between the effective leverage seen in the data and that implied by Lemma 1, v_t, as measurement errors. Note that assuming away leverage shocks does not preclude exogenous shocks to liquidity supply: there can be shocks to the net worth of the financial sector either directly through exogenous changes in equity injections m or through realizations of returns.[8] Our underlying assumption is that shifts in liquidity supply must take the form of shifting banks' net worth.[9] In Appendix D.3 we discuss consequences of violating the identifying assumption. Because a positive shock to effective leverage increases bank lending and drives down future spreads, our estimator would be likely biased downwards.

We measure the aggregate banking sector's effective leverage $d\Theta_t$, and the two yield curves $\mathbb{E}_t[dr^K_{t+h}]$ and $\mathbb{E}_t[dr^B_{t+h}]$ empirically, and estimate the three key parameters, $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}$ and γ using the generalized method of moments. We briefly overview the variables below and provide details of the estimation in Appendix D.2.

Leverage: We obtain the market value of equity and the liquid asset positions of banks from the linked CRSP - Call Report data. We aggregate bank-holding companies' market value and their net supply of liquid assets (liquid liabilities minus liquid assets holdings). The effective leverage of the banking sector is calculated as

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

Real liquid rates: We take the nominal yield curves based on Treasury bonds from the U.S. Treasury and adjust them with inflation expectations over different horizons from the Cleveland Fed to construct the yield curve for real liquid rates.

[8] For example, a capital quality shock, as in Merton (1973), can be a source of exogenous variation in the value of capital and thus net worth.

[9] In the words of the standard demand-supply estimation, we allow for both shifts in the demand and supply curve to drive changes in prices (returns). But we assume that changes in the supply curve are all parallel shifts due to changes in net worth, which we observe in the data. As a result, banks' leverage represents an invariant part of the supply curve that we can identify through changes in prices.

Returns on capital: We use a corporate bond yield curve as a proxy for expected returns on capital over different horizons. The corporate bond yield curve data are derived from high-quality market corporate bonds (grade A and above), also provided by the U.S. Treasury. We adjust them proportionally so that the long-term (20+) yield corresponds to Moody's BAA bond yields, which is close to the rate on prime bank loans. We convert nominal yields into real yields using the same inflation expectations data.

Table 2: Estimated parameters of the financial constraint

	All data	Excluding recessions
size of cross-price elasticities, $\bar{\Theta}_{r^K}$	24.15^{***}	25.57^{***}
	(5.80)	(5.06)
size of own-price elasticities, $\bar{\Theta}_{r^B}$	26.58^{***}	18.33^{***}
	(6.41)	(3.78)
the forward-looking component, γ	0.957^{***}	0.970^{***}
	(0.01)	(0.00)
Observations	243	212

Note: We use monthly data from 2001 January to 2020 April. Optimal weighting matrix and standard errors use heteroskedastic and autocorrelation consistent (HAC) estimators. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

The first column of Table 2 presents our baseline estimates of $\bar{\Theta}_{r^K}$, $\bar{\Theta}_{r^B}$, and γ. Estimates of $\bar{\Theta}_{r^K}$ and $\bar{\Theta}_{r^B}$ are around 25. This means banks increase their effective leverage by 25 percentage points when the quarterly spread between the two returns in the following quarter increases by one percentage point for one quarter. Banks' effective leverage responds to future changes in returns with a discount rate γ around 0.96, which implies a "half-life" of four years: the response to a one-quarter spread increase four years ahead is half as strong as the response to the same change in the spread in the following quarter. To the extent that changes in returns are persistent, banks choose their effective leverage in response to the discounted sum of all future changes in spreads. To alleviate concerns that our results are driven mostly by large movements of effective leverage and spreads in times of financial distress, we estimate them on a restricted sample that excludes all NBER recession months. The estimates, shown in column 2, remain largely unchanged.

Within the class of models nested in our framework, our estimation does not impose any additional restriction on the form of financial constraints. Our semi-structural estimation is flexible in the sense that it is free from the specific micro-foundation and functional forms imposed by fully specified models of financial frictions and therefore avoids the difficulty of measuring deep parameters such as asset diversion rate or monitoring cost. This general formulation allows us to compare the canonical models systematically: Our estimation suggests an important role for a forward-looking component: γ being close to one, which suggests models of the Gertler-Kiyotaki-Karadi type capture an empirically important feature of the financial constraints. Yet, as we show below, standard models that feature a forward-looking constraint impose a restriction on the other parameters, $\bar{\Theta}_{r^K}$ and $\bar{\Theta}_{r^B}$, which are not necessarily the same as those implied by the data. In Section 6, we show differences in these elasticities have significant effects on aggregate responses.

Net worth and steady-state returns: Finally, to complete the calculation of the elasticities of liquidity supply, we supplement our estimates with standard calibration of the net worth process and steady-state returns. We set $f = 0.05$, which is roughly the dividend rate of banks used in the range of values in the literature (Gertler and Karadi (2011), Gertler and Kiyotaki (2010), Lee et al. (2020)).[10] To calibrate the steady-state returns of our model, we set r^B equal to 1.0% (annually), consistent with the average real yield on Treasury debt with maturity around 5-10 years in our sample. The value for r^K is 3.3% per year, corresponding to the average real yield on BAA corporate bonds. The average effective leverage $\bar{\Theta}$ in our sample is 4. Given f and the steady-state real returns and leverage, the parameter m is determined by the steady-state bank net worth (net worth is equal to 13% of annual GDP).

Implied elasticities: Figure 2 shows the implied semi-elasticities, $\frac{\partial \mathcal{D}_t/\partial r_s^K}{\mathcal{D}_t}$ and $\frac{\partial \mathcal{D}_t/\partial r_s^B}{\mathcal{D}_t}$, calculated using the formula in Lemma 2 together with the estimated parameters for $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}, \gamma$ and the calibrated net worth process. Consider an increase in returns r_s^K, r_s^B in period s. The cross-price elasticities show that liquidity supply increases in all periods t prior to s, peaking one period before s. The increases reflect the forward-looking part of the financial constraint. After the peak, liquidity supply drops sharply. Yet, it remains elevated because of the propagation through the net

[10] In Appendix D.4, we show that our conclusion about the relative effectiveness of policies is insensitive to different values of f.

worth process. The own-price elasticities exhibit a similar pattern with the opposite sign. The size of the initial cross-price semi-elasticity $\frac{\partial \mathcal{D}_0/\partial r_1^K}{\mathcal{D}_0}$ is 8.05. This means that a one-percentage unexpected increase in the return on capital one period ahead leads to an increase in liquidity supply by 8.05%. The corresponding number for the own-price semi-elasticity is -8.86.

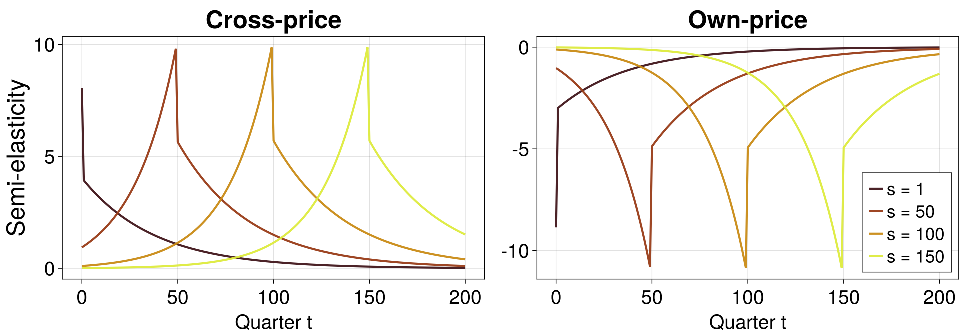

Figure 2: Semi-elasticities of liquidity supply. Each line corresponds to a different period s and shows semi-elasticity of liquidity supply in quarter t with respect to r_s^K and r_s^B.

Model Comparison: Alternative Specifications of Liquidity Supply

We compare our estimates of liquidity supply to three common specifications in the literature, and we contrast their quantitative implications in Section 6.

First, we consider the financial sector in Gertler and Karadi (2011) and Gertler and Kiyotaki (2010). The Gertler-Karadi-Kiyotaki model imposes a tight link between steady-state bank balance sheets and the key elasticities of the financial sector. We calculate the implied elasticities using the banking sector's effective leverage.[11] The three parameters that govern the liquidity supply elasticities are given by the following:

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

The values of these parameters implied by the long-run averages of bank balance sheets and steady-state returns are 11.90, 11.97, and 0.987, respectively. The semi-elasticities are only a half of those in our baseline: $\frac{\partial \mathcal{D}_0/\partial r_1^K}{\mathcal{D}_0}$ is 3.96, $\frac{\partial \mathcal{D}_0/\partial r_1^B}{\mathcal{D}_0}$ is -3.99.

[11] The implied elasticities depend on the discount factor used by banks. The formula below uses $\{r_{t+1}^K\}$ as discount rates. We provide the formula for other alternatives in Appendix C.1.

Second, we consider a case in which the private liquidity supply is perfectly inelastic: $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B} \to \infty$, and $\bar{\Theta}_{r^K}/\bar{\Theta}_{r^B} \to 1$. In this case, the model converges to an economy where the capital and liquid asset markets are linked by a financial sector that responds perfectly elastically to changes in capital returns and liquid rates. Capital returns and liquid rates feature a constant spread. This feature is an important assumption in Auclert et al. (2023). Our result shows that their assumption is equivalent to modeling a financial sector with perfectly elastic supply.

Finally, we consider the case in which $\mathbf{D}_{r^K}$, $\mathbf{D}_{r^B}$, and $\mathbf{D}_y$ are all identically zero. In this case, both liquidity supply and the net worth of banks are constant. This specification is a modified version of Kaplan et al. (2018). The level of bank liquidity supply reflects its empirical counterpart, but the elasticities are kept zero, as in most two-asset HANK models. In Appendix E.2, we provide a detailed discussion of the relationship between our framework, Kaplan et al. (2018), and Auclert et al. (2023).

Calibration for the rest of the model

Preferences: We assume there are two types of households, indexed by s. Their population shares are μ_s. They have a period utility function of the following form:

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

We set the intertemporal elasticity of substitution, $1/\sigma_s$, to $1/2$ for $s = 1$ and to 2 for $s = 2$.[12] The Frisch labor supply elasticity, φ, is set to 1. Parameter ς is set so that steady-state average hours worked equal one-third. Finally, the share of agents with high intertemporal elasticity of substitution is set to 20%.

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: Households cannot have a negative asset position, $\underline{a} = \underline{b} = 0$. Adjustment of

[12] This is consistent with Aguiar et al. (2020).

illiquid assets holdings incurs a real cost similar to Auclert et al. (2021):

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

We set χ_0 to 0.1, and χ_2 to 2. We calibrate discount rates β_s of both types and χ_1 to match three targets: the steady-state ratios of liquid and illiquid assets to GDP, and the share of hand-to-mouth households. In the calibrated model, liquid and illiquid assets to annual GDP are 0.56 and 3.57 respectively. 30.2% of the households are hand-to-mouth, with 16.7% being poor hand-to-mouth (without any liquid assets) in the steady-state.

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 $\bar{\iota}_1, \bar{\iota}_2$ are set to ensure that the steady-state investment to capital ratio equals δ, and the price of capital is 1. We set $\kappa_I = 0.5$, which implies the elasticity of investment to capital price is 2. The elasticity of substitution between varieties of labor, ε_W, is set to 6. The degree of nominal wage rigidities, κ_W, is set to 200, so the slope of the wage Phillips curve is 0.04.

Goverment: The income tax function is given by $\mathcal{T}(y_{i,t}) = y_{i,t} - (1-\tau) y_{i,t}^{1-\lambda}$. We set net tax revenue, T, to 15% of steady-state output. We set liquid assets provided by the government to 15.6% of the annual output. We assume that the government does not hold any illiquid assets in the steady state. Government purchases are determined residually from the budget constraint and amount to 14.8% of GDP. We set λ, the tax system's progressivity parameter, to 0.18.

6 Model and Policy Comparison

Existing models for policy analysis feature implicit assumptions about the liquidity supply elasticities, and these assumptions lead to quantitatively significant differences in aggregate responses. We consider two policy alternatives: asset purchases versus tax cuts. These two are central to the "Wall Street v.s. Main Street" debate. In this debate, the argument for asset purchases usually builds on their effectiveness in stimulating aggregate output. We study aggregate responses to these policies and the relative effectiveness of the two in stimulating aggregate output. We compare policy responses across models that implicitly assume different liquidity supply elasticities and contrast these assumptions with our estimates from Section 5.

6.1 Policy Alternatives: Asset Purchases and Tax Cuts

We consider two alternative policies financed through government debt issuance. The paths of government debt, liquid rates, and government purchases under both policies are identical:

$$db_t^G = \rho \, db_{t-1}^G + \eta^t, \quad dr_t^B = 0, \quad dg_t = 0.$$

We assume the monetary policy sets the real liquid rate constant at its steady-state level (therefore, the nominal rate adjusts one-to-one with expected inflation) and that government purchases are kept constant over time. As we show in Corollary 1, these assumptions are without loss of generality for comparing the relative effectiveness of policies, as long as the two policies feature the same paths for $\{db_t^G, dr_t^B, dg_t\}$. With the same issuance of government debt, we consider two policy alternatives: one directs resources to purchase illiquid assets while the other uses the proceeds for a tax cut.

Asset purchases: We consider a transitory government asset purchase program in which the government's illiquid asset holdings are equal to the injection of liquid government debt: $da_t^G = db_t^G$. This is associated with net asset purchases $d\Delta_t$ and net taxes given by

$$d\Delta_t = da_t^G - (1 + r^A) da_{t-1}^G, \quad dT_t = (r^B - r^A) db_{t-1}^G.$$

Tax cuts: Alternatively, we consider the government keeping its illiquid asset holdings at the steady-state level, $da_t^G = 0$. Instead of asset purchases, the government pays out the proceeds from debt issuance as tax cuts:

$$d\tilde{T}_t = (r^B - r^A) db_{t-1}^G - d\Delta_t.$$

We assume the following parameters for the path of government debt: $\eta = 0.5$ and $\rho = 0.95$. For the government asset purchases program, the government increases its holdings of illiquid assets for four quarters and then starts selling them back to households. For the tax cuts, transfers are received by households mostly within four quarters, and then the government increases taxes to retire government debt. The resulting paths of asset purchases and net tax revenue under the two policies are compared in Figure 3.

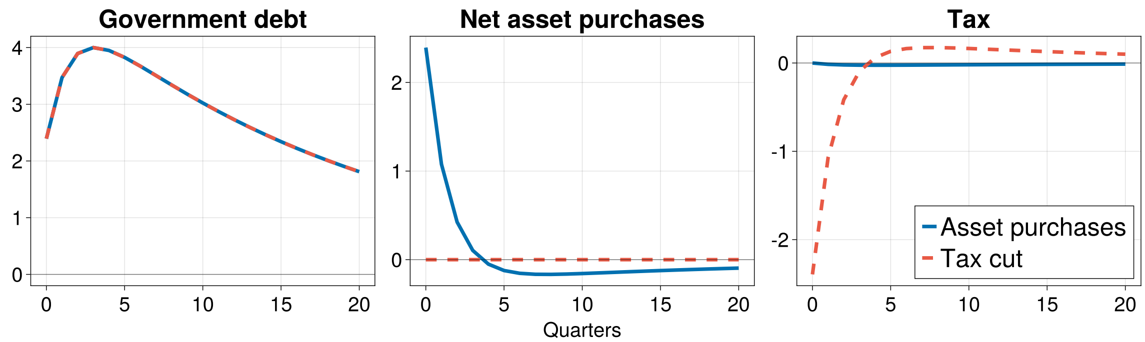

Figure 3: Government debt, net asset purchases, and taxes; x-axis: quarters, y-axis: % of steady-state GDP.

We study the effects of each policy separately under different model specifications featuring different liquidity supply elasticities, including perfectly inelastic and elastic liquidity supply, as well as that implied by the calibration of a Gertler-Karadi-Kiyotaki model and our empirical estimates.

6.2 Targeting the Financial Market: Asset Purchases

Figure 4 shows that output, consumption, investment, and capital price respond positively to the government asset purchase program. The red line represents the response when elasticities of liquidity supply are given by our empirical estimates, $\bar{\Theta}_{r^K} = 24.2$. Yellow shades from light to dark represent models with increasing values for $\bar{\Theta}_{r^K}$ from the Gertler-Karadi-Kiyotaki specification ($\bar{\Theta}_{r^K} = 11.9$) to our empirical estimates.[13] The blue line indicates responses with perfectly inelastic liquidity supply, and the black line indicates responses with perfectly elastic liquidity supply, $\bar{\Theta}_{r^K} \to \infty$. When the financial sector's liquidity supply has low elasticities with respect to $d\boldsymbol{r}^K$, aggregate responses of output, consumption, investment, and asset prices are amplified. Moreover, differences in the output response are mostly driven by differences in investment. Increases in investment are due to firms' responses to capital

[13]The value of γ, the forward-looking component of the financial constraint, is kept at the level corresponding to the empirical estimate of 0.957.

price increases, associated with lower expected returns on capital $d\boldsymbol{r}^K$, consistent with the asset market channel described in Section 4.2.

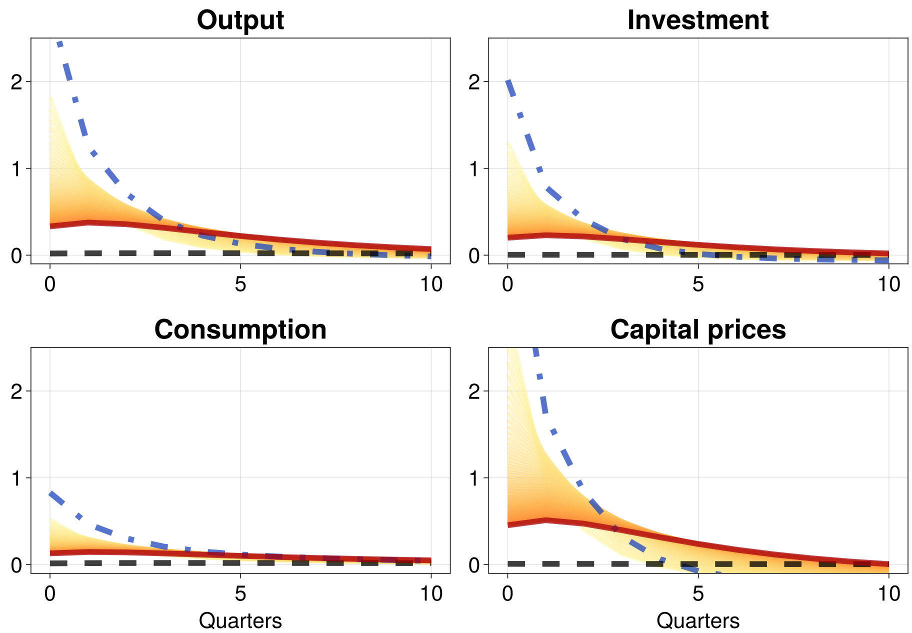

Figure 4: Impulse response functions to government illiquid asset purchases. The y-axis: % of steady-state GDP. Red: empirical elasticities. Light to dark yellow: low to high elasticities starting from GKK. Blue: perfectly inelastic. Black: perfectly elastic.

To understand how the financial sector affects aggregate responses, we decompose the aggregate output response into the three channels in Theorem 1:

$$d\boldsymbol{y} = \underbrace{\left(\mathbf{I} - \boldsymbol{\Psi}_y - \boldsymbol{\Omega}\,\boldsymbol{\epsilon}_y\right)^{-1}}_{(3)} \times \Big(\underbrace{\boldsymbol{\Psi}_T d\boldsymbol{T}}_{(1)} + \underbrace{\boldsymbol{\Omega}\left(d\boldsymbol{b}^G + \boldsymbol{\epsilon}_T d\boldsymbol{T}\right)}_{(2)} \Big),$$

The three panels in Figure 5 show the decomposition of total aggregate output responses into (1) the goods market channel, (2) the asset market channel, and (3) the general equilibrium effect resulting from the modified Keynesian cross, which we plot as the difference between $d\boldsymbol{y}$ and the sum of the first two effects.

The decomposition in Figure 5 shows how each channel is affected by liquidity supply elasticities. First, the goods market channel depends only on the household sector, and is not affected by the specification of the financial sector. This role of this channel is negligible because the policy does not generate large movements in $d\boldsymbol{T}$. On the other hand, the issuance of government debt and households' saving response shift excess liquidity in the economy. The asset market channel depends crucially on the features of the financial sector and drives the differences in output responses in Figure 4. Asset purchases initially lead to an increase in excess liquidity because there is a significant increase in government debt. In response, the rate of return on capital r_{t+1}^K goes down, and the capital price q_t jumps up. It induces banks to reduce liquidity transformation and supply less liquid assets. A substitution effect due to changes in returns shifts household asset holdings from illiquid assets to liquid assets. Yet, an increase in capital price increases consumption and investment hence increasing aggregate demand, and therefore an income effect increases the households' holdings of both assets. When liquidity supply is inelastic, the changes in returns on capital are large, and output responses are strong.

Finally, the general equilibrium effect through the Keynesian cross is generally small. The standard Keynesian cross logic that aggregate income leads to more consumption and investment needs to be modified due to responses through the financial sector. When there is an increase in output, households demand more liquid assets, which leads to a fall in excess liquidity, counteracting the first two channels. The dampening of the Keynesian cross logic is stronger when the financial sector responds inelastically because the capital price and returns need to respond strongly to balance the liquid asset market. This explains why the general equilibrium effect is larger when liquidity supply elasticities are high, despite the asset market channel being weaker.[14]

[14]The ranking of lines in the right panel of Figure 5 reflects both the partial equilibrium response (goods market and asset market channels) and the Keynesian multiplier. For example, the black line shows a smaller GE response than the red lines due to the absence of the asset market channel.

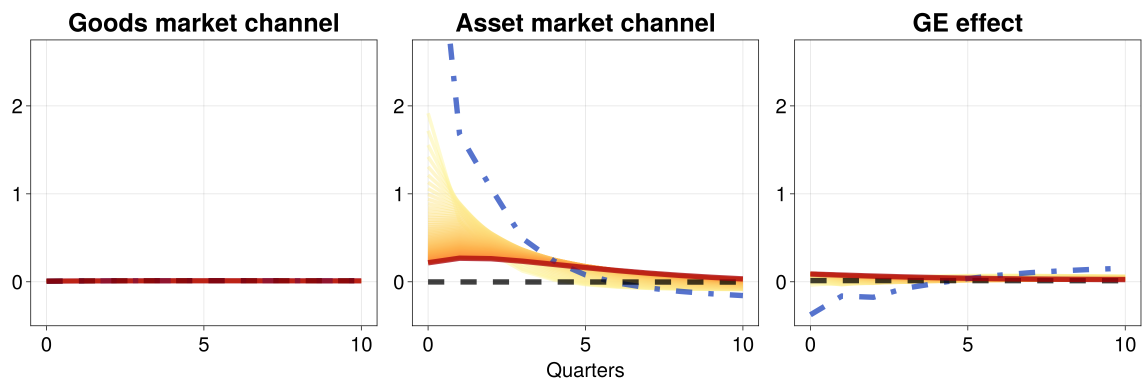

Figure 5: Decomposition of output response to a government illiquid asset purchases, using the formula from Theorem 1. The y-axis: % of steady-state GDP. Red: empirical elasticities. Light to dark yellow: low to high elasticities starting from GKK. Blue: perfectly inelastic. Black: perfectly elastic.

6.3 Targeting Households: Tax Cut

Figure 6 shows aggregate responses of output, consumption, investment, and capital price to the tax cuts, where each line represents the same model specifications as in Section 6.2. Similar to responses to the government asset purchases program, aggregate responses of output, consumption, investment, and asset prices are amplified when the financial sector's liquidity supply features low elasticities in response to $d\boldsymbol{r}^K$. However, the differences in responses are smaller in comparison to responses to the asset purchase program. To understand why the responses to tax cuts are less sensitive to different model specifications, we show the same decomposition of aggregate output response in Theorem 1 in Figure 7.

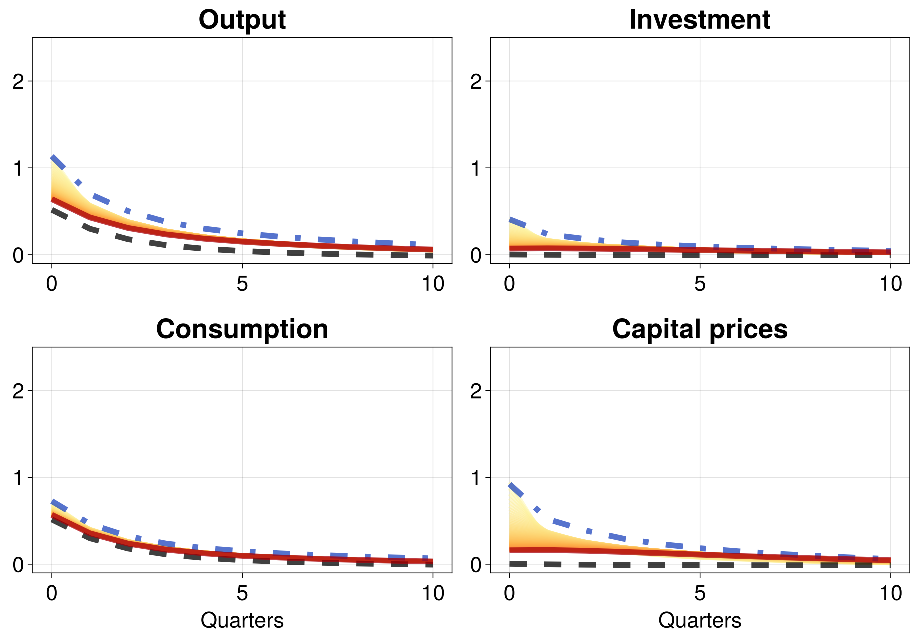

Figure 6: Impulse response functions to a tax cut. The y-axis: % of steady-state GDP. Red: empirical elasticities. Light to dark yellow: low to high elasticities starting from GKK. Blue: perfectly inelastic. Black: perfectly elastic.

The decomposition shows that the asset market channel once again explains most of the differences across model specifications, but the differences are dampened. This is because the tax cut induces households to save in liquid assets and absorbs the excess liquidity created by the corresponding issuance of government debt. With less excess liquidity created by the tax cuts, the asset market channel is not as strong in comparison to the asset purchases program. As a result, the aggregate output responses are less sensitive to the specification of the financial sector. In contrast to responses to the asset purchase program where the goods market channel is negligible, aggregate output response to the tax cut has a noticeable contribution from the goods market channel. When the financial sector features a relatively elastic liquidity supply, the goods market channel becomes the dominant channel that accounts for most of the size of aggregate output responses. The goods market channel represents the direct response of households' consumption to the tax cuts, and its strength is determined by their aggregate intertemporal marginal propensity to consume. Allowing for rich household heterogeneity in our framework gives us the ability to calibrate households' consumption responses to match evidence from the microdata, and thereby pin down the strength of the goods market channel.

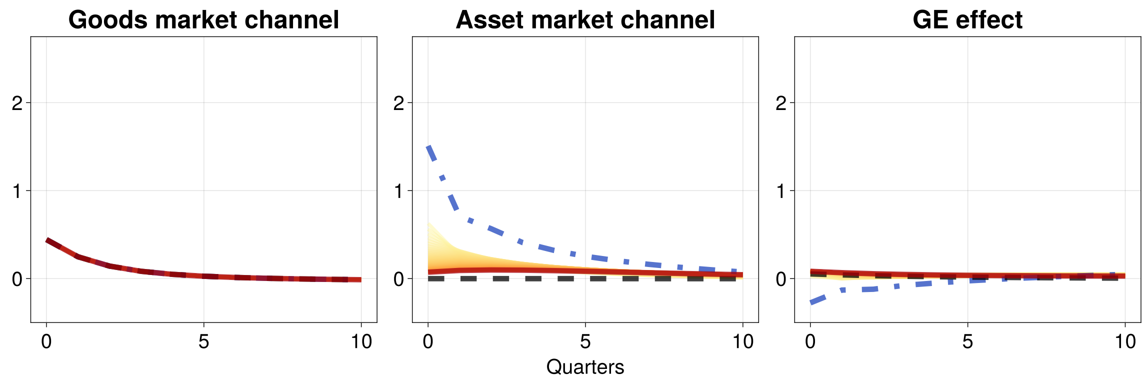

Figure 7: Decomposition of output response to tax cuts, using the formula from Theorem 1. The y-axis: % of steady-state GDP. Red: empirical elasticities. Light to dark yellow: low to high elasticities starting from GKK. Blue: perfectly inelastic. Black: perfectly elastic.

6.4 Relative Effectiveness of Policies

Figure 8 compares the relative effectiveness between asset purchases and tax cuts in stimulating aggregate output across models with different liquidity supply elasticities. For each model, we plot the output differences between the two policies, $\widehat{d\boldsymbol{y}} := d\boldsymbol{y}^{\text{asset}} - d\boldsymbol{y}^{\text{tax cut}}$, as in Corollary 1. The prediction varies widely among the models: At one extreme, models with perfectly inelastic liquidity supply (blue line) predict the asset purchase program has a much stronger effect on aggregate output than tax cuts: aggregate output response on impact is more than twice as large and the difference amount to 1.8% of steady-state output. A model with liquidity supply implied by financial intermediation of the Gertler-Karadi-Kiyotaki type (light yellow) gives a qualitatively similar prediction: asset purchases have a stronger effect on output than tax cuts. At the other extreme, models with perfectly elastic liquidity supply (black line), by assumption, rule out the possibility that asset purchases may have an effect on aggregate output per se. As a result, a comparison between the two policies features nearly no response from asset purchases and predicts a much stronger effect in response to tax cuts. The model featuring liquidity supply elasticities from our empirical estimates generates a non-negligible response to asset purchases but predicts that tax cuts targeting the household sector have a relatively stronger effect on output.

The result is driven by two forces. On the one hand, the high elasticities from our estimation imply modest aggregate output responses through the asset market channel in comparison to a model with perfectly inelastic liquidity supply or financial intermediation of the Gertler-Karadi-Kiyotaki type. Yet, another equally important force driving the result is that the household sector features a relatively strong consumption response through the goods market channel in response to tax cuts. This highlights the importance of accounting for key features of both the financial sector (liquidity supply elasticities) and households (marginal propensity to consume and save) in order to understand the aggregate responses to macroeconomic policies.

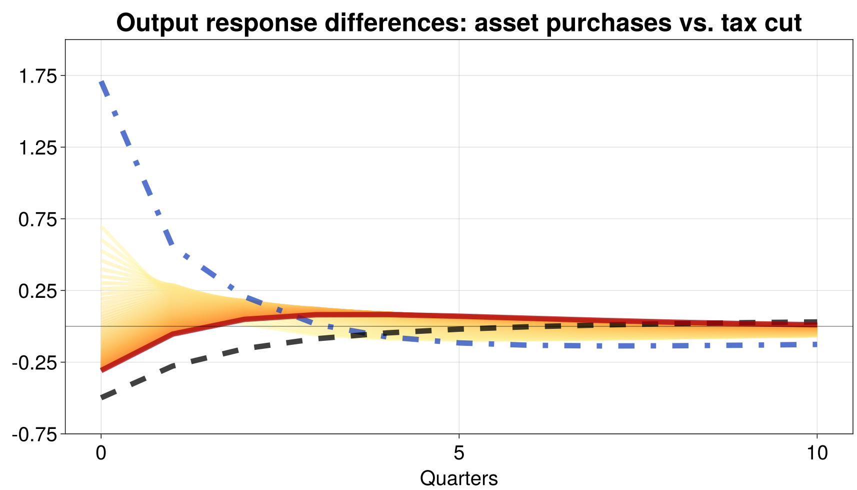

Figure 8: Difference between output response to asset purchases and tax cuts. The y-axis: % of steady-state GDP. Red: empirical elasticities. Light to dark yellow: low to high elasticities starting from GKK. Blue: perfectly inelastic. Black: perfectly elastic.

7 Conclusion

We study how the financial sector affects the effectiveness of macroeconomic policies in a framework that nests models of financial intermediation with various microfoundations and allows for rich household heterogeneity. We characterize aggregate responses with an intertemporal demand-and-supply system, which allows us to isolate how the financial sector affects aggregate responses to policies through different channels in the goods and asset markets. We show that the financial sector's liquidity supply elasticities with respect to expected returns are sufficient statistics that summarize its role in shaping aggregate responses. These elasticities determine the relative effectiveness of policies in stimulating aggregate output, and they are central to the "Wall Street vs. Main Street" debate: whether transferring resources to the financial sector is more effective in stimulating the economy than transferring resources to households. In commonly used setups, aggregate output responses differ by order of magnitudes due to implicit assumptions about these elasticities. Our estimates of elasticities for the U.S. economy imply a modest effect through the asset markets and a relatively strong effect of targeting households.

The importance of these elasticities implores comprehensive empirical measurement beyond the scope of this paper, including the measurement of these elasticities at the micro level as well as the aggregation from micro to macro elasticities. Detailed microfoundations of the financial friction that generate elasticities consistent with empirical measures will be useful to understand how these elasticities are affected by various macroprudential regulations imposed on the financial sector. 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.

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.

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist, "The financial accelerator in a quantitative business cycle framework," in J. B. Taylor and M. Woodford, eds., *Handbook of Macroeconomics*, Vol. 1 of *Handbook of Macroeconomics*, Elsevier, December 1999, chapter 21, pp. 1341–1393.

Bianchi, Javier and Enrique G Mendoza, "Optimal time-consistent macroprudential policy," *Journal of Political Economy*, 2018, *126* (2), 588–634.

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*, 30–65.

Cúrdia, Vasco and Michael Woodford, "The central-bank balance sheet as an instrument of monetarypolicy," *Journal of Monetary Economics*, 2011, *58* (1), 54–79. Carnegie-Rochester Conference Series on Public Policy: The Future of Central Banking April 16-17, 2010.

den Heuvel, Skander J. Van, "The welfare cost of bank capital requirements," *Journal of Monetary Economics*, March 2008, *55* (2), 298–320.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl, "The deposits channel of monetary policy," *The Quarterly Journal of Economics*, 2017, *132* (4), 1819–1876.

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.

Eggertsson, Gauti B., Ragnar E. Juelsrud, Lawrence H. Summers, and Ella Getz Wold, "Negative Nominal Interest Rates and the Bank Lending Channel," NBER Working Papers 25416, National Bureau of Economic Research, Inc January 2019.

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

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

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

_ **and Peter Karadi**, "A model of unconventional monetary policy," *Journal of monetary Economics*, 2011, *58* (1), 17–34.

Gornemann, Nils, Keith Kuester, and Makoto Nakajima, "Monetary policy with heterogeneous agents," Working Papers 12-21, Federal Reserve Bank of Philadelphia 2012.

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, Benjamin Moll, and Giovanni L Violante, "Monetary policy according to HANK," *American Economic Review*, 2018, *108* (3), 697–743.

Karadi, Peter and Anton Nakov, "Effectiveness and addictiveness of quantitative easing," *Journal of Monetary Economics*, 2021, *117* (C), 1096–1117.

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

_ **and** _ , "Liquidity, business cycles, and monetary policy," *Journal of Political Economy*, 2019, *127* (6), 2926–2966.

Koby, Yann and Christian K. Wolf, "Aggregation in heterogeneous-firm models: Theory and measurement," Technical Report 2020.

Krishnamurthy, Arvind and Annette Vissing-Jorgensen, "The aggregate demand for treasury debt," *Journal of Political Economy*, 2012, *120* (2), 233–267.

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

_ , Emi Nakamura, and Jón Steinsson, "The Power of Forward Guidance Revisited," *American Economic Review*, October 2016, *106* (10), 3133–3158.

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

Merton, Robert C., "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 1973, *41* (5), 867–887.

Ottonello, Pablo, Diego J Perez, and Paolo Varraso, "Are collateral-constraint models ready for macroprudential policy design?," *Journal of International Economics*, 2022, *139*, 103650.

Schroth, Josef, "On the Distributional Effects of Bank Bailouts," *Review of Economic Dynamics*, April 2021, *40*, 252–277.

Townsend, Robert M., "Optimal contracts and competitive markets with costly state verification," *Journal of Economic Theory*, October 1979, *21* (2), 265–293.

Uribe, Martin and Vivian Z. Yue, "Country spreads and emerging countries: Who drives whom?," *Journal of International Economics*, June 2006, *69* (1), 6–36.

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.

Woodford, Michael, "Simple Analytics of the Government Expenditure Multiplier," *American Economic Journal: Macroeconomics*, January 2011, *3* (1), 1–35.

A Proofs and Derivations

A.1 Proof of Lemma 2

Proof. To save on notation define $\Theta_t := \Theta\big(\{r^B_{s+1}, r^K_{s+1}\}_{s\geq t}\big)$. To get the response of liquidity supply, recall that $d_t = (\Theta_t - 1)\, n_t$, so

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

Totally differentiating 2 and evaluating at the steady state results in

$$dn_t = (1-f)\left(\left[\left(\bar{r}^K - \bar{r}^B\right) d\Theta_{t-1} + \left(dr^K_t - dr^B_t\right)\bar{\Theta} + dr^B_t\right]\bar{n} + \left[\left(\bar{r}^K - \bar{r}^B\right)\bar{\Theta} + \left(1+\bar{r}^B\right)\right] dn_{t-1}\right),$$

where

$$d\Theta_t = \sum_{u=1}^{\infty}\left(\frac{\partial \Theta_t}{\partial r^K_{t+u}} dr^K_{t+u} + \frac{\partial \Theta_t}{\partial r^B_{t+u}} dr^B_{t+u}\right),$$

because

$$\frac{\partial \Theta_t}{\partial r^K_{s+1}} = \frac{\partial \Theta_t}{\partial r^B_{s+1}} = 0, \quad \forall s \leq t.$$

Define $G := (1-f)\left[\left(\bar{r}^K - \bar{r}^B\right)\bar{\Theta} + \left(1+\bar{r}^B\right)\right] \geq 0$ to write

$$dn_t = (1-f)\sum_{u=0}^{t} G^u \left[\left(\bar{r}^K - \bar{r}^B\right) d\Theta_{t-1-u}\bar{n} + \left(dr^K_{t-u} - dr^B_{t-u}\right)\bar{\Theta}\bar{n} + dr^B_{t-u}\bar{n}\right].$$

Now, consider a particular variation such that $dr^K_s = 1$ and $dr^K_u = 0$ for all $u \neq s$, and $dr^B_u = 0$ for all u. We have

$$dn_t = \begin{cases} \bar{n}\,(1-f)\left(\bar{r}^K - \bar{r}^B\right)\sum_{u=0}^{t-1} G^u \frac{\partial \Theta_{t-1-u}}{\partial r^K_s}, & s > t, \\ \bar{n}\,(1-f)\left(r^K - r^B\right)\sum_{u=t-s}^{t-1} G^u \frac{\partial \Theta_{t-1-u}}{\partial r^K_s} + \bar{n}\,(1-f)\, G^{t-s}\bar{\Theta}, & s \leq t. \end{cases}$$

The assumption about the structure of Θ_t implies

$$\frac{\partial \Theta_{t-1-u}}{\partial r^K_s} = \begin{cases} \gamma^{s-t+u}\bar{\Theta}_{r^K}, & s > t-1-u, \\ 0, & s \leq t-1-u. \end{cases}$$

which allows us to write

$$dn_t = \begin{cases} \bar{\Theta}_{r^K} \bar{n} \left(1 - f\right) \left(\bar{r}^K - \bar{r}^B\right) \gamma^{s-t} \sum_{u=0}^{t-1} \left(\gamma G\right)^u dr_s^K, & s > t, \\ \bar{\Theta}_{r^K} \bar{n} \left(1 - f\right) \left(r^K - r^B\right) G^{t-s} \sum_{l=0}^{s-1} \left(\gamma G\right)^l dr_s^K + \bar{n} \left(1 - f\right) G^{t-s} \bar{\Theta} dr_s^K, & s \leq t. \end{cases}$$

Finally, define $\Sigma(s) \coloneqq (1-f)(\bar{r}^K - \bar{r}^B)\frac{1-(\gamma G)^s}{1-\gamma G}$ and divide by $\left(\bar{\Theta} - 1\right)\bar{n}$ to get

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

Derivation of $\frac{\partial \mathcal{D}_t / \partial r_s^K}{\mathcal{D}_t}$ follows the same steps. □

A.2 Proof of Lemma 3

Proof. We first show how we obtain the aggregate demand and supply functions and then demonstrate that if the goods market and the liquid asset market clear, then by Walras' law the illiquid asset market clears as well. We begin by showing that

$$(1 - \tau_t) \left(\frac{W_t}{P_t} z_{i,t} h_{i,t}\right)^{1-\lambda} = \frac{z_{i,t}^{1-\lambda}}{\int_0^1 z_{i,t}^{1-\lambda} di} \left[(1 - \alpha) y_t - T_t\right]$$

Recall that we have $\frac{W_t}{P_t} h_t = (1 - \alpha) y_t$ and $h_{i,t} = h_t$ so

$$(1 - \tau_t) \left(\frac{W_t}{P_t} z_{i,t} h_{i,t}\right)^{1-\lambda} = (1 - \tau_t) \left[(1 - \alpha) y_t z_{i,t}\right]^{1-\lambda}.$$

Now, since

$$T_t = \frac{W_t}{P_t} h_t - (1 - \tau_t) \int \left(\frac{W_t}{P_t} z_{i,t} h_{i,t}\right)^{1-\lambda} di,$$

we have

$$(1 - \tau_t) \left[(1 - \alpha) y_t z_{i,t}\right]^{1-\lambda} = \frac{z_{i,t}^{1-\lambda}}{\int_0^1 z_{i,t}^{1-\lambda} di} \left[(1 - \alpha) y_t - T_t\right].$$

Using this in the household budget constraint, we see that adjustment costs and optimal policy rules for consumption and savings in each type of asset depend on the aggregates only through the path of output $\{y_t\}_{s=0}^{\infty}$, taxes $\{T_t\}_{s=0}^{\infty}$ and returns on both types of assets $\left\{r_t^A, r_t^B\right\}_{s=0}^{\infty}$. Therefore given the initial distribution of assets and productivity, we obtain $\mathcal{A}_t\left(\left\{y_s, r_s^A; r_s^B, T_s\right\}_{s=0}^{\infty}\right)$, $\mathcal{B}_t\left(\left\{y_s, r_s^A; r_s^B, T_s\right\}_{s=0}^{\infty}\right)$ and $\mathcal{C}_t\left(\left\{y_s, r_s^A; r_s^B, T_s\right\}_{s=0}^{\infty}\right)$.

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, after rearranging, can be solved to obtain capital in each period as a function of the path of output, r^K and 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 $q_t := \mathcal{Q}_t\left(\left\{y_s, r_s^K\right\}_{s=0}^{\infty}\right)$. Similarly, given $\left\{r_t^K\right\}_{t\geq 0}$ and $\left\{r_t^B\right\}_{t\geq 0}$ we obtain the liquidity supply function $\mathcal{D}_t\left(\left\{r_s^K, r_s^B\right\}_{s=0}^{\infty}\right)$.

We now derive the function $\mathcal{R}_t^A\left(\cdot\right)$ using Equation 3 as follows:

$$\begin{aligned} 1 + r_t^A =& 1 + \frac{1}{a_{t-1}}\left(r_t^K q_{t-1} k_{t-1}^F + r_t^N n_{t-1}\right) \\ =& \frac{1}{q_{t-1}k_{t-1} - d_{t-1}}\left(\left(1 + r_t^K\right) q_{t-1} k_{t-1} - \left(1 + r_t^B\right) d_{t-1}\right). \end{aligned}$$

Define $L_t := d_t / \left(q_t k_t\right)$. This variable can be interpreted as a *liquidity transformation ratio*. As explained before, we have $d_t = \mathcal{D}_t\left(\left\{r_s^K, r_s^B\right\}_{s=0}^{\infty}\right)$, $q_t = \mathcal{Q}_t\left(\left\{y_s, r_s^K\right\}_{s=0}^{\infty}\right)$, and $k_t = \mathcal{K}_t\left(\left\{y_s, r_s^K\right\}_{s=0}^{\infty}\right)$ so we can write

$$L_t = \mathcal{L}_t\left(\left\{y_s, r_s^K, r_s^B\right\}_{s=0}^{\infty}\right), \quad 1 + r_t^A = \frac{1}{1 - \mathcal{L}_{t-1}\left(\cdot\right)}\left(1 + r_t^K\right) - \frac{\mathcal{L}_{t-1}\left(\cdot\right)}{1 - \mathcal{L}_{t-1}\left(\cdot\right)}\left(1 + r_t^B\right).$$

r_t^A depends on $\left\{r_s^K, r_s^B\right\}_{s=0}^{\infty}$. We can write it in a more compact way as $r_t^A :=$

$\mathcal{R}_t^A\left(\left\{r_s^K, r_s^B; \mathcal{D}_\sqcup\right\}_{s=0}^\infty\right)$. Because $\int b_{i,t}di = \mathcal{B}_t\left(\left\{y_s, r_s^A; r_s^B, T_s\right\}_{s=0}^\infty\right)$ and $d_t = \mathcal{D}_t\left(\left\{r_s^K, r_s^B\right\}_{s=0}^\infty\right)$

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

means that the liquid asset market clears. Since government debt satisfies Equation 4, the government budget constraint is satisfied. We can now obtain illiquid asset demand in the same way $\int a_{i,t}di = \mathcal{A}_t\left(\left\{y_s, r_s^A, r_s^B; T_s\right\}_{s=0}^\infty\right)$ for all t. By the Walras law, the illiquid asset market clears $\mathcal{A}_t\left(\left\{y_s, r_s^A, r_s^B; T_s\right\}\right) = q_t k_t - d_t - a_t^G$.

□

A.3 Proof of Proposition 1.

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 B.2. Rearrange and left-multiply by the inverse of $-\boldsymbol{\epsilon}_{r^K}$ to obtain Equation 5.

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

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

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

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

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

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

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

The condition is satisfied for $d\boldsymbol{r}^K = \varsigma d\boldsymbol{r}^B$. □

A.4 Proof of Theorem 1.

Proof. We define aggregate demand 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 B.2.

Let $\boldsymbol{\Omega} := \boldsymbol{\Psi}_{r^K}(-\boldsymbol{\epsilon}_{r^K})^{-1}$, and use Proposition 1 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

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

which is the formula in Theorem 1. □

B Additional Derivations

B.1 Time 0 returns.

We can eliminate r_0^K 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. 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 $\left\{y_s, r_s^K\right\}_{s=0}^{\infty}$.

B.2 Linearized equilibrium conditions

We use the following notation: $d\boldsymbol{r}^B$ represents $\{dr_{s+1}^B\}_{s=0}^{\infty}$. The same convention applies to other rates of return. We use $d\boldsymbol{y}$ to represent $\{dy_s\}_{s=0}^{\infty}$. Our notation is the same for other variables that are not rates of return. These are column vectors. 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 start by obtaining some auxilliary results. Define

$$\mathbf{S}_{+1} := \begin{bmatrix} 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}, \mathbf{S}_{-1} := \begin{bmatrix} 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}.$$

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

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

We can write it as

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

Therefore

$$d\boldsymbol{q} = \frac{\bar{q}'}{k}\left(\mathbf{I}-\mathbf{S}_{-1}\right)d\boldsymbol{k}$$

so

$$d\boldsymbol{q} = \mathbf{Q}_y d\boldsymbol{y} + \mathbf{Q}_{r^K} d\boldsymbol{r}^{\boldsymbol{K}}, \quad d\boldsymbol{x} = \bar{\iota}'(\mathbf{I}-\mathbf{S}_{-1})d\boldsymbol{k} + \bar{\iota} d\boldsymbol{k}$$

which allows us to write

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

We also have

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

Capital price at $t=0$ can change only if the investment rate ι_0 changes. That depends on function $\mathcal{X}_t(\cdot)$. In a matrix form we can write

$$dr_0^K = \frac{\alpha}{\bar{k}}\mathbf{e}_1 d\boldsymbol{y} + (1-\delta)\left(\mathbf{q}_y d\boldsymbol{y} + \mathbf{q}_{r^K} d\boldsymbol{r}^K\right), \tag{7}$$

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

Banks

We now turn to the financial sector of the economy and we characterize derivatives of $\mathcal{D}_t(r_0^K(\boldsymbol{y}, \boldsymbol{r}^K), \boldsymbol{r}^K, \boldsymbol{r}^B)$. We represent them as matrices

$$\begin{aligned}
\mathbf{D}_{r^K} &= \bar{\Theta}_{r^K}\mathbf{N}(\gamma) + \mathbf{N}_0 + \mathbf{n}_0(1-\delta)\mathbf{q}_{r^K}, \\
\mathbf{D}_{r^B} &= -\bar{\Theta}_{r^B}\mathbf{N}(\gamma) - \frac{\bar{\Theta}-1}{\bar{\Theta}}\mathbf{N}_0, \\
\mathbf{D}_y &= \mathbf{n}_0\left[\frac{\alpha}{\bar{k}}\mathbf{e}_1 + (1-\delta)\mathbf{q}_y\right]
\end{aligned}$$

Let $\mathbf{D}_{r^K}$ be a matrix of total derivatives of $\mathcal{D}_t(r_0^K(\boldsymbol{y}, \boldsymbol{r}^K), \boldsymbol{r}^K, \boldsymbol{r}^B)$ with respect to rates of return on capital. Its $(t+1, s)$ entry is a total derivative of liquidity supply at time t with respect to r_s^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 liquidity supply at time $t+1$ with respect to y_{s+1}. To populate these matrices we use formulas from Lemma 2 and the dependence of time-0 return on capital on future returns on capital and output from Appendix B.1.

Recall the definition of G from Appendix A.1. Matrix $\mathbf{N}_0$ consists of terms $G^{t-s}(1-f)\,\bar{n}$, present only for $s \leq t$. It captures the effect of net worth accumulation on liquidity supply, holding the leverage ratio constant. Its $(t+1, s)$-th entry is $G^{t-s}(1-f)\,\bar{n} \geq 0$.

$$\mathbf{N}_0 = (1-f)\left(\bar{\Theta}-1\right)\bar{\Theta}\bar{n}\begin{bmatrix} 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & \cdots \\ G & 1 & 0 & 0 & \cdots \\ G^2 & G & 1 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

Matrix $\mathbf{N}(\gamma)$ consists of all other terms. Its $(t+1, s)$-th entry captures the effect of r_s^K on liquidity supply in period t through changes in the leverage ratio (both in period t and in the past). Define $P := (1-f)\left(\bar{\Theta}-1\right)\left(\bar{r}^K - \bar{r}^B\right) \geq 0$.

$$\mathbf{N}(\gamma) = \bar{n}\begin{bmatrix} 1 & \gamma & \gamma^2 & \cdots \\ P & 1+\gamma P & \gamma + \gamma^2 P & \cdots \\ PG & P + \gamma PG & 1 + \gamma P + \gamma^2 PG & \cdots \\ PG^2 & PG + \gamma PG^2 & P + \gamma PG + \gamma^2 PG^2 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

All entries of this matrix are non-negative. If $\gamma = 0$, then $\mathbf{N}(\gamma)$ is a lower-triangular matrix with ones on the diagonal.

Let us turn to the effect of changes in r_0^K. The sum $\bar{\Theta}_{r^K}\mathbf{N}(\gamma) + \mathbf{N}_0$ allows to capture the effects of changes in return on capital in periods $s = 1, 2, \ldots$, but ignores the effect of r_0^K. Changes in liquidity supply due to dr_0^K can be summarized as

$$\mathbf{n}_0 = (1 - f)\left(\bar{\Theta} - 1\right)\bar{\Theta}\bar{n}\begin{bmatrix} 1 \\ G \\ G^2 \\ \cdots \end{bmatrix},$$

a vector such that its t-th element correponds to the $(t, 1)$-th entry of $\mathbf{N}_0$. The total effect of $d\boldsymbol{r}^K$ on liquidity supply is therefore

$$\mathbf{D}_{r^K} = \bar{\Theta}_{r^K}\mathbf{N}(\gamma) + \left[\mathbf{N}_0 + (1-\delta)\mathbf{n}_0\mathbf{q}_{r^K}\right].$$

where the $(1-\delta)\mathbf{n}_0\mathbf{q}_{r^K}$ term describes how returns on capital in the future move q_0 and therefore r_0^K.

$$\mathbf{D}_y = \mathbf{n}_0\left[\frac{\alpha}{\bar{k}}\mathbf{e}_1 + (1-\delta)\mathbf{q}_y\right]$$

reflects the fact that q_0 (and thus r_0^K) depends also on the path of output. Note that $d\boldsymbol{y}$ matters for liquidity supply only because it affects r_0^K.

Derivation of $\mathbf{D}_{r^B}$ follows the same steps. The main difference is that dr_t^B enters the law of motion for net worth with a coefficient $1 - \bar{\Theta}$ instead of $\bar{\Theta}$.

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. We have $dr_0^A = dr_0^K / (1 - L)$ where L is the steady state ratio d/qk.

We now proceed to eliminate $dr_1^A, dr_2^A, \ldots$ by using the condition that links returns on illiquid assets and on capital, Equation 3. We have

$$dr_t^A = \sum_{s=0}^{\infty} \frac{\partial \mathcal{R}_t^A}{\partial y_s} dy_s + \sum_{s=0}^{\infty} \frac{\partial \mathcal{R}_t^A}{\partial r_{s+1}^K} dr_{s+1}^K + \sum_{s=0}^{\infty} \frac{\partial \mathcal{R}_t^A}{\partial r_{s+1}^B} dr_{s+1}^B$$

They capture the effect of changes in rates of return and changes in output on the liquidity transformation ratio. As shown in Appendix A.2, $\mathcal{R}_t^A$ depends on the liquidity transformation ratio L_t. Since $L_t = \frac{d_t}{q_t k_t}$ we have $dL_t = -\frac{L}{q} dq_t - \frac{L}{k} dk_t + \frac{L}{d} dd_t$. Define

$$\mathbf{L}_{r^K} := \begin{bmatrix} \frac{\partial L_0}{\partial r_1^K} & \frac{\partial L_0}{\partial r_2^K} & \cdots \\ \frac{\partial L_1}{\partial r_1^K} & \frac{\partial L_1}{\partial r_2^K} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}, \mathbf{L}_{r^B} := \begin{bmatrix} \frac{\partial L_0}{\partial r_1^B} & \frac{\partial L_0}{\partial r_2^B} & \cdots \\ \frac{\partial L_1}{\partial r_1^B} & \frac{\partial L_1}{\partial r_2^B} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}, \mathbf{L}_y := \begin{bmatrix} \frac{\partial L_0}{\partial y_0} & \frac{\partial L_0}{\partial y_1} & \cdots \\ \frac{\partial L_1}{\partial y_0} & \frac{\partial L_1}{\partial y_1} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}.$$

These matrices satisfy

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

Therefore

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

Households

Define the following matrices

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

Use Equation 7 to define

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

These matrices fully capture the effect of $d\boldsymbol{y}$ and $d\boldsymbol{r^K}$ on consumption and asset demand through dr_0^A.

Now, 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.

Let

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

We define matrices that contain derivatives of $\mathcal{B}$ is the same way and we obtain:

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

C Nested Models and Extensions

C.1 Nested Models of Financial Frictions

We show how our framework nests some commonly used models of financial frictions by appropriately choosing the financial constraint $\Theta\left(\left\{r_{s+1}^B, r_{s+1}^K\right\}_{s\geq t}\right)$. We also demonstrate that in all these models financial frictions result in $\Theta_t(\cdot)$ that has the special structure we use in Lemma 2.

Gertler-Karadi-Kiyotaki[15]

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.

[15] These microfoundations of financial frictions have been used in the recent literature studying interactions between the financial sector and household heterogeneity, for example in Lee et al. (2020) and Lee (2021).

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 a bank can receive from depositors depends on the franchise value $v_{j,t}(n_{j,t})$, where $n_{j,t}$ is net worth — bank j must be better off continuing instead of running away. The optimization problem is:

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

subject to

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

The first constraint is the incentive compatibility constraint resulting from the agency problem. $\Lambda_{t,t+s}$ is the discount factor used by banks. The recursive formulation of the problem is:

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

subject to

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

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

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

First order condition with respect to $\psi_{j,t}$ is

$$\Lambda_{t,t+1} \left(f + (1-f)\eta_{j,t+1}\right) \left(r^L_{t+1} - r^D_{t+1}\right) = \frac{1}{\theta}\lambda_{j,t}$$

so

$$\eta_{j,t} = \frac{1}{1-\lambda_{j,t}} \Lambda_{t,t+1} \left(f + (1-f)\,\eta_{j,t+1}\right) \left(1 + r_{t+1}^{D}\right).$$

The guess that $v_{j,t}(n_{j,t}) = \eta_{j,t} n_{j,t}$ 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

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

If the incentive compatibility constraint is binding, we have

$$\eta_{j,t} = \Lambda_{t,t+1} \left(f + (1-f)\,\eta_{j,t+1}\right) \left[\left(r_{t+1}^{K} - r_{t+1}^{B}\right)\eta_{j,t}\theta + \left(1 + r_{t+1}^{B}\right)\right]$$

which can be rearranged as

$$\eta_{j,t} = \frac{\Lambda_{t,t+1} \left(f + (1-f)\,\eta_{j,t+1}\right)\left(1 + r_{t+1}^{B}\right)}{1 - \Lambda_{t,t+1}\left(f + (1-f)\,\eta_{j,t+1}\right)\left(r_{t+1}^{K} - r_{t+1}^{B}\right)\theta}. \tag{8}$$

As all banks face the same rates of return, the marginal value of net worth $\eta_{j,t}$ is the same for them, η_t. It follows that, if the incentive compatibility constraint is binding, $l_{j,t}^{B} = \theta\eta_t n_{j,t}$ and so if $\Lambda_{s-1,s} = 1/\left(1 + r_s^{B}\right)$ or $\Lambda_{s-1,s} = 1/\left(1 + r_s^{K}\right)$ we can write

$$l_{j,t}^{B} = \Theta\left(\left\{r_{s+1}^{B}, r_{s+1}^{K}\right\}_{s \geq t}\right) n_{j,t}.$$

Aggregating individual banks $\int_0^1 l_{j,t}^{B} dj = q_t k_t^{B}$ and $\int_0^1 n_{j,t} dj = n_t^{B}$ we obtain

$$q_t k_t^{B} = \Theta\left(\left\{r_{s+1}^{B}, r_{s+1}^{K}\right\}_{s \geq t}\right) n_t$$

which coincides with the solution to the bank's problem described in Section 2.3. In this model, 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 $\Lambda_{s-1,s} = 1/\left(1 + r_s^B\right)$, then

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

Finally, if $\Lambda_{s-1,s}$ is a constant (for example equals a household discount factor β, as in Lee et al. (2020)), we have

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

Here $\Theta = \theta\eta$, the steady state leverage ratio. We obtain these expressions by differentiating Equation 8 with respect to returns and evaluating the resulting expression at the steady state.

Bernanke, Gertler, Gilchrist (1999)

In Bernanke et al. (1999) financial frictions arise because of "costly state verification" (Townsend (1979)). 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 $\psi(1) = 1$.[16] 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

[16]There is no aggregate uncertainty in our framework and this explain 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), Eggertsson et al. (2019), Chi et al. (2021) and Cúrdia and Woodford (2011) 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, Equation 2, remains the same. More specifically, assume that the bank maximizes

$$r_{t+1}^N n_t = \max_{k_t^B, d_t} r_{t+1}^K q_t k_t^B - r_{t+1}^B d_t - \Upsilon_t \left(\frac{q_t k_t^B}{n_t} \right) n_t + \bar{\Upsilon}_t$$

subject to balance sheet $q_t k_t^B = d_t + n_t$.

Here $\Upsilon_t \left(\frac{q_t k_t^B}{n_t} \right) n_t$ 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 $\psi_t := q_t k_t^B / n_t$. First order condition is

$$r_{t+1}^K - r_{t+1}^B = \Upsilon_t' \left(\frac{q_t k_t^B}{n_t} \right),$$

which can be rewritten as

$$q_t k_t^B = \Upsilon_t'^{-1} \left(r_{t+1}^K - r_{t+1}^B \right) n_t.$$

If we define $\Theta \left(\left\{ r_{s+1}^K, r_{s+1}^B \right\}_{s \geq t} \right) := \Upsilon_t'^{-1} \left(r_{t+1}^K - r_{t+1}^B \right)$, 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 $\theta < 1$ 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

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

By using the balance sheet, we can rewrite it as

$$q_t k_t^B \leq \frac{1 + r_{t+1}^B}{1 + r_{t+1}^B - \theta \left(1 + r_{t+1}^K\right)} n_t. \tag{9}$$

We can map it to our framework by defining

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

and we have

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

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

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

Using the bank balance sheet, we have

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

The constraint differs from the one in Equation 9 in that $1 + r_{t+1}^K$ in the denominator is replaced by $\frac{q_{t+1}}{q_t}$. This form of collateral constraint is not nested in our framework exactly because $\frac{q_{t+1}}{q_t}$ is generally a function both returns on capital $\{r_s^K\}$ and output $\{y_s\}$. Yet, we expect the two collateral constraints to generate similar dynamics when most of the changes in $1 + r_{t+1}^K$ are driven by capital gain $\frac{q_{t+1}}{q_t}$.

Current-value collateral constraints

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

$$q_t k_t^B \le \frac{1}{1-\theta} n_t.$$

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, $\bar{\Theta}_{r^K} = 0, \bar{\Theta}_{r^B} = 0, \gamma = 0$.

D Bringing Model to Data

D.1 Balance Sheets

We obtain balance sheet data from the Financial Accounts of the United States (FoF), 2000Q2-2020Q2. We refer to variables with their serial numbers. For bank balance sheet, we use the Call Report data provided by Drechsler et al. (2017) on their website, which allows us to link it to the CRSP data for the market valuation of bank equity. We refer to variables from these two dataset with their variable names.

Banks: We use variables from the Call Report data, the CRSP data, and the FoF data, linking the Call Report data to CRSP using 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. We use the aggregate FoF series for the banking sector to construct the following adjustment factor

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

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."
- *market value of bank net worth:* 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."
- *effective leverage:* We construct the effective leverage of the banking sector as

$$\Theta_t \coloneqq 1 + \frac{\text{liquid liabilities - liquid assets}}{\text{market value of bank equity}}.$$

Money market funds (mmf):

- *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 (MMMFFAA027N) into liquid liabilities and equity, and impute the net worth of mmf by assuming the same effective leverage as the banking sector:

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

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 (BOGZ1FL193020005A), time and saving accounts (BOGZ1FL193030205A), 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 (BOGZ1FL193034005A), and households' holdings of treasury debt, calculated as the total government and municipal securities (BOGZ1FL193061005A) net of municipal securities (HNOMSAA027N).

- *net illiquid assets*: We calculate households' net illiquid asset holdings as their total assets (BOGZ1FL192000005A) net of liquid asset holdings defined above and their liabilities (BOGZ1FL194190005A). 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.[17]

Accouting for corporate deposits:

- 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

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

not affect the consolidated balance sheet of the fund. This is because holding a combination of these deposits in the illiquid account with the corresponding net worth of banks supplying these deposits is equivalent to directly holding capital of the same value. Specifically, consider the following modification to the model: (1) the banking sector has net worth $(1+\chi)n_t$ instead of n_t, (2) the illiquid account passively holds extra deposits χd_t that correspond to the corporate deposits in the data, and (3) capital in the illiquid account is $q_t k_t^F - \chi(n_t + d_t)$ instead of k_t^F

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

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

  Since both the goods market clearing and the liquid asset market clearing conditions are not affected, Lemma 3 implies that aggregate responses with the modifications above are identical to that from the model in Section 2.

Table 3 provides a breakdown of liquid asset positions of the household sector, the banking sector, and money market funds.

D.2 Estimation of $\Theta_{r^K}, \Theta_{r^B}$ and γ

Variable construction

Leverage ($d\Theta_t$):

- We use the linked Call-Report-CRSP-FoF data to construct the measure of effective leverage as discussed in Section D.1.
- 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-aggregateing

Table 3: Liquid asset positions

	liquid assets		liquid liabilities	
households	deposits	0.42		
	mmf shares	0.10		
	treasury	0.06		
banks	cash & reserves	0.04		
	fed funds and repo (net)	0.03		
	treasury	0.02		
			deposits	0.44
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 2000Q2 to 2020Q2.

daily market value of bank equity.

- Deviation of effective leverage away from the steady state, $d\Theta_t$, is calculated as the deviation of effective leverage from a quadratic time trend.

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

- We obtain the yield curve data on Treasury debt and corporate bonds (HQM) from the U.S. Treasury on this website (Treasury yields) and this website (HQM yields).

- We adjust the HQM yields with a constant factor so that the 30-year yield corresponds to Moody's BAA bond yields (series BAA from FRED), which better reflects the rate on prime bank loans. We obtain the adjustment factor as the coefficient from regressing BAA yields on 30-year HQM yields.

- We use yields on securities with maturity of 6 months and 1, 2, 3, 5, 7, 10, 20, and 30 years, aggregating observations to a monthly frequency.

- We construct real yields by subtracting expected inflation from nominal yields. We use inflation expectations data from the Cleveland Fed on this website.

- We calculate deviations of real yields from a quadratic trend, and we add back the means.

- We calculate forward rates between the maturities we observe and extend the forward rates to all horizons with a left-continuous step function.
- 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 rate from the mean.

Table 4 shows summary statistics of selected variables we constructed and used in our estimation:

Table 4: Standard Deviation of Detrended Effective Leverage and Forward Rates

$d\Theta_t$	$\mathbb{E}_t\left[dr^K_{t+1\text{ y}}\right]$	$\mathbb{E}_t\left[dr^K_{t+5\text{ y}}\right]$	$\mathbb{E}_t\left[dr^K_{t+10\text{ y}}\right]$	$\mathbb{E}_t\left[dr^B_{t+1\text{ y}}\right]$	$\mathbb{E}_t\left[dr^B_{t+5\text{ y}}\right]$	$\mathbb{E}_t\left[dr^B_{t+10\text{ y}}\right]$
0.997	0.106%	0.066%	0.057%	0.084%	0.040%	0.032%

Estimation

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

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

where $X_t = \left\{\mathbb{E}_t[dr^K_{t+\tilde{h}}], \mathbb{E}_t[dr^B_{t+\tilde{h}}]\right\}_{\tilde{h}\in\mathcal{H}}$, $\mathcal{H} = \{6 \text{ months}, 1, 2, 3, 5, 7, 10, 20, 30 \text{ years}\}$. As explained in Section 5.2, the identification assumption we make is that the effective leverage can move only in response to macroeconomic conditions only through its response to returns. We do not allow for shocks to $d\Theta_t$ that would be correlated with returns. Any deviation of $d\Theta_t$ from the formula implied by Lemma 1 must be attributed to measurement error (uncorrelated with returns). Note that this does not mean that we rule out all shocks to the financial sector – liquidity supply is allowed to move also in response to shocks that directly affect net worth.

For the estimation result in Table 2:

- We use a two-step GMM with the optimal weighting matrix, and a quadratic spectral kernel to compute the covariance matrix of the vector of sample moment conditions.

- We search for the minimum of the objective function by applying the following procedure. (1) We create a coarse grid of 75 equidistant points between 5 and 100 for $\bar{\Theta}_{r^K}$ and $\bar{\Theta}_{r^B}$, and between 0.0 and 0.999999 for γ. (2) We perform a grid search to minimize the sum of squared moment conditions (this corresponds to using an identity matrix as a weighting matrix). (3) We then create a denser grid: 75 points between 10 and 30 for $\bar{\Theta}_{r^K}$ and $\bar{\Theta}_{r^B}$ and between 0.92 and 0.99999 for γ. This new grid contains the minimum found in the previous step. (4) We repeat the grid search. We then use the minimum found in the second step as a starting point and use simulated annealing to estimate $\bar{\Theta}_{r^K}, \bar{\Theta}_{r^B}, \gamma$ with a two-step GMM with the optimal weighting matrix.

D.3 Leverage Shocks and Estimation Bias

We now discuss the estimation bias due to violation of Equation 6. We show the direction of bias for estimates of $\bar{\Theta}_{r^K}$ and $\bar{\Theta}_{r^B}$ in a simplified case with $\gamma = 0$ as in Bernanke et al. (1999).

Christiano et al. (2014) extend Bernanke et al. (1999) and introduce shocks to the variance of idiosyncratic returns. These shocks shift the spread-leverage relationship and, up to the first order approximation, are ismorphic to adding an exogeneous shifter θ_t in

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

Christiano et al. (2014) label these shocks "risk shocks" and argue, through the lens of a medium-scale New Keynesian DSGE model, that these shocks are an important driver of business cycle fluctuations.

For simplicity, assume that $\bar{\Theta}_{r^K} = \bar{\Theta}_{r^B}$ and define $ds_{t+1} \coloneqq \mathbb{E}_t\left[dr_{t+1}^K - dr_{t+1}^B\right]$. Let θ_t be an exogenous shock that affects the leverage ratio. Assume it is uncorrelated with all other shocks. The true relationship is $d\Theta_t = \bar{\Theta}_{r^K} ds_{t+1} + \theta_t$, and its empirical counterpart

$$d\Theta_t = \bar{\Theta}_{r^K} ds_{t+1} + \upsilon_t.$$

The OLS estimator for $\bar{\Theta}_{r^K}$ satisfies

$$\mathbb{E}\left[\hat{\bar{\Theta}}_{r^K}\right] = \frac{\mathbb{E}\left[d\Theta_t \cdot ds_t\right]}{\mathbb{E}\left[ds_{t+1}^2\right]} = \bar{\Theta}_{r^K} + \frac{\mathbb{E}\left[\theta_t \cdot ds_{t+1}\right]}{\mathbb{E}\left[ds_{t+1}^2\right]}$$

Let $\boldsymbol{z}_t$ be the vector of all state variables except θ_t. These might include the stock of capital, distributions of assets, and various structural shocks. Assume $\mathbb{E}\left[\theta_t \mid \boldsymbol{z}_t\right] = 0$ Up to a first order approximation, in equilibrium

$$ds_{t+1} = \alpha_\theta \theta_t + \sum_k \alpha_{z_k} z_{k,t}.$$

It immediately follows that

$$\mathbb{E}\left[\hat{\bar{\Theta}}_{r^K}\right] = \bar{\Theta}_{r^K} + \alpha_\theta \frac{\mathbb{E}\left[\theta_t^2\right]}{\mathbb{E}\left[ds_{t+1}^2\right]}$$

and the sign of α_θ determines the direction of the bias. The bias is positive if $\alpha_\theta > 0$ and negative if $\alpha_\theta < 0$. The first case corresponds to leverage shocks increasing the expected spread, and the second case to leverage shocks decreasing the expected spread. Holding net worth constant, higher leverage should lead to more capital accumulation and reduce the expected spread. Moreover, a fall in the expected rate of return on capital is consistent with an increase in the price of capital in the current period, which increases net worth. This is indeed what Christiano et al. (2014) shows in Figure 4 of their paper. We conclude that $\alpha_\theta < 0$ is more likely to be empirically relevant and thus

$$\mathbb{E}\left[\hat{\bar{\Theta}}_{r^K}\right] - \bar{\Theta}_{r^K} < 0.$$

D.4 Robustness Check with respect to Parameter f

Figure 9 shows the relative effectiveness of asset purchases and tax cuts for different values of f. The red line represents our baseline specification in Figure 8. The gray shades from light to dark represent deviations from the baseline model for $f \in [0.01, 1]$. For comparison, the blue and black lines are the perfectly inelastic and perfectly elastic cases in Figure 8, which do not depend on f. The proximity of the gray shades and the red line indicates that for the comparison of the two alternative policies, the specification of parameter f is inconsequential. This is in sharp contrast to the differences in output responses resulting from $\bar{\Theta}_{r^K}$ as shown in Figure 8.

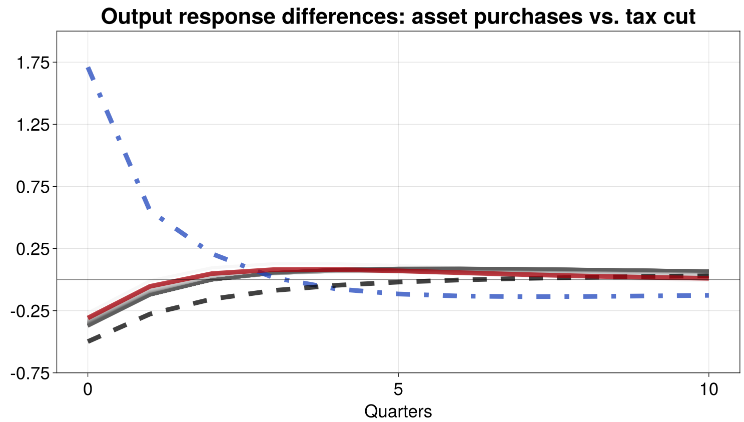

Figure 9: Difference between output response to asset purchases and tax cuts. The y-axis: % of steady-state GDP. Red: empirical elasticities. Light to dark gray: $f \in [0.01, 1]$. Blue: perfectly inelastic. Black: perfectly elastic.

E Generalization and Extension

E.1 Generalized net worth process

State-dependent exogenous equity injection

So far, we assumed that equity injections are constant, m. We now relax this assumption and allow $m_t = \xi \left(1 + r_t^K\right) q_{t-1} k_{t-1}^B$ with $\xi \geq 0$, as in Gertler and Kiyotaki (2010). Totally differentiating Equation 2 and evaluating at the steady state results in

$$dn_t = (1-f)\left[\left(\bar{r}^K - \bar{r}^B\right) d\Theta_{t-1} + \left(dr_t^K - dr_t^B\right)\bar{\Theta} + dr_t^B\right]\bar{n} \\ + (1-f)\left[\left(\bar{r}^K - \bar{r}^B\right)\bar{\Theta} + \left(1 + \bar{r}^B\right)\right] dn_{t-1} + dm_t$$

where $dm_t = \xi \left[\bar{\Theta}\bar{n} dr_t^K + \left(1 + \bar{r}^K\right)\left(\bar{n} d\Theta_{t-1} + \bar{\Theta} dn_{t-1}\right)\right]$. The linearized law of motion for n_t is

$$dn_t = (1-f)\sum_{u=0}^{t} G^u \left[\left(\bar{r}^K - \bar{r}^B - \xi \frac{1+\bar{r}^K}{1-f}\right) d\Theta_{t-1-u}\bar{n} + \left(\left(1 + \frac{\xi}{1-f}\right) dr_{t-u}^K - dr_{t-u}^B\right)\bar{\Theta}\bar{n}\right] \\ + (1-f)\sum_{u=0}^{t} G^u dr_{t-u}^B \bar{n},$$

where $G := (1-f)\left[\left(\bar{r}^K - \bar{r}^B\right)\bar{\Theta} + \left(1+\bar{r}^B\right) + \xi\frac{1+\bar{r}^K}{1-f}\bar{\Theta}\right] \geq 0$. Observe that the form of the above expression is the same as with $m_t = m$. The only difference is in coefficients. Consider a particular variation such that $dr_s^K = 1$ and $dr_u^K = 0$ for all $u \neq s$, and $dr_u^B = 0$ for all u. We have

$$dn_t = \begin{cases} \bar{n}\,(1-f)\left(\left(\bar{r}^K - \bar{r}^B - \xi\frac{1+\bar{r}^K}{1-f}\right)\sum_{u=t-s}^{t-1} G^u \frac{\partial\Theta_{t-1-u}}{\partial r_s^K} + \left(1+\frac{\xi}{1-f}\right)G^{t-s}\bar{\Theta}\right), & s \leq t, \\ \bar{n}\,(1-f)\left(\bar{r}^K - \bar{r}^B - \xi\frac{1+\bar{r}^K}{1-f}\right)\sum_{u=0}^{t-1} G^u \frac{\partial\Theta_{t-1-u}}{\partial r_s^K}, & s > t. \end{cases}$$

Finally, define $\tilde{\Sigma}(s) := (1-f)(\bar{r}^K - \bar{r}^B - \xi\frac{1+\bar{r}^K}{1-f})\frac{1-(\gamma G)^s}{1-\gamma G}$ and divide by $\left(\bar{\Theta}-1\right)\bar{n}$ to get

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

Endogenous equity injection

Karadi and Nakov (2021) solve a version of Gertler-Karadi-Kiyotaki model with optimal equity injections. The optimization problem in their model gives $m_t = \xi_{t-1}n_{t-1}$, where $\xi_{t-1} = \zeta\Lambda_{t-1,t}(1-f)(\eta_t - 1)$, and η_t denotes the marginal value of net worth in a Gertler-Karadi-Kiyotaki model, as defined in Appendix C.1.

Linearization gives

$$dm_t = \zeta\Lambda(1-f)nd\eta_t + \xi dn_{t-1} + \zeta\,(1-f)\,(\eta-1)\,nd\Lambda_{t-1,t}.$$

From Appendix C.1, the marginal value of net worth in a Gertler-Karadi-Kiyotaki model satisfies $d\Theta_t = \theta d\eta_t$, and therefore $d\xi_{t-1} = \zeta\Lambda(1-f)\frac{1}{\theta}d\Theta_t$. Let $\psi := \zeta\Lambda\frac{1}{\theta}$ and $\omega := \zeta\,(\eta-1)$; we can write

$$\begin{aligned} dn_t = {} & (1-f)\left[\left(r^K - r^B\right)d\Theta_{t-1} + \Theta\left(dr_t^K - dr_t^B\right) + dr_t^B + \psi d\Theta_t + \omega d\Lambda_{t-1,t}\right]n \\ & + (1-f)\left[\left(r^K - r^B\right)\Theta + 1 + r^B + \xi\right]dn_{t-1}. \end{aligned}$$

If the bank's discount rate is $\Lambda_{t-1,t} = 1/\left(1+r_t^K\right)$, then $d\Lambda_{t-1,t} = -1/(1+r^K)^2 dr_t^K$, and

$$dn_t = (1-f)\left[\left(r^K - r^B\right) d\Theta_{t-1} + \Theta\left(dr_t^K - dr_t^B\right) + dr_t^B + \psi d\Theta_t + \tilde{\omega} dr_t^K\right] n$$
$$+ (1-f)\left[\left(r^K - r^B\right)\Theta + 1 + r^B + \xi\right] dn_{t-1}, \quad \tilde{\omega} \coloneqq -\omega/(1+r^K)^2.$$

Let $G \coloneqq (1-f)\left[\left(r^K - r^B\right)\Theta + 1 + r^B + \xi\right]$, define $\sigma(s) \coloneqq \frac{1-(G\gamma)^s}{1-G\gamma}$ and use

$$\frac{\partial \Theta_{t-u}}{\partial r_s^K} = \begin{cases} 0, & s \leq t-u, \\ \gamma^{s-t+u-1}\bar{\Theta}_{r^K}, & s > t-u, \end{cases}$$

to write

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

E.2 Limiting Cases: 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 $a_t^G = 0$,[18] 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.

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

In our quantitative study in Section 6 we follow a different strategy. We want to keep the steady state the same for all models to isolate the role of liquidity supply elasticities. This would not be possible with $d_t = 0$. We set the matrices $\mathbf{D}_{r^K}, \mathbf{D}_{r^B}, \mathbf{D}_y$ to be identically zero. This can be done by assuming $f = 1$ (which ensures that net worth remains constant) 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,[19] the firms' problem is

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

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

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

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

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

[19] 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).

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 1, 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 $1 + r^K_{t+1} = \frac{1}{1-\zeta}(1 + r^B_{t+1}) = 1 + r^A_{t+1} \quad \forall t \geq 0$.
2. Servicing one unit of government debt (in time t goods) issued at time t costs $(1 + r^B_{t+1})/(1 - \zeta)$ units of goods in period $t + 1$.
3. The goods market clearing requires $c_t + x_t + g_t + \frac{\zeta}{1-\zeta}(1 + r^B_t) \int b_{i,t-1} di = y_t$.

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 $r^A_{t+1} = r^K_{t+1}$ even with $d_t > 0$. Assume that the passive mutual fund holding capital directly and bank equity has intermediation cost

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

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

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

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

The sum of intermediation costs in period t is

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

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.