SIEPR | STANFORD INSTITUTE FOR ECONOMIC POLICY RESEARCH

THE SHORT RATE DISCONNECT IN A MONETARY ECONOMY

Moritz Lenel
Princeton University

Monika Piazzesi
Stanford University & NBER

Martin Schneider
Stanford University & NBER

July, 2019

Working Paper No. 19-026

The short rate disconnect in a monetary economy *

Moritz Lenel	Monika Piazzesi	Martin Schneider
Princeton	Stanford & NBER	Stanford & NBER

July 15, 2019

Abstract

In modern monetary economies, most payments are made with inside money provided by payment intermediaries. This paper studies interest rate dynamics when payment intermediaries value short bonds as collateral to back inside money. We estimate intermediary Euler equations that relate the short safe rate to other interest rates as well as intermediary leverage and portfolio risk. Towards the end of economic booms, the short rate set by the central bank disconnects from other interest rates: as collateral becomes scarce and spreads widen, payment intermediaries reduce leverage, and increase portfolio risk. We document stable business cycle relationships between spreads, leverage, and the safe portfolio share of payment intermediaries that are consistent with the model. Structural changes, especially in regulation, induce low frequency shifts, such as after the financial crisis.

*Email addresses: lenel@princeton.edu, piazzesi@stanford.edu, schneidr@stanford.edu.

We thank Hengjie Ai, Alan Moreira, Tyler Muir, Ricardo Reis, Oreste Tristani, and Mike Woodford for helpful comments. We also thank the NSF for a research grant that supported this project.

1 Introduction

Current research on monetary policy relies heavily on standard asset-pricing theory. Indeed, it assumes the existence of real and nominal pricing kernels that can be used to value all assets. Moreover, the central bank's policy rate is typically identified with the short rate in the nominal pricing kernel. With nominal rigidities as in the New Keynesian framework, the central bank then has a powerful lever to affect valuation of all assets – nominal and real – and hence intertemporal decisions in the economy. Focus on this lever makes the pricing kernel a central element of policy transmission.

In spite of its policy relevance, empirical support for monetary asset pricing models has been mixed at best. Models that fit the dynamics of long duration assets, such as equity and long term bonds, often struggle to also fit the policy rate. This is true not only for consumption based asset-pricing models that attempt to relate asset prices to the risk properties of growth and inflation, but also for more reduced-form approaches to describe the yield curve. The finding is typically attributed informally to a convenience yield on short term debt. We refer to it as the "short rate disconnect".[1]

This paper proposes and quantitatively assesses a theory of the short rate disconnect that is based on the role of banks in the payment system. We start from the fact that short safe instruments that earn the policy rate are predominantly held by payment intermediaries, in particular commercial banks and money-market mutual funds. We argue that these intermediaries, which we call "banks" throughout this paper, are on the margin between short safe debt and other fixed income claims. We derive new asset-pricing equations that relate the short rate to bank balance-sheet ratios. We show that those equations account quite well for the short rate disconnect, especially at business cycle frequencies.

Our asset-pricing equations follow from the fact that banks issue short nominal debt used for payments. In our model, leverage requires collateral, and the ideal collateral to back short nominal debt is in turn short nominal debt. When such debt becomes more scarce, its equilibrium price rises and the short interest rate falls. In particular, the market short rate disconnects from the short rate of the nominal pricing kernel used to value other assets, such as long term bonds or equity.

Empirically, our approach places restrictions on the joint dynamics of the yield curve and bank-balance sheets that we evaluate with US data since the 1970s. Our measure of the short

[1] In the literature on arbitrage-free factor models of the yield curve, the idiosyncratic variation of the short rate was first documented by Duffee (1996). Piazzesi (2005) shows that accounting for this variation in the policy rate jointly with longer rates requires a carefully-chosen fourth factor beyond the traditional three factors that capture well the dynamics of longer rates. Piazzesi and Schneider (2007) study a consumption-based pricing model with recursive utility that works well for rates one year and longer, but not at the short end.

rate disconnect is the spread between a "shadow" short rate – measured as the short end of a yield-curve model estimated with only medium and long maturity Treasury rates – and the three month T-bill rate. This "shadow spread" captures a cost of safety for banks and rises consistently at the end of booms. As safe collateral becomes scarce and its cost increases, banks increase their share of risky collateral and thereby have a riskier portfolio overall. At the same time, banks lower risk by reducing their leverage in booms, as our theory predicts.

An important feature of the model is that banks choose both their leverage ratios and the share of short safe bonds in their asset portfolio. Optimal leverage trades off low funding costs (due to the liquidity benefit of inside money) against an increasing marginal cost of leverage, as in models with bankruptcy costs. The optimal safe portfolio share trades off the higher return on risky instruments against the cost of backing inside money with worse collateral. Leverage and portfolio safety optimally go together: a safe bank faces a lower cost of leverage and can more cheaply produce inside money.

Adjustment along both margins is crucial for our model to account for the comovement of the safe share, leverage, and the shadow spread in the data. Indeed, a higher shadow spread makes safe bonds more costly to hold, and banks shift their asset portfolio towards risky instruments. As a result, leverage becomes more costly and is optimally reduced. In contrast, a lower shadow spread pushes banks towards safer collateral and higher leverage. Through this mechanism, the model successfully captures the dynamics of bank-balance sheets at business cycle frequencies: a procyclical shadow spread goes along with countercyclical leverage and procyclical portfolio risk taking.

We also study lower frequency movements in balance sheet ratios and spreads, with an emphasize on the role of macro-prudential regulation. The model predicts low bank leverage in the 1980s when the shadow spread was particularly high. It does so even assuming that bank balance sheet costs remained unchanged for the last four decades. At the same time, our results suggests that the 2008 financial crisis triggered a structural break: after 2008, higher balance sheet costs induced banks to hold more safe assets – including much larger share of reserves – and as a result produce more inside money relative to assets. More generally, once we allow for slow moving shifts in balance sheet cost parameters over the whole sample, the model captures the low frequency trends in bank leverage while maintaining the fit at business cycle frequencies.

Our results call into question the traditional account of how monetary policy is transmitted to the real economy. The existence of a volatile shadow spread implies that the central bank does not control the short rate of the nominal pricing kernel. Interest rate policy thus relies on pass-through from the policy rate to the shadow rate and hence to intertemporal decisions in the economy. Our bank-based asset-pricing equations suggest that the transmission of monetary policy works at least to some extent through bank-balance sheets. As a result, monetary policy and macroprudential policy are likely to both matter for the course of interest rates.[2]

Formally, our model describes the behavior of a banking sector that maximizes shareholder value subject to financial frictions. We capture the rest of the economy by two standard elements: a pricing kernel used by investors to value assets – in particular bank equity – and a broad money demand equation that relates the quantity of deposits to their opportunity cost. Our approach is thus in the spirit of consumption-based asset pricing pioneered by Breeden (1979) and Hansen and Singleton (1983): we test valuation equations that must hold in general equilibrium, without taking a stand on many other features of the economy, in particular the structure of the household sector, and the technology and pricing policy of firms.[3]

In our model, the key friction faced by banks is that delegated asset management is costly, and more so if it is financed by debt. We assume that a bank financed by equity requires a proportional balance-sheet cost per unit of assets. If the bank also issues deposits, the resource cost per unit of assets is increasing in bank leverage, so banks' marginal cost of debt is upward sloping. One interpretation is that debt generates the possibility of bankruptcy, which entails deadweight costs proportional to assets. An upward sloping marginal cost of debt implies that the portfolio choice of levered banks looks *as if* the banks are more risk averse than their shareholders.[4] In particular, since banks issue short nominal debt, they value safe short nominal debt as collateral. It is this collateral benefit of short debt that generates the short rate disconnect in our model.

We solve banks' optimization problem and evaluate their first order conditions, given an incomplete asset market structure: banks can invest in reserves, short safe bonds that earn the policy rate, as well as risky assets that stand in for other fixed income claims available to banks such as loans. We show that there is no short rate disconnect when bank assets are safe, that is, banks only hold reserves and short nominal bonds. More generally, however, the collateral

[2]Piazzesi, Rogers and Schneider (2019) study the transmission of monetary policy in a New Keynesian model with banks. They show that the short rate disconnect tends to dampen the effect of interest rate policy. Moreover, pass-through from the policy rate to the shadow rate depends on bank-balance sheets, and in particular on the role of reserves: it is stronger in a corridor system with scarce reserves as opposed to a floor system with ample reserves.

[3]In particular, our approach is equally consistent with the supply side of a real business cycle and that of a New Keynesian model. In both cases, a pricing kernel and a money demand function can be derived from representative agent optimization. Our model is also consistent with heterogeneous agent models as long as there is a set of state prices that determines the shareholder value of banks.

[4]We emphasize that this property holds even though banks face no adjustment costs of equity. Moreover, in contrast to much of the literature our model does not rely on segmentation of equity markets that gives rise to "bankers" who are risk averse agents distinct from "households". Instead, household own banks and choose to hold some assets directly and others indirectly through banks, as in Piazzesi and Schneider (2018).

benefit of short bonds generates a wedge between the market short rate and the short rate in the nominal pricing kernel. The resulting shadow spread is higher when banks have a larger share of their portfolio invested in risky assets: risky banks place a particularly high value on short nominal bonds relative to other investors, such as bank shareholders. The banks' optimization problem also implies that when the shadow spread is high, banks counteract the increase in risk on their asset side by reducing risk on their liability side. During these times, banks thus reduce their leverage.

Quantitative assessment of our theory requires data on balance sheets. To measure the positions of payment intermediaries, we consolidate banks and money market funds: both institutions offer payments services to households and corporations. We further define safe assets as assets with short maturity that are nominally safe (such as reserves, vault cash, and government bonds). Finally, we define leverage as the ratio of inside money to total fixed income assets net of other debt that can be viewed as senior to inside money. To measure inside money, we use a broad concept of money that includes money market accounts. The raw fact that provides evidence for our mechanism is that payment intermediaries have a portfolio share of safe assets as well as a leverage ratio that are strongly negatively correlated with the shadow spread, both at business cycle frequencies and over longer periods. [5]

Our approach builds on the idea that bonds earn a convenience yield, pioneered by Patinkin (1956) and Tobin (1963)). Recent examples include Bansal and Coleman (1996), Krishnamurthy and Vissing-Jorgensen (2012), Venkateswaran and Wright (2014), Andolfatto and Williamson (2015), Nagel (2016), and Woodford (2016). In these models, the convenience yield reflects a nonpecuniary benefit to investors who hold the bonds, analogously to a convenience yield on money: for example, bonds enter the utility function or relax cash-in-advance constraints. In our model, in contrast, investors receive a nonpecuniary benefit from inside money, but do not hold short bonds directly. Indeed, they perceive short bonds as too expensive because the short rate reflects a convenience yield earned by banks that supply inside money. Investors receive the convenience yield on short bonds – and the nonpecuniary benefit of lower balance-sheet cost – only indirectly as bank shareholders.

Our model thus contributes to the growing literature on intermediary-based asset pricing that studies equilibrium relationships between asset prices and balance-sheet ratios.[6] However, while the literature has focused on assets that are held by intermediaries because of their

[5]Our analysis thus treats money market funds as highly levered banks that also hold risky and riskfree assets. The key facts we emphasize also hold for commercial banks alone.

[6]Examples include Brunnermeier and Pedersen (2009), He and Krishnamurthy (2013), Brunnermeier and Sannikov (2014), Adrian, Etula and Muir (2014), Greenwood and Vayanos (2014), Koijen and Yogo (2015), Bocola (2016), Moreira and Savov (2017), He, Kelly and Manela (2017), Muir (2017), Haddad and Sraer (2018), and Haddad and Muir (2018).

complexity – for example, mortgage-backed securities or credit-default swaps – our intermediaries price what is arguably one of the simplest assets: short nominal bonds. At the same time, our approach is not inconsistent with the presence of convenience yields in assets other than short bonds. For example, the model in Lenel (2018) incorporates a convenience yield on long bonds earned by hedge funds that use such bonds as collateral. Through the lens of the current model, this convenience yield is incorporated into the pricing kernel of investors.

Our theory is based on the scarcity of safe short assets available to banks, measured by the shadow spread. A related, but distinct, concept is the scarcity or reserves, measured by the spread between a market short rate (such as the three month T-bill rate) and the interest rate on reserves. The distinction has come into sharp focus recently as central bank operating procedures have changed. Indeed, after quantitative easing increased the quantity of reserves in 2008, the spread between T-bill and reserve rates turned negative. In contrast, the short rate disconnect we document is present both before and after 2008. Our paper is thus only tangentially related to work on bank liquidity management that relates bank behavior to the level of the short rate (for example, Bhattacharya and Gale 1987, Whitesell 2006, Cúrdia and Woodford 2011, Reis 2016, Bianchi and Bigio 2014, Drechsler, Savov and Schnabl 2018, De Fiore, Hoerova and Uhlig 2018.) In general, we would expect bank Euler equations for both safe and liquid assets to hold jointly. For example, Piazzesi and Schneider (2018) consider a model that incorporates both bank liquidity management and a scarcity of bank collateral as in the present paper and derive its implications for monetary policy.

In the wake of the recent financial crisis, a growing literature studies monetary policy when banks face financial frictions. One strand assumes that banks have a special ability to lend, and hence add value via positions on the asset side of their balance sheets (for example, Cúrdia and Woodford 2010, Gertler and Karadi 2011, Gertler, Kiyotaki and Queralto 2012, Christiano, Motto and Rostagno 2014, Brunnermeier and Sannikov 2016, Christiano, Motto and Rostagno 2012, Del Negro, Eggertsson, Ferrero and Kiyotaki 2017, and Brunnermeier and Koby (2018). These papers also distinguish assets priced by banks – for example bank loans – from assets priced by households, which include the policy instrument. Policy transmission depends on pass-through from the policy rate (which aligns with households' expected marginal rate of substitution) to the loan rate and hence to bank-dependent borrowers.

Our paper assumes that banks have a special ability to provide inside money as a medium of exchange. We share this "liability centric" view of banking with e.g. Williamson (2012), Hanson, Shleifer, Stein and Vishny (2015), Williamson (2016), Begenau (2019), and Diamond (2019). As in these papers, banks' portfolio choice in our model is shaped by banks' ability to fund themselves with deposits. In our case, banks value short safe debt as particularly good collateral for inside money, which serves as the only medium of exchange.

2 The short rate disconnect in the data

Our theory implies that the interest rate on nominal safe short bonds reflects valuation by payment intermediaries, whereas other bonds – including longer Treasuries – may be priced directly by investors. The shadow rate – the short rate in investors' pricing kernel – is thus not directly observable in the market since investors do not hold short bonds.

However, we can derive an estimate of the shadow rate from the prices of longer safe bonds. Indeed, interest rates of different maturities are connected: long rates should be risk-adjusted expectations of averages of future short rates. This principle motivates parsimonious yield-curve models that exploit the strong factor structure in data on interest rates of different maturities. These models jointly describe the dynamics of short and long rates in terms of a few factors. Estimation of such models does not require data on all maturities, but instead exploits their strong comovement.

Figure 1 shows the average nominal Treasury yield curve in the US for a quarterly sample from 1973 to 2018. The data for maturities of 1 year and longer is from Gurkaynak, Sack and Wright (2007). The data on the 3-month T-bill rate is from FRED. The plot shows that the average curve is upward sloping and concave. The plot also reveals that, on average, the short end of the Treasury curve is far below the longer maturity rates.

To obtain a shadow short rate, we use the model and estimates of Gurkaynak, Sack and Wright (2007) who construct zero-coupon bond yields from data on Treasury *bonds* but not Treasury *bills*. The paper estimates a five-factor model of the yield curve with only data on Treasury bonds. This approach thus leaves out precisely those instruments that payment intermediaries like to hold as collateral for inside money. It also restricts attention to bonds with remaining maturity longer than three months. We compute the short rate from their estimated curve. Details are in the Online Appendix.

Figure 1 shows the average shadow rate as a grey dot, roughly 30 basis points above the average 3-month T-bill rate. The plot thus illustrates the familiar short-rate disconnect derived from arbitrage-free yield-curve models: when these models are fitted to only data on longer-maturity rates, they imply a short rate that is higher than the observed short rate. Our strategy here is similar to Greenwood, Hanson and Stein (2015) who want to measure the convenience yield of T-bills relative to longer Treasury bonds.

We emphasize that the disconnect does not imply that there does not exist a pricing kernel that could jointly price all Treasury bills and bonds. Instead, the evidence in Figure 1 illustrates the well-known finding in the arbitrage-free yield-curve literature that a model that is fit to data on exclusively long rates will do poorly in matching the short rate. An example of an arbitrage-free model that is able to simultaneously match movements in short and long rates is Piazzesi (2005), who introduces a special factor that is designed to capture movements that are idiosyncratic to short rates. Consistent with this evidence, our model below will derive a pricing kernel for banks that will likewise price all Treasuries jointly.

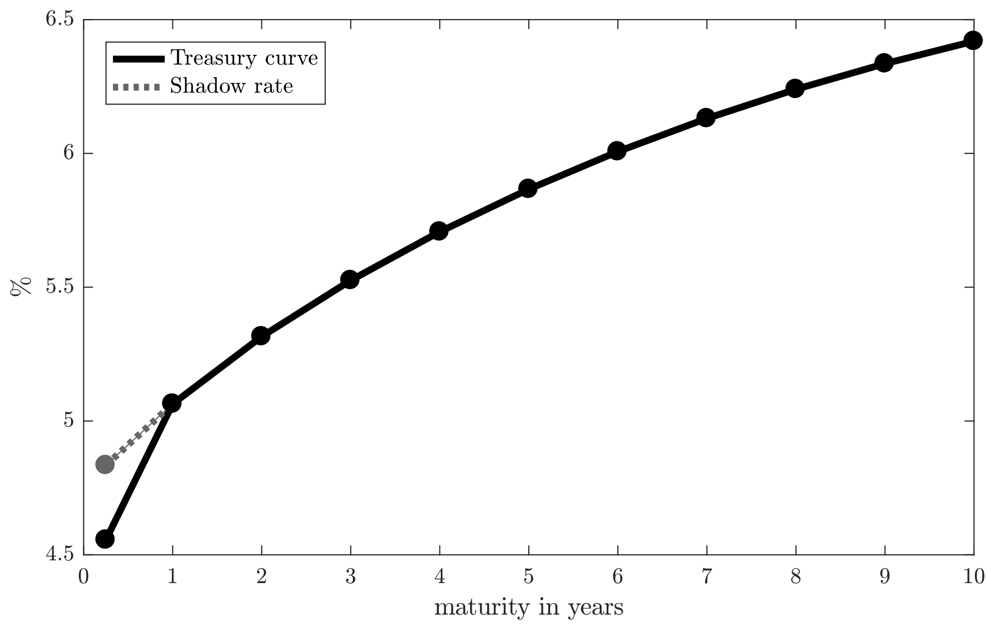

Figure 1: Average Treasury yield curve in quarterly data as black solid line, 1973-2018. The grey dot is our measure of the shadow rate, the 3-month rate implied by the estimated yield-curve model in Gurkaynak, Sack and Wright (2007).

Figure 2 plots the shadow spread, the difference between the shadow rate and the 3-month T-bill rate, measured on the left axis. The spread is higher towards the beginning of the sample, tends to rise at the end of booms, and falls during the shaded NBER recessions. The spread shares these properties with the level of the T-bill rate, shown as a grey line measured on the right axis. We emphasize also that the shadow spread remains positive during the zero lower bound period post-2008: in fact it is not unusually low during this period.

For robustness, we obtain the analogue to Figure 1 with constant maturity Treasury rates from FRED. The resulting plot also shows a short rate disconnect. Moreover, we compute alternative measures of the shadow rate. For example, we linearly extrapolate the 1-year and 2-year Treasury rate to a 3-month maturity, or we fit a cubic polynomial through the longer-maturity Treasury rate and then extrapolate the 3-month rate. These alternative measures lead to shadow spreads that are highly correlated with the series in Figure 2.

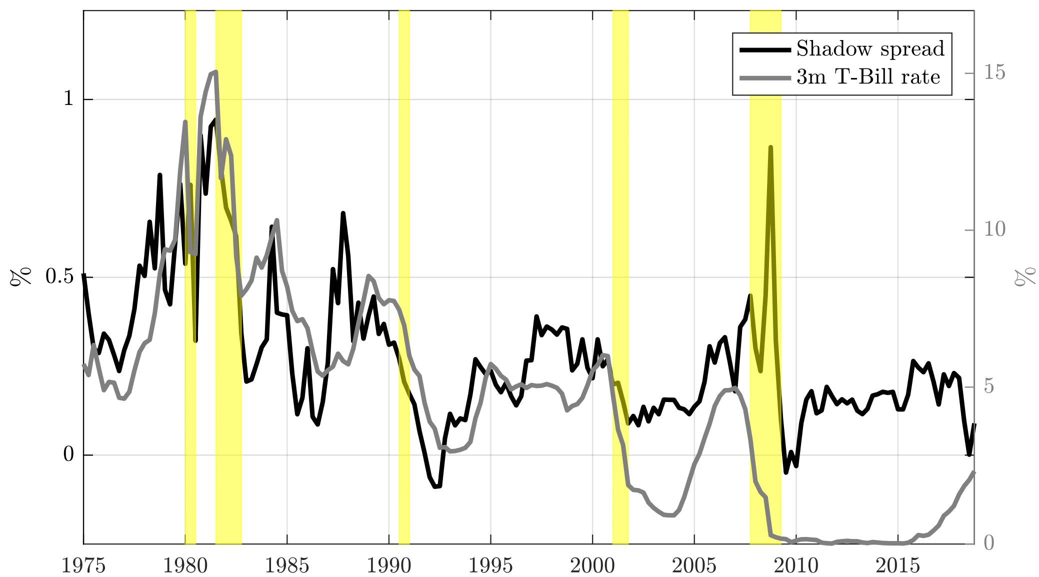

Figure 2: Shadow spread and 3-month T-bill rate. The black line is the difference between the shadow rate and the 3-month T-bill rate in percent measured along the left vertical axis. The grey line is the 3-month T-bill rate in percent measured along the right vertical axis. NBER recessions are shaded.

3 A model of the short rate disconnect

We study an economy with a single consumption good and an infinite horizon. There is a group of agents, whom we will call "investors", who hold bank equity as well as other risky assets directly. Investors use inside money as a payment instrument. The inside money is provided by competitive banks. We do not model in detail what the investor sector does: Section 3.1 simply summarizes how that sector values assets, including inside money. With this approach, we can focus on a model mechanism that is robust to what exactly the "real economy" looks like. Section 3.2 then lays out the problem of the banking system, and Section 3.3 derives the key asset-pricing conditions that must hold in equilibrium.

3.1 Environment and preferences

Let M_{t+1} denote the real pricing kernel for investors. It is a random variable that represents the date t value, in consumption goods, of contingent claims that pay off one unit of the consumption good in various states of the world at date $t+1$, normalized by the relevant conditional probabilities. For example, in an economy with a representative household, M_{t+1} is equal to the household's marginal rate of substitution between wealth at dates t and $t+1$.

The price of any asset held by investors in equilibrium is given by the present value of payoffs – in consumption goods – discounted with the pricing kernel. In particular, the value of a bank is given by the present value of its payout to shareholders, to be described below. Moreover, we think of this pricing kernel as determining real intertemporal decisions in the economy.

Since we are interested in nominal interest rates, it is helpful to introduce additional notation for the valuation of nominal claims. Let P_t denote the price of goods in terms of dollars and define the nominal pricing kernel as $M^{\$}_{t+1} = M_{t+1}P_t/P_{t+1}$. With this change of numeraire, $M^{\$}_{t+1}$ represents (normalized) date t values, in dollars, of contingent claims that pay off one dollar in various states of the world at date $t+1$. We also define a nominal one period safe interest rate by

$$1 = E_t\left[M^{\$}_{t+1}\right](1 + i^S_t). \tag{1}$$

We refer to i^S_t, the short rate in the nominal pricing kernel, as the *shadow rate*.

We assume that investors cannot borrow at the nominal rate on short safe debt, denoted i^B_t. This assumption is sensible as long as private agents cannot issue perfectly safe debt; only the government can do that. It implies that the shadow rate serves as an upper bound on the market nominal rate on short safe debt. The two rates are equal only if investors directly hold short safe debt. The short rate disconnect occurs when the market rate drops below the shadow rate. In this case, investors perceive short nominal bonds as too expensive and do not hold them directly. As we will see, this scenario is consistent with equilibrium because banks may value short nominal bonds more than investors.

Finally, consider the valuation of inside money, or deposits, by investors. We assume that investors rely on deposits to make transactions and are therefore willing to accept an interest rate on deposits i^D_t that is below the shadow rate. The opportunity cost of money $i^S_t - i^D_t$ reflects the value of money for making payments. It is declining in real balances held by the rest of the economy: the marginal benefit of payment instruments is declining in the overall quantity held. Formally, we model the payment benefits as a decreasing, convex "money demand" function v:

$$v_t(D_t/P_t) = \frac{i^S_t - i^D_t}{1 + i^S_t}, \tag{2}$$

where D_t denotes the dollar value of deposits, or inside money. The dependence on t here stands in for other forces that affect money demand, for example the level of consumption.

3.2 Payment intermediaries

Payment intermediaries provide inside money to investors. In the U.S. economy, they consist not only of traditional depository institutions but also of money-market funds. We consolidate all payment intermediaries and refer to them as "banks" for short.[7] Banks issue nominal deposits D_t to the rest of the economy and purchase assets worth A_t dollars to back those deposits. They maximize shareholder value. We allow shareholders to freely adjust equity every period and hence focus on a one period ahead portfolio and leverage choice.

Banks have access to three classes of assets: short nominal safe debt that pays the market rate i_t^B, reserves and risky bonds. Reserves are short safe bonds that pay a nominal reserve rate i_t^M set by the central bank. Risky bonds deliver a stochastic real rate of return r_{t+1}^L. We describe a bank's portfolio by its share of reserves α_t^M in total assets as well as the share of other short safe bonds α_t^B in assets. The real rate of return r_{t+1}^α on the bank's asset portfolio is a weighted average of the returns on reserves, safe bonds, and risky bonds. We also define bank leverage at date t as the ratio of promised deposit payoffs to assets

$$\ell_t = \frac{D_t(1+i_t^D)}{A_t}. \tag{3}$$

All ingredients of the leverage ratio are known as of date t, so ℓ_t is part of the description of bank policy at date t.

Banks' technology is described by two cost functions. First, we introduce a cost of delegated portfolio management. The idea is that agency problems always entail costs, but that those are compounded when the value of assets falls short of the promised payoff on debt. We thus assume that, for each dollar of assets acquired at date t, the bank incurs a *balance sheet cost* of $k(\tilde{\ell}_{t+1})$ dollars at date $t+1$, where $\tilde{\ell}_{t+1}$ is an *ex post* measure of leverage, namely the ratio of deposits to the *stochastic* payoff on assets at $t+1$:

$$\tilde{\ell}_{t+1} = \frac{\ell_t}{(1+r_{t+1}^\alpha)P_{t+1}/P_t}. \tag{4}$$

For given leverage chosen at date t, ex post leverage is high if the nominal return on assets in the denominator is low – a shortfall of assets relative to promised debt.

The function k is strictly increasing and convex in $\tilde{\ell}$.[8] It starts at $1 > k(0) > 0$: even an all equity-financed bank incurs an operating cost. Leverage then raises costs at an increasing rate and a bank without equity is not viable. Convexity of the cost function thus effectively makes the bank more averse to risk than what would be implied by shareholders' pricing kernel M_{t+1}

[7] In practice, money-market mutual funds keep their assets at custodian banks and rely on the latter's access to Fedwire and other payment systems for their payment services. For an aggregate approach that distinguishes only between payment intermediaries and an investor sector, it thus makes sense to consolidate.

[8] We impose no condition here to ensure that $\tilde{\ell}$ is below one so that bank equity is positive. Nevertheless, we focus throughout on interior solutions with that property. In our quantitative application, we specify a cost function that slopes up sufficiently quickly for banks to choose leverage below one, as in the data.

alone. This type of cost can be microfounded by a setup with bankruptcy costs: suppose, for example, banks incur a deadweight cost – a share of assets is lost in reorganization – whenever the return on assets falls below a multiple of debt.

Our second cost function captures the idea that reserves are liquid instruments that help banks meet liquidity shocks. Banks face such shocks because their debt is inside money used for payments. We assume that, for each dollar of deposit issued at date t, the bank incurs a *liquidity cost* of $f(m_t)$ dollars, where m_t is the ratio of reserves to average depositors' transactions

$$m_t := \frac{\alpha_t^M A_t}{\zeta_t D_t}. \tag{5}$$

The average propensity to use deposits for payments ζ_t is known to the bank at date t. The function f is convex, and converges to zero as m_t becomes large. Its slope $f'(m_t)$ is negative for low values of m_t, as more reserves help banks to manage liquidity. For large value of m_t, the slope approaches a positive number $\rho_t > 0$, which may vary over time. This number captures that reserves are an asset that cannot be used as collateral in repo transactions, and therefore do not provide as much liquidity as other assets, such as safe short bonds, which can be used in repos. The presence of liquidity costs is not essential for the short rate disconnect to obtain. They are useful, however, to contrast the scarcity of short safe debt that gives rise to the short rate disconnect in our model to the scarcity of reserves that ended with quantitative easing programs.

At date t, a bank acquires A_t dollars worth of assets and issues D_t dollars worth of deposits; shareholders' equity is $A_t - D_t$. It chooses nonnegative assets, deposits as well as nonnegative balance-sheet ratios α_t^M, α_t^B and ℓ_t with $\alpha_t^M + \alpha_t^B \leq 1$ in order to maximize the discounted value of payoffs

$$\begin{aligned} &\left(E_t\left[M_{t+1}^{\$}\left(1-k\left(\tilde{\ell}_{t+1}\right)\right)\left(1+r_{t+1}^{\alpha}\right)\left(P_{t+1}/P_t\right)\right]-1\right)A_t \\ &\qquad +\left(1-E_t\left[M_{t+1}^{\$}(1+i_t^D)\right]-\zeta_t f(m_t)\right)D_t. \end{aligned} \tag{6}$$

Here the portfolio weights α_t^M and α_t^B enter into the return on assets r_{t+1}^{α} and together with leverage determine m_t and ex post leverage $\tilde{\ell}_{t+1}$ according to equation (4). The first term is then the return on assets net of balance-sheet costs and the second term is the interest payment on deposits plus liquidity costs. The bank's objective is homogeneous of degree one in its asset and liability positions; optimal policy determines only balance-sheet ratios.

3.3 Bank optimization and bank Euler equations

Shareholder value maximization means that the bank compares returns on potential assets and liabilities to its cost of capital. In a setup with risk, the cost of capital is state-dependent and captured by shareholders' pricing kernel M_{t+1}. For each asset and liability position, the bank thus computes the risk-adjusted return. At an optimum, the risk-adjusted gross return on each asset position has to be less than or equal to one – otherwise the bank could issue an infinite amount of equity in order to buy the asset. If the risk-adjusted return is strictly below one, the bank holds zero units of the asset; while it would like to go short, it is not allowed to do so. The risk-adjusted return thus has to be equal to one for all assets that the bank holds in equilibrium. Analogously, the risk-adjusted return on deposits has to be larger than or equal to one – otherwise the bank would issue an infinite amount of deposits. Banks issue deposits if their risk-adjusted return is equal to one.

A key feature of our model is that the balance-sheet cost affects risk-adjusted returns. To see this, consider for example the first-order condition for assets A_t. Taking the derivative of shareholder value, we have that the risk-adjusted overall return on bank assets must be equal to one:

$$E_t \left[M_{t+1} \left(1 - k\left(\tilde{\ell}_{t+1}\right) + k'\left(\tilde{\ell}_{t+1}\right)\tilde{\ell}_{t+1}\right)\left(1 + r^{\alpha}_{t+1}\right)\right] - \alpha^M_t f'\left(m_t\right) = 1. \tag{7}$$

The balance-sheet cost enters in two ways. First, it proportionally lowers the return on assets – this is true even if leverage is zero. Second, an additional dollar of realized return has a *marginal collateral benefit* $k'\left(\tilde{\ell}_{t+1}\right)\tilde{\ell}_{t+1}$: it lowers ex post leverage and hence the balance-sheet cost. In other words, backing deposits with assets makes deposit production cheaper.

Since all individual assets incur balance-sheet costs and contribute collateral, the cost k enters all bank optimality conditions. To concisely write those conditions, we define the *bank pricing kernel*

$$M^B_{t+1} = M_{t+1}\left(1 - k\left(\tilde{\ell}_{t+1}\right) + k'\left(\tilde{\ell}_{t+1}\right)\tilde{\ell}_{t+1}\right). \tag{8}$$

In other words, banks value payoffs more in states of the world in which their ex-post leverage is higher. Intuitively, this random variable describes how bank shareholders value contingent claims held inside the bank. There are two differences to the pricing kernel M_{t+1}: the proportional balance-sheet cost is subtracted, whereas the marginal collateral benefit is added.

The bank pricing kernel clarifies what states of the world are "bad" for the bank (that is, high M^B_{t+1}), and hence what assets represent bad risks for the purposes of bank portfolio choice. Since the bank owes short nominal debt, it is entirely safe if and only if it is "narrow", that is, it holds only short nominal bonds or reserves. In this case, the leverage ratio $\tilde{\ell}_{t+1}$ as defined in equation (4) is constant across states at $t+1$. Indeed, for a narrow bank, the nominal return on bank assets in the denominator is a weighted sum of predetermined nominal interest rates. Short nominal debt is thus good collateral for the bank in the sense that it does not worsen its risk profile. More generally, states are even worse for the bank than for shareholders if the return on bank assets is low. Since the balance-sheet cost function is convex, the bank pricing kernel is higher in states with higher ex post leverage, which is when nominal returns are lower.

Using the real bank pricing kernel together with its nominal counterpart $M_{t+1}^{B,\$}$, we rearrange the bank first order conditions with respect to A_t, α_t^M and α_t^B to derive a set of "bank Euler equations". For each of the three available assets – risky bonds, safe short bonds and reserves – the Euler equation says that the risk-adjusted expected return should be less or equal to one, with equality if the bank indeed holds the asset:

$$E_t\left[M_{t+1}^B(1+r_{t+1}^L)\right] \leq 1, \tag{9}$$

$$E_t\left[M_{t+1}^{B,\$}\right](1+i_t^B) \leq 1, \tag{10}$$

$$E_t\left[M_{t+1}^{B,\$}\right](1+i_t^M) = 1+f'(m_t). \tag{11}$$

The bank Euler equation for reserves must hold with equality in any equilibrium since only banks can hold reserves. Reserves differ from short safe bonds because of their marginal liquidity benefit $-f'(m_t)$. As a result, banks may wish to hold both in equilibrium: if the bank Euler equation for bonds holds with equality, then

$$\frac{i_t^B - i_t^M}{1+i_t^B} = -f'(m_t), \tag{12}$$

that is, the liquidity benefit is equated to the discounted spread between the bond rate and the reserve rate. An increase in the quantity of reserves in times when banks have a low liquidity ratio m_t reduces liquidity costs, and the spread $i_t^B - i_t^M$ shrinks.[9] When banks have ample reserves, as they had in recent years after the QEs, the slope $f'(m_t)$ approaches a positive number ρ_t, and the spread becomes negative. The negative spread reflects the fact that reserves cannot be used in repo transactions, while safe short bonds can be used in repos.

[9]Piazzesi and Schneider (2018) present a model in which a counterpart of f is derived from banks' liquidity shock distribution. Their formulation implies a threshold for the ratio m_t beyond which f remains constant so that the spread is literally zero. They use this setup to distinguish the ample reserve regime after 2008 with the scarce reserve regime prevalent before the financial crisis. In the present paper the focus is not on reserve management so this distinction is not critical.

Finally, consider the bank's first order condition with respect to deposits:

$$\frac{i_t^S - i_t^D}{1 + i_t^S} = E_t \left[M_{t+1}^{\$} \, k' \left(\tilde{\ell}_{t+1} \right) \, (1 + i_t^D) \right] + \zeta_t \, f(m_t) - \zeta_t \, f'(m_t) \, m_t. \tag{13}$$

The left hand side is the opportunity cost of deposits to the rest of the economy, or the value of the liquidity provided by deposits. The right hand side is the marginal cost of producing an additional unit of deposits. It consists of a marginal balance-sheet cost as well as marginal liquidity cost. Competitive banks thus equate the price of inside money to its marginal cost.

The presence of the balance-sheet and liquidity cost functions together with the liquidity benefit of deposits for households implies that our model has determinate interior solutions for leverage and portfolio weights. The choice of leverage works much like in the trade-off theory of capital structure. On the one hand, deposits are a cheap source of funds for banks, since their interest rate is below the short rate in the nominal pricing kernel. On the other hand, issuing debt incurs costs. An interior optimal leverage trades off the two forces. Moreover, portfolio choice is determinate because it affects portfolio risk and hence expected cost.[10]

3.4 Conduct of monetary and macro-prudential policy

The central bank conducts conventional monetary policy by setting the nominal rate i_t^B. To understand how the price level is determined, we consider first the case with ample reserves. In this case, the two Euler equations for risky assets and short safe bonds, (9) and (10), determine leverage ℓ_t and the safe portfolio share, which is the share of the portfolio invested in reserves and safe short bonds. Banks' optimality condition for deposits then sets the deposit spread, which determines the investors' demand for real balances. Given a nominal quantity of bonds and reserves, the price level adjusts so that the real quantity of reserves and short bonds allows to produce the desired quantity of real balances.

In the presence of liquidity benefits, the quantity of reserves relative to bonds additionally pins down the spread between the interest rate on reserves i_t^M and the policy rate i_t^B. If the central bank wants to set both rates, it therefore needs to adjust the nominal quantity of reserves relative to bonds. A large expansion in the aggregate quantity of reserves is associated with a reduction in the spread between the nominal short rate and interest on reserves, but may have no effect on the price level as long as the shadow spread falls to induce banks to hold more collateral. Our model can therefore speak to the regime change in US monetary policy around 2008, moving from a zero interest on reserves environment to the current setting with ample reserves.

[10]The four equations in (9)-(11) and (13) jointly restrict the three bank balance-sheet ratios α_t^M, α_t^B and ℓ_t. An equilibrium in which the bank holds all assets thus requires that interest rates align to allow a solution.

The model also allows us to think about the conduct of macro-prudential policy, which we can model as changes in the balance-sheet cost function. Stricter capital requirements correspond to an increase in the general cost of leverage, while the curvature of the cost function governs banks' risk-sensitivity and can capture changes in risk weights. To model larger operating cost associated with, for example, additional reporting requirements of asset holdings, we can uniformly increase balance-sheet cost independent of the leverage choice. Our quantitative exercise will suggest that variations in the latter cost component capture the dynamics of regulatory changes well, in particular after 2008.

3.5 The short rate disconnect in equilibrium

We focus on equilibria such that the risky bond is priced by the pricing kernel of investors. This might be because investors can go both long and short in the risky bond, or alternatively because the outstanding quantity of risky bonds is so large that not only banks hold risky bonds but also investors hold them directly. It follows that banks also hold risky bonds, then their pricing kernel must similarly price them. Since the bank pricing kernel is generally different from that of shareholders, the balance-sheet ratios of banks must respond appropriately.

To clarify the relationship between the scarcity of short safe assets and the short rate disconnect, we characterize equilibria in which banks hold a fixed supply of reserves A_t^M as well as nominal short bonds A_t^B. We think of these quantities as being endogenously determined in general equilibrium by the interaction of government policy and the demand from other intermediaries who hold short bonds. Our focus here is on the relationship between quantities and prices implied by banking sector optimization and partial equilibrium in the reserve and deposit markets.

At given prices, all banks in our model choose the same leverage and portfolios. We define a symmetric *equilibrium* as a tuple $\left(\ell_t, \alpha_t^M, \alpha_t^B, P_t, D_t, i_t^D, i_t^B\right)$ that solves the bank first-order conditions (9)-(11) and (13), the money demand equation (2) as well as the market clearing conditions

$$D_t \ell_t^{-1} \left(1 + i_t^D\right) \alpha_t^M = A_t^M, \tag{14}$$

$$D_t \ell_t^{-1} \left(1 + i_t^D\right) \alpha_t^B = A_t^B. \tag{15}$$

In principle, there could be two types of equilibria. In a *narrow banking equilibrium*, banks do not invest in risky bonds, $\alpha_t^M + \alpha_t^B = 1$. With narrow banking, there is no short rate disconnect. Indeed, from (8), the pricing kernel of a narrow bank is proportional to that of investors, with factor $(1+i_t^S)/\left(1+i_t^B\right)$.[11] Since investors price the risky bond, a positive shadow spread would violate the bank condition (9): if a narrow bank were to earn less than the short rate, then it always makes sense to take a little risk. This is a version of Arrow's "local risk neutrality" result, here applied to the case of banks. In a narrow banking equilibrium, we have $i_t^S = i_t^B$, and the bank pricing kernel is then the same as that of investors.

When can a narrow banking equilibrium exist? There must be a large enough (real) quantity of short safe collateral that can back inside money demanded at the cost implied by narrow banking. From (9) and the fact that there is no disconnect, the optimal leverage ratio with narrow banking depends only on the balance-sheet cost:

$$k\left(\ell^*\right) = k'\left(\ell^*\right)\ell^*. \tag{16}$$

Leverage adjusts so that the balance-sheet cost is exactly offset by the marginal collateral benefit from short safe bonds. This leverage ratio together with the optimal reserve share from (11) implies a deposit rate by (13) and an equilibrium real quantity of deposits by (2). Market clearing for reserves thus requires a large enough real quantity of short safe collateral.

In a *risky banking equilibrium*, banks buy risky bonds, so (9) holds with equality. Such an equilibrium is consistent with a quantity of inside money that is large relative to the quantity of short safe collateral. Importantly, however, a risky banking equilibrium does not require that the short rate i_t^B equals the shadow rate i_t^S. To see this, we use the definition of the bank pricing kernel to rearrange the Euler equation for bonds as

$$\frac{1}{1+i_t^B} = \frac{1}{1+i_t^S} + E_t\left[M_{t+1}\left(-k\left(\tilde{\ell}_{t+1}\right) + k'\left(\tilde{\ell}_{t+1}\right)\tilde{\ell}_{t+1}\right)\right]. \tag{17}$$

In general, there is a spread between the short rate and the shadow rate given by the risk-adjusted difference between the marginal collateral benefit and the asset cost.

To sum up, if the bank is narrow, that is, it holds no risky bonds, then ex post leverage $\tilde{\ell}_{t+1}$ is predetermined and the spread is zero. In other words, in an economy with narrow banks, there is no short rate disconnect. More generally, however, for a risky bank the balance-sheet cost induces a wedge between the two interest rates. In the next section, we use a particular functional form for the cost function to work out its empirical implications.

[11] Ex post leverage for a narrow bank is not random, which means that

$$E_t\left[M_{t+1}^{B,\$}\right]\left(1+i_t^B\right) = \left(1-k\left(\tilde{\ell}_{t+1}\right)+k'\left(\tilde{\ell}_{t+1}\right)\tilde{\ell}_{t+1}\right)E_t\left[M_{t+1}^{\$}\right]\left(1+i_t^B\right) = 1.$$

Since $\left(1+i_t^S\right) = 1/E_t\left[M_{t+1}^{\$}\right]$, we get $M_{t+1}^{B,\$}/M_{t+1}^{\$} = (1+i_t^S)/(1+i_t^B)$.

4 Quantitative evaluation

In this section we connect the model to the data. Section 4.1 provides evidence on a key assumption of the model, that short safe bonds are not held directly by households, but are held through intermediaries, in particular payment intermediaries. Section 4.2 provides measures of bank balance-sheet ratios and uses them to test the bank Euler equations under the assumption that the regulatory environment remains constant. Finally, Section 4.3 extends the model to allow for changes in regulation that shift the balance-sheet costs of banks.

4.1 Who holds short safe bonds?

Our theory is based on the idea that payment intermediaries value short safe bonds as collateral. We now consider evidence on asset positions that support this view. The ideal data to make our point would be sectoral accounts that track Treasury bills by maturity. Unfortunately, such data is not available for the US financial system. The available data do, however, allow several conclusions.

We first note that households do not buy T-bills from the government in the primary market. There has been a recent effort to sell T-bills directly to the public via the TreasuryDirect website. We can rule out, however, that the public purchases a sizeable share of T-bills through this channel. Indeed, for the period between 2008 and 2016, on average only 1.1% of all T-bills sales went through TreasuryDirect directly to households, and only 1.6% was sold non-competitively in total. The remaining T-bills were sold in a competitive auction process to primary dealers and other financial institutions.

Our second source of information about T-bill holdings are data from the Financial Accounts of the United States. Unfortunately, we observe a breakdown of the overall instrument "Treasuries" into short-term bills and long-term notes and bonds only for a subset of sectors: money-market funds, insurance companies, mutual funds (since 2010), the monetary authority, and the rest of the world. While we do not see a breakdown for nonfinancial corporations, it makes sense to assume that their Treasury holdings consist of T-bills held for liquidity purposes.

Figure 3 uses this information to take a first stab at the composition of T-bill holdings by the domestic private sector. The total here is outstanding Treasury bills less holdings by the monetary authority and the rest of the world. The top shaded area in the figure consists of T-bills held by "Others" – the remaining T-bills outstanding that are not accounted for by holdings of specific sectors. This category in particular contains holdings of commercial banks, as well as those of households or institutions lumped in by the Financial Accounts with the (residual) household sector, in particular hedge funds.

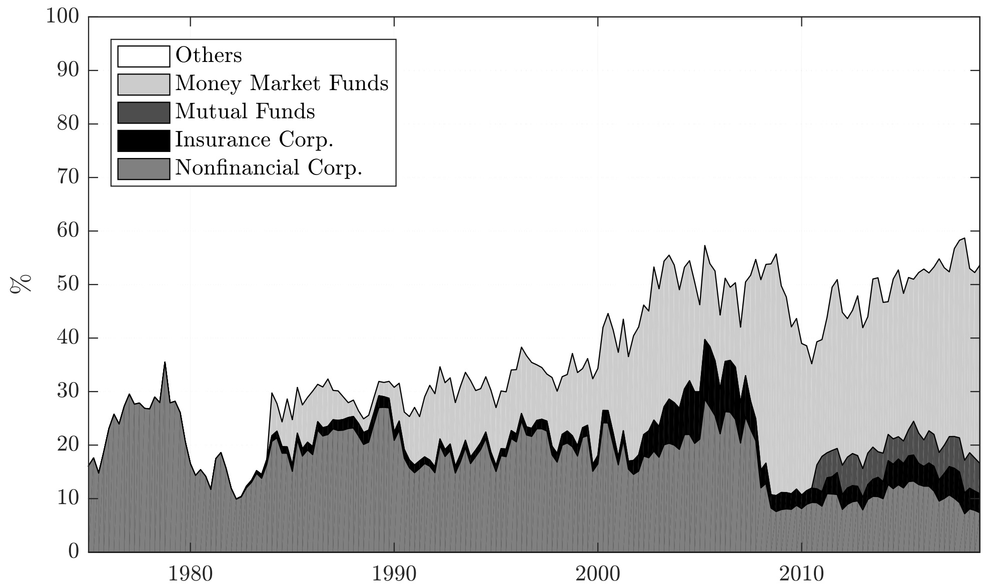

Figure 3: Holdings of T-Bills by Money-Market Funds, Mutual Funds, Insurance Corporations, Nonfinancial Corporations and Others. Quarterly data from the Flow of Funds.

Our theory suggests that a large chunk of the "Other" category of T-bill holdings should consist of holdings of commercial banks. To assess this possibility, Figure 4 compares the time series of T-bill holdings by "Others" and money-market funds (in red) with all the Treasury holdings of payment intermediaries (in blue). Here payment intermediaries include depository institutions, credit unions, and banks. Importantly, "Treasuries" now include bonds, not only bills. The two lines should coincide if (i) no other domestic sector (except those in Figure 3) holds bills and (ii) all Treasuries held by banks are bills (as opposed to bonds or notes).

Several facts emerge from Figure 4. First, payment intermediaries' holdings are typically higher than "Other" bill holdings not yet accounted for, consistent with all bills being held by payment intermediaries. The exception is the period around the recent boom and bust, where one might expect more participation of broker-dealers and hedge funds in the bill market. Second, the cyclical movements in bill holdings is closely aligned with payment intermediaries holdings, again with the exception of the recent boom-bust episode. Together we view these patterns as supportive of an approach that treats bills as held by payment intermediaries (as well as possibly other intermediaries), and not directly by investors.

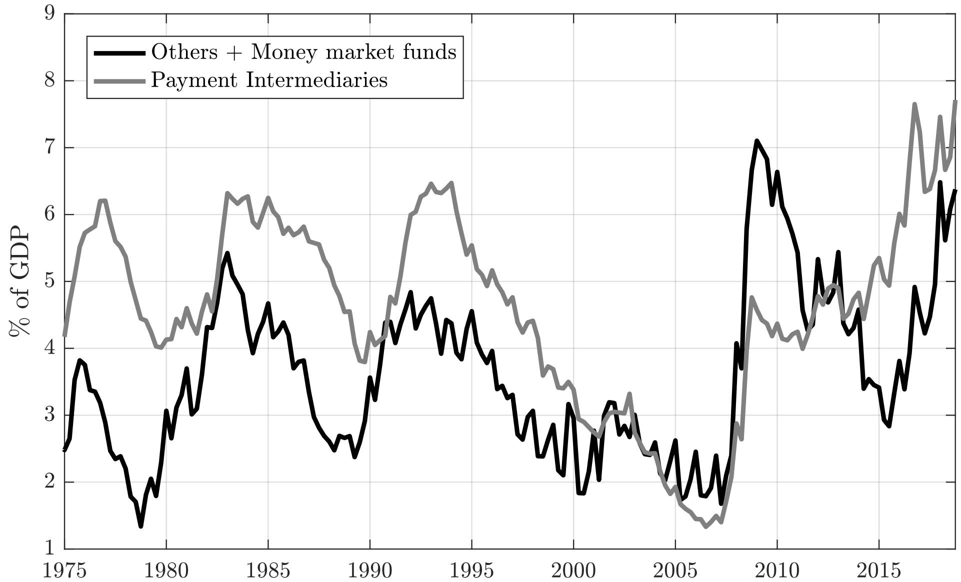

Figure 4: T-Bills held by Others and Money-Market Funds, together with all Treasury holdings by Payment Intermediaries.

4.2 Bank Euler equations

In order to obtain transparent versions of the bank Euler equations that can be taken to the data, we make a number of simplifying assumptions. We work with a balance-sheet cost function of the form

$$k(\tilde{\ell}_{t+1}) = b\left(\bar{k} + \tilde{\ell}_{t+1}^{\gamma}\right), \tag{18}$$

where b and $\bar{k}$ are strictly positive and $\gamma > 1$. The balance-sheet cost thus consists of an operating cost (per dollar of assets) $b\bar{k}$ plus a power function that captures the sensitivity to leverage. The parameter b scales the overall cost, while an increase in $\bar{k}$ regulates the *relative* importance of the operating cost. The parameter γ governs the curvature of the cost function.

The pricing kernel of the bank now becomes

$$M_{t+1}^{B,\$} = M_{t+1}^{\$}\left(1 - b\left(\bar{k} + (1-\gamma)\tilde{\ell}_{t+1}^{\gamma}\right)\right). \tag{19}$$

Since $\gamma > 1$, the cost function is convex and the pricing kernel is increasing in ex-post leverage $\tilde{\ell}_{t+1}$. A bad state of the world for the bank (high pricing kernel) occurs when the return on its asset portfolio is low and ex-post leverage is high. As a result, a bank places a higher value on assets that pay off more in those states of the world.

We introduce additional notation that helps decompose ex-post leverage $\tilde{\ell}_{t+1}$ into the bank's leverage chosen at date t and its stochastic (nominal) return on assets:

$$\begin{aligned} \ell_t &:= (1+i_t^D)D_t/A_t, && (20) \\ 1+r_{t+1}^{\alpha,\$} &:= (1+r_{t+1}^{\alpha})P_{t+1}/P_t. && (21) \end{aligned}$$

The leverage ratio ℓ_t measures the *promised* payment on deposits (principal plus interest) as a fraction of assets. It is known to the bank as of date t. Ex-post leverage can then be written as the ratio $\tilde{\ell}_{t+1} = \ell_t/(1+r_{t+1}^{\alpha,\$})$, thus separating the leverage decision from asset choice and the realization of returns.

We follow Campbell and Viceira (1999) in exploiting conditional lognormality to approximate optimal portfolio choice. In our case, the bank takes into account shareholders' valuation via a pricing kernel, so we assume joint conditional lognormality of the gross return on the risky bond and the pricing kernel. We denote by σ_t the conditional volatility of the risky bond return given date t information. This distributional assumption is motivated by continuous-time setups with Brownian motion; the Campbell-Viceira approach approximates the elegant solutions that obtain in such setups. The key approximation step is a second-order Taylor approximation of the bank's portfolio return around the riskfree return. We then obtain the following characterization of bank balance-sheet ratios, derived formally in the Online Appendix:

Proposition 1 *The bank's optimal portfolio share of safe assets is*

$$\alpha_t \approx 1 - \frac{1}{\gamma\sigma_t^2}\log\left(1+\frac{i_t^S - i_t^B}{b\bar{k}}\right), \tag{22}$$

Optimal bank leverage is

$$\ell_t \approx \exp\left(i_t^{\alpha} + 0.5(1-\alpha_t^2)\sigma_t^2\right)\exp\left(-\frac{1}{2}\gamma\sigma_t^2(1-\alpha_t)^2\right)\ell^*, \tag{23}$$

where $\ell^* = \left(\bar{k}/(\gamma-1)\right)^{1/\gamma}$, $\alpha_t = \alpha_t^B + \alpha_t^M$ *is the safe portfolio share, which combines the portfolio shares invested in safe short bonds and reserves, and* $i_t^{\alpha} = (1-\alpha_t)i_t^S + \alpha_t^B i_t^B + \alpha_t^M i_t^M$.

The approximate formulas clarify the trade-offs faced by a bank in our model and how balance sheets respond to the environment. Consider first the safety of the asset portfolio. If there is no short rate disconnect, then the optimal bank portfolio consists only of safe assets: we obtain a "narrow" bank with $\alpha_t = 1$. Since the bank makes money from issuing short safe nominal deposits, and the convex balance-sheet cost penalizes risk, it makes sense for banks to avoid any risk. Only if risk avoidance is costly – because of a positive shadow spread – does it become optimal to back inside money with risky collateral.

For a risky bank, the safe portfolio share is decreasing in the shadow spread and increasing in risk as well as the fixed component and the curvature of the balance-sheet cost. To draw a connection to standard portfolio choice theory, we can rearrange the formula to resemble that for optimal myopic portfolio choice with a risky and riskfree asset:

$$1-\alpha_t \approx \frac{-\log(b\bar{k}) - \left(-\log\left(b\bar{k} - \left(i_t^S - i_t^B\right)\right)\right)}{\gamma\sigma_t^2}. \tag{24}$$

Here the denominator is the product of risk and curvature in the objective; here γ works like risk aversion for a power utility investor. The numerator is an expected (risk-adjusted) excess return to shareholders of placing a risky versus a riskfree asset in the bank. In both cases, they incur the balance-sheet cost $b\bar{k}$, and for the riskfree asset they further incur the shadow spread. The expected return on the risky bond does not matter because shareholders can also hold it directly – this is why shareholders compare risk-adjusted and not raw returns.

The formula for optimal leverage has three components. The constant ℓ^* determines leverage (up to an interest factor) if the bank is safe (that is, $\alpha_t = 1$). It follows from the properties of the balance-sheet cost function alone. For a risky bank, two things change. First, the risk-adjusted expected return on the portfolio increases since the bank avoids the shadow spread and the mean return increases by Jensen's inequality. This effect – captured by the first exponent – tends to make leverage increase with the safe portfolio share. At the same time, however, a riskier bank incurs a higher certainty equivalent leverage cost: the second exponent says that riskier banks should reduce leverage.

Quantitatively, a key force in our model is that higher bond risk *increases* leverage. The formulas show why this effect is powerful: an increase in bond return variance σ_t^2 leads banks to optimally reduce the share $1-\alpha_t$ of risky bonds in proportion with variance. As a result, total portfolio risk $\sigma_t^2(1-\alpha_t)^2$ declines – the risk reduction implied by optimal portfolio choice always outweighs the increase in bond risk. A safer bank then optimally increases leverage, with the strength of the effect driven by the curvature of the cost function. At the same time, the potentially offsetting force that works through the mean return tends to be quantitatively small, as moderate percentage point movements in α_t meet the modest shadow spread.

The formulas also clarify the role of the shadow spread for bank leverage and portfolio choice. A higher shadow spread makes safe banking more costly and induces more risk taking. For leverage, there are again two forces: small changes in the mean return and an incentive to lower leverage in the face of higher portfolio risk. This second force is key for our account of the data below: as the shadow spread increases towards the end of booms, banks choose riskier portfolios and reduce leverage. We note that this effect would obtain even if risk were constant. Our results below suggest that risk and spreads move together in equilibrium.

Data on balance-sheet ratios To form the balance-sheet ratios ℓ_t and α_t, we need data counterparts for three bank positions in the model: deposits, short safe bonds, and total assets. The mapping between model and data needs to take into account that the model has one type of payment intermediary, while in the data both banks and money-market mutual funds are providers of inside money. At the same time, some payment intermediaries in the data issue claims that cannot be identified with inside money, for example repo borrowing by commercial banks.

Our theory makes predictions about aggregate balance-sheet ratios, in particular how much collateral backs outstanding inside money. We thus construct an aggregate payment intermediary sector by consolidating banks and money-market funds. In practice, money-market fund companies keep their portfolios at custodian banks with whom they also contract for payment services that they sell to shareholders – for example, money-market fund companies do not directly participate in Fedwire, the key gross settlement system used for interbank payments. Our consolidation thus treats these contracts as occurring within one large firm.

Formally, the data counterpart of inside money D_t is Money of Zero Maturity (MZM), provided by the Federal Reserve Bank of St. Louis. The MZM series is a broad measure of money that incorporates those types of deposits and money-market fund shares that are sufficiently liquid to provide immediate payment services. An advantage of this series is its stable money-demand relationship to interest rates, as documented by Teles and Zhou (2005).

Our measure of total assets held by payment intermediaries is derived from the U.S. Financial Accounts (Z.1), where we aggregate depository institutions (Table L.110) and money-market funds (Table L.121). We emphasize that we work at the level of individual institutions, not bank holding companies. We thus treat any wholesale funding of bank holding companies as occurring within the investor sector. We view this approach as appropriate since our theory is about banks producing inside money, whereas bank holding companies also own intermediaries with very different business models, in particular broker dealers.

To further address wholesale funding at the commercial bank level, we subtract repurchase agreements which can be viewed as senior to deposits because they are tied to specific collateral. In other words, our measure of total assets contains only the haircut on repo collateral, not the total value of the securities. This approach takes care of the most important funding source that is not explicitly in our model. We treat commercial paper the same way – while seniority here is not so clear, the adjustment is also relatively small.

As the data counterpart of leverage ℓ_t, we calculate the ratio of MZM and aggregate payment intermediary asset holdings. We need to multiply this ratio by the deposit interest rate, since we have defined leverage ℓ_t as the ratio of promised repayment in the next period relative to current asset holdings. We use the MZM own rate provided by the Federal Reserve Bank of St. Louis – it reflects a weighted average of rates on the different flavors of inside money that make up MZM.

To sum up, our measure of leverage differs from other statistics of bank leverage discussed in the literature in three ways. First, within the banking sector, we only consider depository institutions for our calculations, not a broader set of banks such as brokers and dealers. Second, we include money-market funds, which hold for our purposes highly leveraged but safe portfolios. Third, and most importantly, we only consider deposits in the numerator of our leverage measure, not a broader set of liabilities.

Our measure of short safe bonds aggregates the subset of those assets that are of short maturity and nominally safe. For depository institutions, we assume that vault cash, reserve, and Treasury holdings fall into this category. For money-market funds, we add holdings of Treasuries, municipal bonds, and government agency debt. To the sum of those two measures we also add the net-repo holdings of both sectors, consistent with having subtracted repo liabilities from the total asset measure. The fraction of those safe assets relative to total asset holdings yields our time series of α_t.

Stylized facts on bank balance sheets and the shadow spread As a first look at how the shadow spread evolves, we just plot the raw data. The top panel of Figure 5 plots the time series of the safe portfolio share α_t in black against the shadow spread in grey over the sample 1975 to 2018. Even in the raw data, one can detect the negative co-movement between the two time series. The same can be said about the time series of leverage ℓ_t which is depicted in the bottom panel of the same figure. Qualitatively, our model gives predictions that are consistent with the data, namely that episodes of high shadow spreads are associated with a lower safe asset share on banks' balance sheets and lower bank leverage. This co-movement is also present in the period after the financial crisis of 2008, which sees an increase in both the safe asset share and bank leverage.

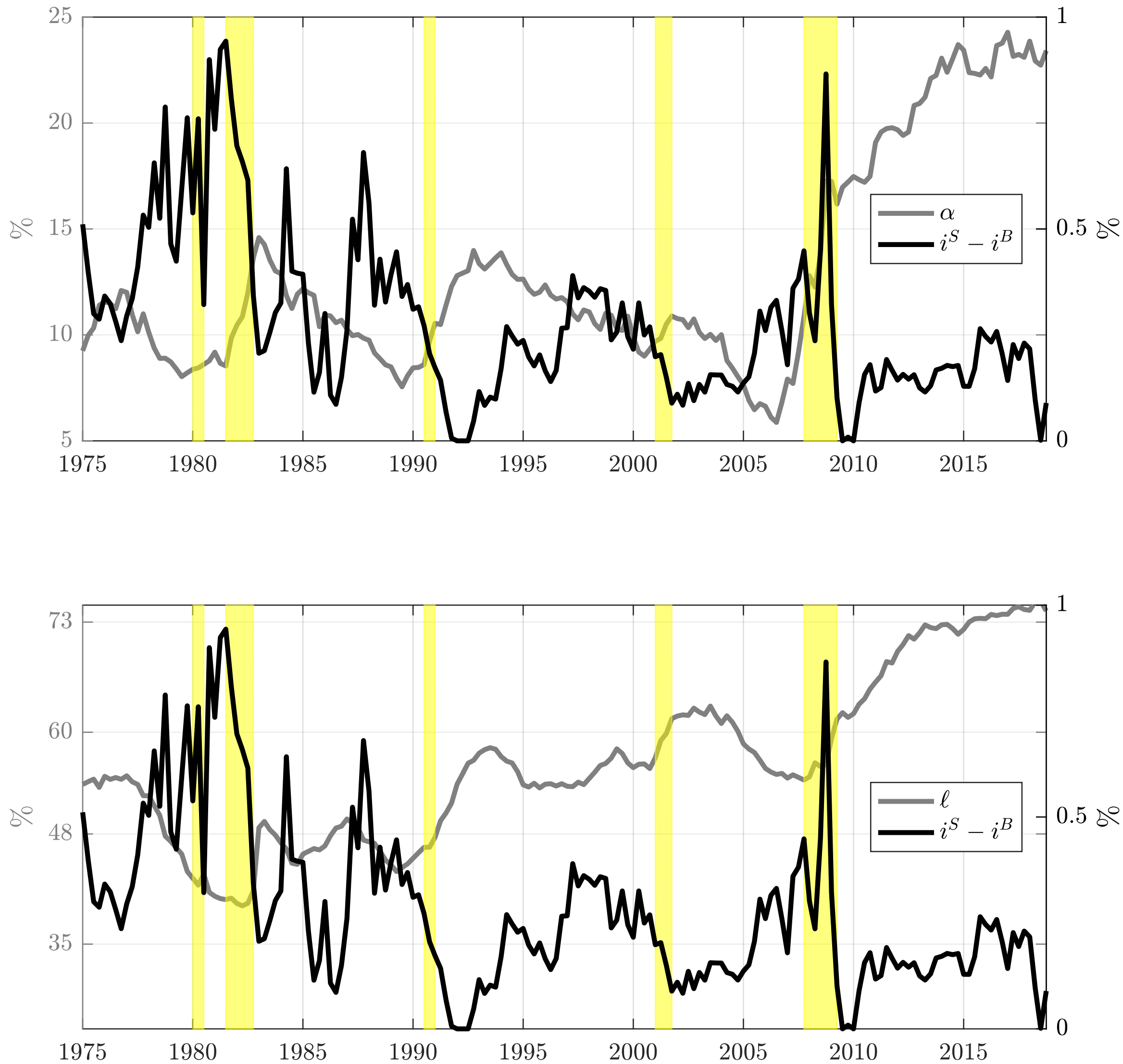

Figure 5: **Top panel:** Safe portfolio share (left axis) and shadow spread (right axis). **Bottom panel:** Leverage ℓ_t (left axis) and shadow spread (right axis). **Data:** i_t^B is the 3 month T-bill rate, i_t^S the shadow rate, data on leverage based on MZM (St. Louis Fed) and payment intermediary asset holdings measured from the U.S. Financial Accounts (Z.1).

The evolution of leverage after 2008 highlights the differences between our leverage measure and, for example, the asset to equity ratio. While capital regulation has forced banks to lower their liability to asset ratio since 2008, the same is not true for the deposit to asset ratio. This observation does not rely on our definition of payment intermediaries, but also holds for commercial banks alone, whose liability to asset ratio has decreased from about 90% before the crisis to about 89% after, but whose deposit to asset ratio has increased from about 64% in the years preceding the crisis to more than 71% in 2017.[12] From our model's perspective, which focuses on the amount of assets that are available to back deposits, the latter is the relevant statistic.

While the co-movements of these three time series are at least qualitatively consistent with the model's mechanisms, these figures do not allow us to evaluate the model's quantitative success. In the next section, we therefore study the empirical fit of the model, which will also allow us to back out the time series of return risk σ_t^2, that also affects the leverage and portfolio choice in the model.

Graphical assessment of bank Euler equations Section 4.2 derived two equations for the portfolio share and leverage in terms of the shadow-bond spread and the risky asset's return variance σ_t^2. While payoff risk is an unobserved latent factor, we can use the equation of the portfolio share to replace $\gamma\,\sigma_t^2$ in the leverage equation. We then find that

$$\ell_t = \exp\left(i_t^\alpha + 0.5(1-\alpha_t^2)\sigma_t^2\right)\exp\left(-\frac{1}{2}(1-\alpha_t)\log\left(1+\frac{i_t^S - i_t^B}{b\bar{k}}\right)\right)\ell^*. \tag{25}$$

The first component is approximately the bank's expected nominal portfolio return. The second component states that leverage is, holding the portfolio share fixed, decreasing in the shadow spread, and, holding the spread fixed, increasing in the safe asset share.

We make two assumptions so we can estimate equation (25) with data on α_t, ℓ_t, i_t^S, and i_t^B alone. First, we impose that $i_t^\alpha \approx i_t^S - \alpha(i_t^S - i_t^B)$. This approximation works well in our sample, because α_t^M is small before 2008, while after 2008 the difference between i_t^M and i_t^B is small. Second, we assume that the Jensen's inequality term in the mean return is negligible, that is, $\exp(0.5\left(1-\alpha_t^2\right)\sigma_t^2) \approx 1$. This is certainly true if the risky asset return is similar to that of common risky assets such as stock indices. For example, even with $\sigma_t = .1$ and $\alpha_t = 0$, we have $\exp\left(0.5\left(1-\alpha_t^2\right)\sigma_t^2\right) = 1.005$. Variation in the Jensen's inequality term is thus too small to affect the mean. An assumption on the scale of the latent variable σ_t does not otherwise restrict the estimation: apart from entering the Jensen's inequality term, σ_t appears only through the product $\gamma\sigma_t^2$, which we back out from the data below.

We estimate the two parameters, $b\bar{k}$ and ℓ^*, by minimizing the sum of squared residuals of equation (25). The black line in the middle panel of Figure 6 depicts the time series of leverage predicted by the model. While the fit is far from perfect, we find that the model captures the

[12]Data: Board of Governors of the Federal Reserve System, H.8 Assets and Liabilities of Commercial Banks in the United States.

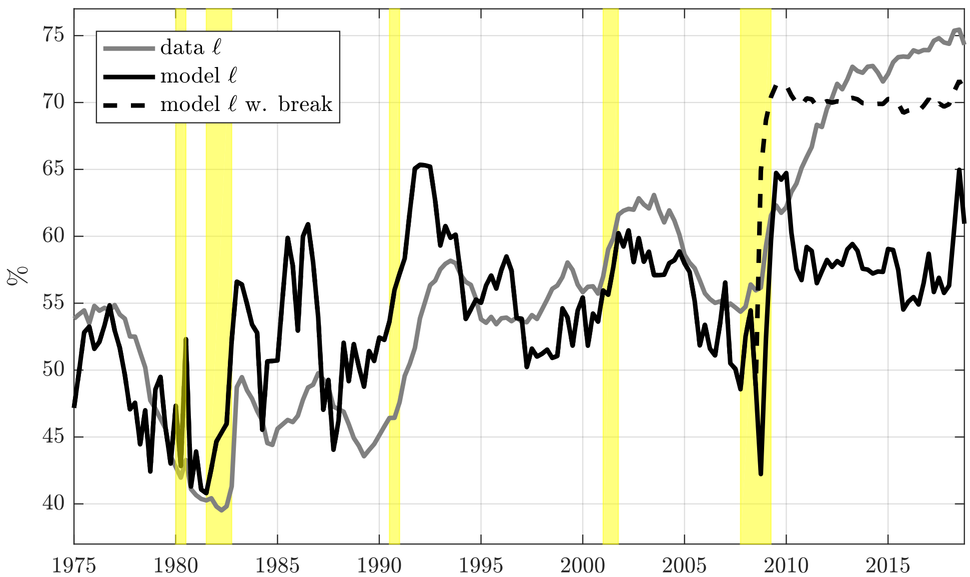

Figure 6: Leverage ℓ_t of payment intermediaries in the data (grey) and model predicted (black) as a function of $i_t^S - i_t^B$ given parameter estimates for $b\bar{k}$, ℓ^* and $\gamma\sigma_t^2$. The dashed line depicts the model fit with a re-estimated set of parameters post-2008. **Data:** i_t^B is the 3 month T-bill rate, i_t^S the shadow rate, data on leverage based on MZM (St. Louis Fed) and payment intermediary asset holdings measured from the U.S. Financial Accounts (Z.1), see text and appendix.

dynamics of leverage variation, at least up to the financial crisis in 2008. This can be seen even better when focusing on the cyclical component of leverage in both data in model. To do so, we use a bandpass filter on both the data and the series predicted by the model. The filter isolates business-cycle fluctuations that persist for periods between 1.5 and 8 years. The resulting cyclical components of the two series are shown in Figure 7. The correlation between the cyclical components of data and model is 72%. The estimated operating cost of banks, $b\bar{k}$, has an annualized value of 0.5%. Since our estimation does not restrict the range of this parameter, it is surprising to find that the estimated operating cost has indeed a sensible order of magnitude. The estimated level of optimal leverage ℓ^* for a safe bank is 67%.[13]

Without a structural change in parameters, the model is necessarily unable to fit the level of leverage after the crisis, since the level in the data post-2008 reaches close to 75%, but is, in the model, bounded by ℓ^*. The deviations in the trend components of the two series

[13]Since $\bar{\ell}^*$ denotes the deposit to asset ratio when banks only holds safe assets, this estimate implies that the available quantity of short safe bonds would need to be about 50% higher than the deposit demand to reach this equilibrium. Deposit demand would be larger at that point, because abundant collateral lowers the cost of deposit provision. We cannot quantify the equilibrium amount of deposits without specifying the functional forms of "money demand" $v(\cdot)$ and liquidity cost $f(\cdot)$.

could be driven by shifts in structural parameters of the banks' balance-sheet cost. Given the regulatory changes in the banking environment, this is plausible, and in the next section we explore which parameter changes can, for example after the financial crisis of 2008, explain the observed deviations of model and data.

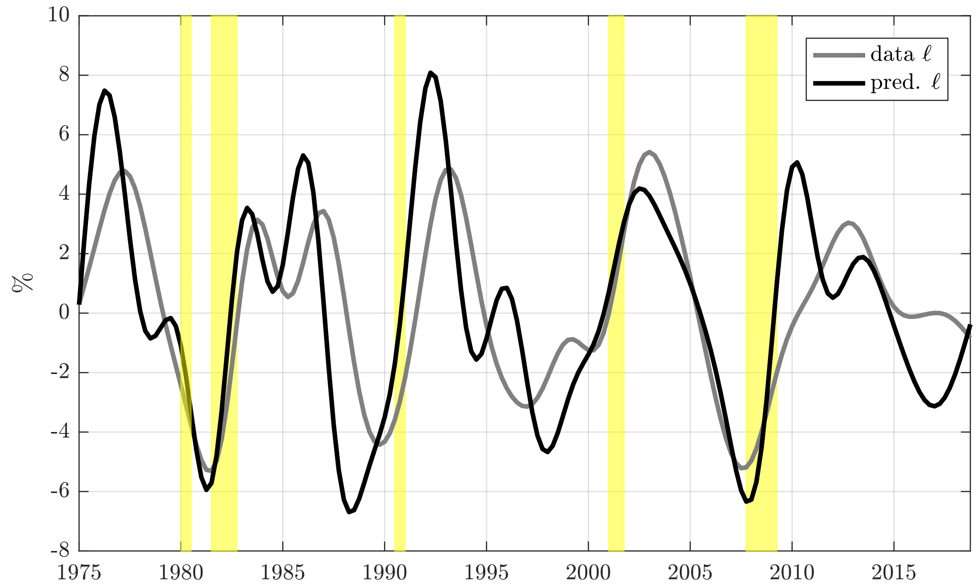

Figure 7: Bandpass filtered time series of data and model-predicted ℓ_t.

Return risk We can use our estimate for $b\bar{k}$ to find an implied measure of $\gamma\sigma_t^2$, which provides a scaled measure of return risk σ_t^2. Rearranging the equation for the portfolio share we have that

$$\gamma\sigma_t^2 = \frac{1}{1-\alpha_t} \log\left(1 + \frac{i_t^S - i_t^B}{b\bar{k}}\right). \tag{26}$$

Given the shadow spread and our parameter estimates, we find the estimated time series of $\gamma\sigma_t^2$ depicted in Figure 8. The series spikes in episodes of distress in financial markets, namely during the second oil price shock in 1979, the recession episodes and banking crisis of the early 1980s, the stock market crash in 1987, the 1994 peso crisis, the 1997/98 episode of financial turmoil associated with Asia, Russia and LTCM and finally in the years leading up to the financial crisis of 2007/08. As one would expect, this measure is correlated with the shadow spread, since in times of higher risk, safe collateral becomes more valuable and the shadow spread widens.

While we have no direct estimate for the curvature of the cost function γ, we can use a plausible level of return risk to back out a likely range for γ. Over the sample, $\gamma\sigma_t^2$ is on

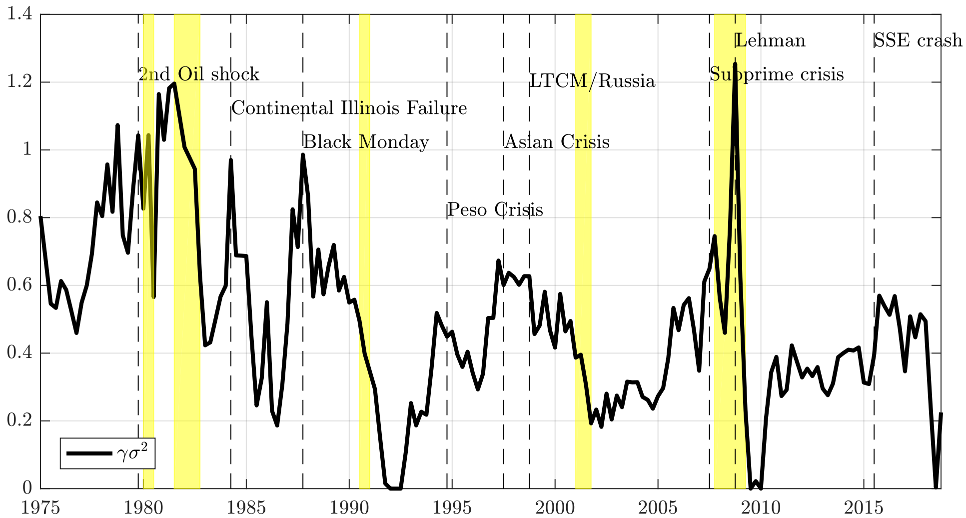

Figure 8: Estimated return variance of the risky asset scaled by the curvature of the balance-sheet cost function, $\gamma\sigma_t^2$, and episodes of financial distress.

average 0.47 and reaches up to 1.19. We use the quarterly return volatility of the S&P 500 stock index as a benchmark and find that the standard deviation of quarterly returns is on average 7.7% over the sample period and reaches up to 10.7% when calculated over 5-year rolling windows. If the bank faces similar return risk on the risky share of its portfolio, a value of $\gamma = 72$ matches the average volatility and implies a maximum volatility of 11.0% over 5-year rolling windows. In case the bank's risky assets had half the return volatility of the S&P 500, our estimated γ would be about 290, and we would find $\gamma = 18$ if the banks' risky return volatility would be twice that of the S&P 500 index. The higher is γ, the lower is ℓ^γ for small values of ℓ, but the steeper it increases as ℓ approaches 1, so that such high values of γ correspond to a more kinked cost function.

4.3 Structural changes in banks' balance-sheet cost

While our model is successful in capturing the cyclical components in the joint co-movement of safe asset share, leverage, and the shadow spread, the overall level of model implied leverage shows at times larger deviations from the data, in particular post-2008. A natural extension of our analysis is to allow for structural changes in the banks' balance-sheet cost function, which seems particularly relevant after the financial crisis of 2008, which was followed by a set of regulatory changes in the banking system. We therefore study to what extent a one-time change in parameters in the last quarter of 2008 can improve the model fit. We choose this quarter as a break point under the premise that the bankruptcy of Lehman Brothers triggered the following changes in the banking system. The dashed black line in Figure 6 depicts the fit of this re-estimated curve post-2008. Our results imply an increased operating cost of $b\bar{k} = 4.6\%$ as well as an increased level of $\ell^* = 70\%$.[14]

One unifying explanation of upward shifts in both, $b\bar{k}$ and $\ell^* = (\bar{k}/(\gamma-1))^{1/\gamma}$, is an increase of $\bar{k}$, which implies an increase in the operating cost of banking. This parameter change seems plausible because the regulatory changes imposed by the Dodd-Frank Act have been associated with increases in bank's balance-sheet cost. In the model, an increase in the operating cost leads to an increase in leverage, as it becomes more costly to hold assets in the bank as collateral. Assuming that b and γ remain fixed, our estimates for $b\bar{k}$ and $b\bar{k}_{post\ 2008}$ let us back out the curvature parameter as $\gamma = 44$. This estimate is in same order of magnitude as our estimates based on the return volatility targets presented in the previous section.

Smooth parameter shifts A one-time structural break in parameters gives us a first indication of the type of changes in bank's cost function that can improve the model fit. However, there have been several changes in bank regulation in the past decades, usually implemented in stages over time. Furthermore, other structural adjustment, for example through technological innovations, will presumably also induce slow moving adjustments in bank's balance-sheet cost. To allow for such slow moving changes in cost function parameters, we re-estimate equation (25), but now allowing for time-variation in its two parameters, namely the operating cost $\overline{bk}_t$, and the optimal level of leverage ℓ_t^* which banks choose when the shadow spread is zero. We are therefore interested in estimating

$$\ell_t = \exp(i_t^S + \alpha_t(i_t^B - i_t^S)) \exp\left(-\frac{1}{2}(1-\alpha_t)\log\left(1 + \frac{i_t^S - i_t^B}{\overline{bk}_t}\right)\right)\ell_t^* + \sigma_R \epsilon_t, \tag{27}$$

where ϵ_t is an independent, standard normally distributed measurement noise shock. We have to recover $\overline{bk}_t$ and ℓ_t^* as latent factors, and assume that both follow random walk processes:

$$\overline{bk}_t = \overline{bk}_{t-1} + \sigma_{bk}\,\eta_t^1, \tag{28}$$

$$\ell_t^* = \ell_{t-1}^* + \sigma_\ell\,\eta_t^2, \tag{29}$$

where η_t^1 and η_t^2 are independent shocks with a standard normal distribution.

Given the non-linear measurement equation, we use the Unscented Kalman filter to back out the time series of $\overline{bk}_t$ and ℓ_t^*. While it would be possible to jointly estimate the stochastic parameters σ_{bk}, σ_ℓ and σ_R and the time series of the latent factors using maximum likelihood

[14] Re-estimating the model pre-2008 has only small effects on our estimates for that period and for visual simplicity we therefore do not depict the updated time-series.

estimation, we find a calibration approach more sensible given the likely misspecification of our simple model. We choose to set σ_{bk} so that the annual standard deviation of the annualized operating cost $\overline{bk}_t$ is $10bp$, while σ_ℓ is set so that the annual standard deviation of ℓ_t^* is 1%. Both choices are meant to ensure that these parameters will only vary slowly over time in order to not affect the cyclical fit of the model. We choose starting values $\overline{bk}_0$ and ℓ_0^* by minimizing the sum of squared residuals in the measurement equation (27), and find σ_R iteratively as the value that matches the resulting standard deviation of the residuals in the measurement equation.

The resulting time series of the latent factors are depicted in the upper panel of Figure 9. The thicker black line reflects the time series of operating cost $\overline{bk}_t$, which is in annual terms initially slightly higher than 2%, but quickly declines in the late 1970s to about 1%. From there we see a slight increase during the late 1980s and early 1990s to about 1.3%, a level that is roughly constant until the financial crisis of 2008, after which the cost is going back up. The grey line denotes the time variation in optimal leverage ℓ_t^*, which banks would choose if the shadow spread was zero, i.e. if collateral was abundant. This series shares the overall dynamics of the evolution of $\overline{bk}_t$.

The lower panel of Figure 9 compares the new model fit to the data. As can be seen, the slow moving structural changes in the cost function parameters lead to an improved model fit over the whole sample, while maintaining the cyclical fit from the previous section. The correlation between leverage in data and model is now 97%. Importantly, the model is able to match the increase in our leverage measure after 2008. The top panel shows that this increase in leverage is in our results driven by an increase in both $\overline{bk}_t$ and ℓ_t^*. As with the one-time structural break, it is again plausible to associate the joint movements between the two series with an increase in $\bar{k}_t$, although the other two parameters must also have changed to jointly explain both series over the whole sample.

The estimated parameter series still feature some cyclical fluctuations, in particular during the 1980s. In order to make sure that these movements are not driving the improved model fit, we use smoothing splines to further remove any higher frequency parameter changes. The smoothed estimates are depicted as thin lines in the upper panel, and a thin black line shows the model prediction using those parameters in the lower panel. As can be seen, the model fit is virtually unchanged.

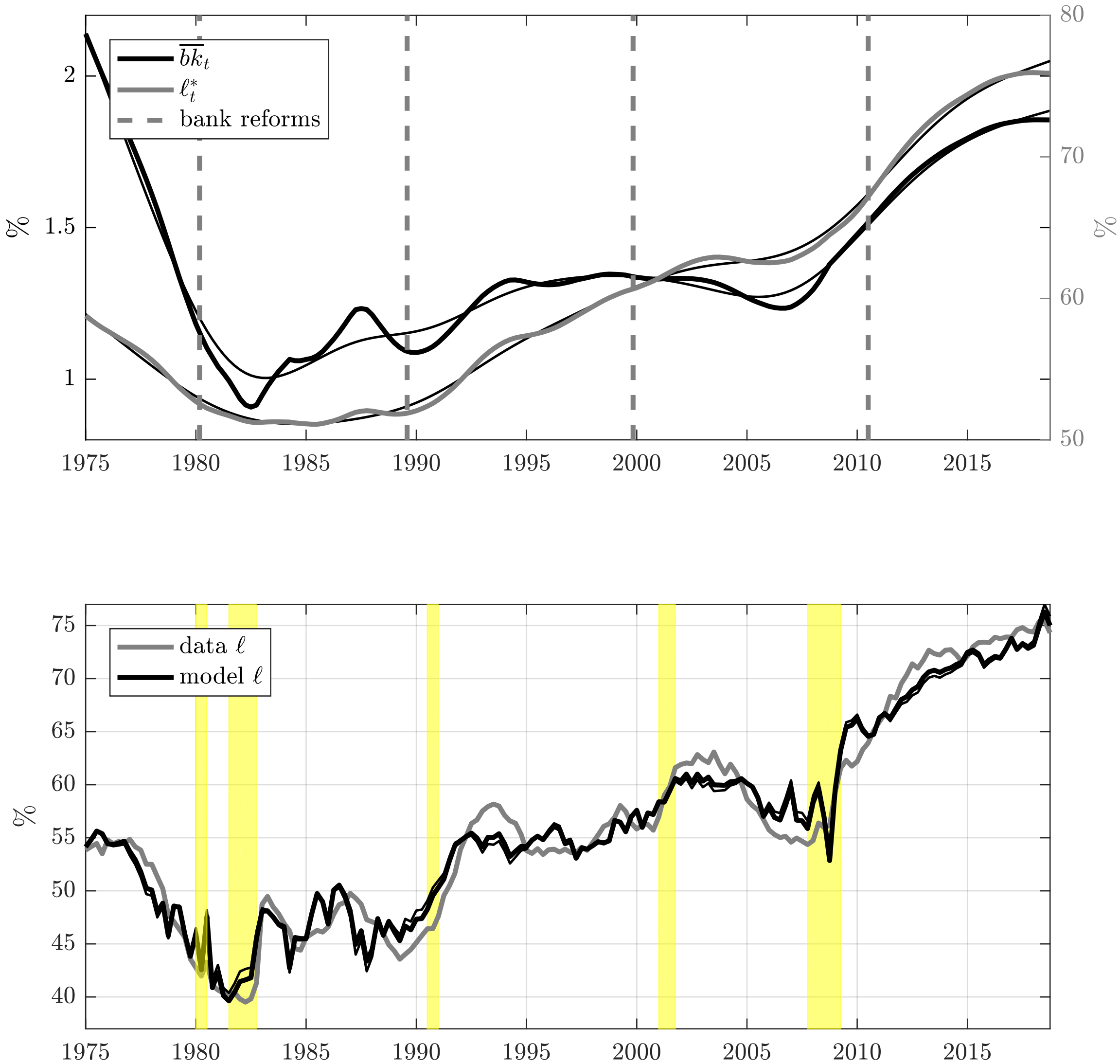

Figure 9: Top panel: Estimated time series of annual operating cost $\overline{bk}_t$ and maximum leverage ℓ_t^* (thick lines). The thin black lines are generated with smoothing splines to further remove any higher frequency parameter movements. The grey dotted lines mark the 1980 "Depository Institutions Deregulation and Monetary Control Act", the 1989 "Financial Institutions Reform and Recovery Act", the 1999 "Gramm-Leach-Billey Act" and the 2010 "Dodd-Frank Act". Bottom panel: Leverage of payment intermediaries in the data (grey) and model (black). The thick black line is generated using the parameter series from the original estimation, while the thin black line is generated using the smoothed estimates.

An increase in $\bar{k}_t$ provides as before a rationale for why we might observe an increase in leverage ℓ_t after 2008: if bank regulation induces higher operating cost $\overline{bk}_t$, both in absolute and relative terms through increases in $\bar{k}_t$, holding assets inside the bank becomes more expensive so that households reduce asset holdings inside the bank by lowering bank equity. This economized production of deposits may also explain the increase of ℓ_t^* after the 1989 "Financial Institutions Reform and Recovery Act", which also tightened bank regulation. Overall we observe that break points in the two latent factor series are roughly associated with the four major bank reforms in the data, lending support to our idea of capturing structural changes in banks' balance-sheet cost.

Estimating return risk As in the previous section, we can use our estimates to back out the evolution of $\gamma\sigma_t^2$. Figure 8 depicts the resulting time series as a grey line, which looks similar to the series estimated without structural change, if somewhat smaller and more stable. When we again back out a value for γ by imposing that the average level of σ_t has to match the level of return volatility of the S&P 500 stock index, we now find that a lower curvature level of $\gamma = 31.7$ can match the return volatility target.

Banks' balance-sheet cost With our estimate of γ at hand, we can also evaluate whether our estimated cost function is economically sensible. We use the definition of $\ell_t^* = \left(\frac{\bar{k}_t}{\gamma-1}\right)^{1/\gamma}$ to derive a time series estimate of $\bar{k}_t$. We then find $b_t = \overline{bk}_t/\bar{k}_t$, which we can use to calculate the balance-sheet cost for any level of realized leverage $\tilde{\ell}_{t+1}$ as $k(\tilde{\ell}_{t+1}) = \overline{bk}_t + b_t\tilde{\ell}_{t+1}^{\gamma}$. Figure 10 depicts the balance-sheet cost for historic levels of leverage, portfolio shares and asset returns, given the estimates γ, b_t, $\bar{k}_t$ and σ_t. The three lines depict balance-sheet cost given the realization of the expected return (black), the realization of a one standard deviation negative return shock (grey) and the realization of a two standard deviation negative return shock (dashed black). Periods in which costs are relatively robust to return shocks are either times of low leverage, high safe asset shares or low return volatility, or a combination of those factors. We find that our estimated cost function yields reasonable levels of balance-sheet cost for plausible return scenarios. For even more negative return shocks the cost can quickly increase due to the high curvature in the cost function.

5 Conclusion

The results presented in this paper support the idea that financial intermediaries value short bonds as safe collateral to back the issuance of payment instruments. The emerging collateral premium drives a wedge between the policy rate and the short rate associated with the pricing kernel of non-bank investors. In our model, this short-rate disconnect has implications for banks' balance-sheet decisions, which are consistent with data. Our findings question the

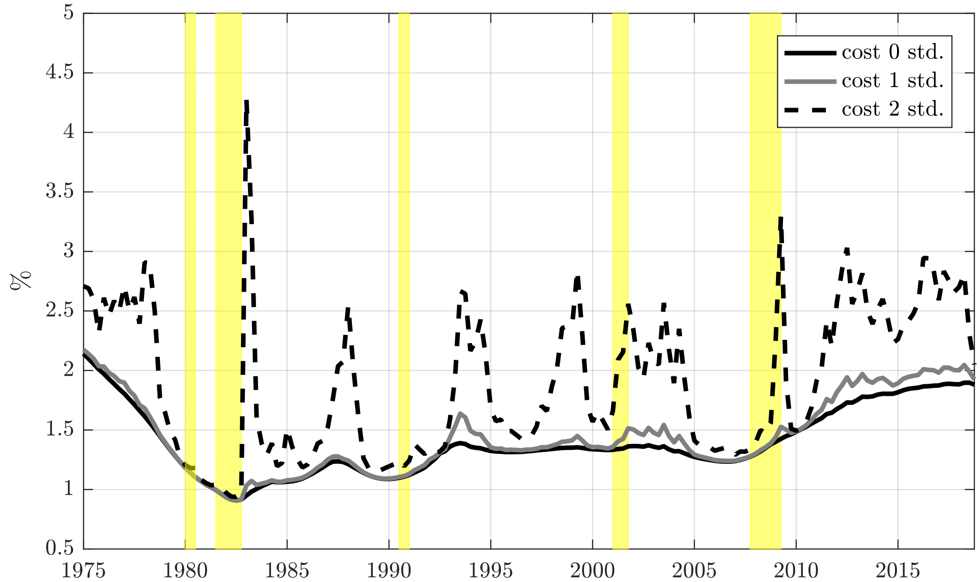

Figure 10: Estimated balance-sheet cost given parameter estimates γ, b_t and $\bar{k}_t$ and given historic choices of leverage ℓ_t, portfolio shares α_t, bond returns, shadow spread and σ_t.

standard assumption in current models of monetary policy, that the policy rate has an immediate link to the pricing kernel of non-bank investors. Motivated by these results, Piazzesi, Rogers and Schneider (2019) explore the implications of the short rate disconnect in a New Keynesian model, and find that this loss of immediacy can indeed fundamentally change the transmission of monetary policy. Further analysis of these mechanisms provides a promising avenue for future research.

References

Adrian, Tobias, Erkko Etula, and Tyler Muir, "Financial Intermediaries and the Cross-Section of Asset Returns," *Journal of Finance*, 2014, *LXIX*, pp. 2557–2596.

Andolfatto, David and Stephen Williamson, "Scarcity of safe assets, inflation, and the policy trap," *Journal of Monetary Economics*, 2015, *73* (C), pp. 70–92.

Bansal, Ravi and Wilbur John Coleman, "A Monetary Explanation of the Equity Premium, Term Premium, and Risk-Free Rate Puzzles," *Journal of Political Economy*, 1996, *104* (6), pp. 1135–1171.

Begenau, Juliane, "Capital Requirements, Risk Choice, and Liquidity Provision in a Business Cycle Model," 2019. Forthcoming *Journal of Financial Economics*.

Bhattacharya, Sudipto and Douglas Gale, "Preference shocks, liquidity, and central bank policy," in William A. Barnett and Kenneth J.Editors Singleton, eds., *New Approaches to Monetary Economics: Proceedings of the Second International Symposium in Economic Theory and Econometrics*, International Symposia in Economic Theory and Econometrics Cambridge University Press 1987, pp. 69–88.

Bianchi, Javier and Saki Bigio, "Banks, Liquidity Management and Monetary Policy," SSRN Scholarly Paper ID 2496246, Social Science Research Network, Rochester, NY 2014.

Bocola, Luigi, "The Pass-Through of Sovereign Risk," *Journal of Political Economy*, 2016, *124* (4), pp. 879–926.

Breeden, Douglas T., "An intertemporal asset pricing model with stochastic consumption and investment opportunities," *Journal of Financial Economics*, 1979, *7*, pp. 265–296.

Brunnermeier, Markus and Yann Koby, "The Reversal Interest Rate," 2018. Working Paper, Princeton.

_ **and Yuliy Sannikov**, "A macroeconomic model with a financial sector," *American Economic Review*, 2014, pp. pp. 379–421.

Brunnermeier, Markus K. and Lasse Heje Pedersen, "Market Liquidity and Funding Liquidity," *Review of Financial Studies*, 2009, *22* (6), pp. 2201–2238.

_ **and Yuliy Sannikov**, "The I Theory of Money," SSRN Scholarly Paper ID 2827562, Social Science Research Network, Rochester, NY 2016.

Campbell, John Y. and Luis M. Viceira, "Consumption and Portfolio Decisions When Expected Returns Are Time Varying," *The Quarterly Journal of Economics*, 1999, *114* (2), pp. 433–495.

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

Christiano, Lawrence, Roberto Motto, and Massimo Rostagno, "Financial Factors in Economic Production," 2012. Working paper, Northwestern.

Cúrdia, Vasco and Michael Woodford, "Credit Spreads and Monetary Policy," *Journal of Money, Credit and Banking*, 2010, *42* (s1), 3–35.

_ and _ , "The central-bank balance sheet as an instrument of monetarypolicy," *Journal of Monetary Economics*, 2011, *58* (1), pp. 54–79.

Del Negro, Marco, Gauti Eggertsson, Andrea Ferrero, and Nobuhiro Kiyotaki, "The Great Escape? A Quantitative Evaluation of the Fed's Liquidity Facilities," *American Economic Review*, 2017, *107* (3), pp. 824–857.

Diamond, William, "Safety Transformation and the Structure of the Financial System," 2019. Working Paper, Wharton.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl, "A Model of Monetary Policy and Risk Premia," *The Journal of Finance*, 2018, *73* (1), pp. 317–373.

Duffee, Gregory R., "Idiosyncratic Variation of Treasury Bill Yields," *The Journal of Finance*, 1996, *51* (2), pp. 527–551.

Fiore, Fiorella De, Marie Hoerova, and Harald Uhlig, "Money Markets, Collateral and Monetary Policy," Working Paper 25319, National Bureau of Economic Research 2018.

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

_ , Nobuhiro Kiyotaki, and Albert Queralto, "Financial crises, bank risk exposure and government financial policy," *Journal of Monetary Economics*, 2012, *59* (S), pp. 17–34.

Greenwood, Robin and Dimitri Vayanos, "Bond Supply and Excess Bond Returns," *Review of Financial Studies*, 2014, *27*, pp. 663–713.

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

Gurkaynak, Refet, Brian Sack, and Jonathan Wright, "The U.S. Treasury yield curve: 1961 to the present," *Journal of Monetary Economics*, 2007, *54*, pp. 2291–2304.

Haddad, Valentin and Tyler Muir, "Do Intermediaries Matter for Aggregate Asset Prices?," 2018. Working Paper, UCLA.

Haddad, Valtentin and David A. Sraer, "The Banking View of Bond Risk Premia," *Working Paper, Princeton and Berkeley*, 2018.

Hansen, Lars Peter and Kenneth J. Singleton, "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy*, 1983, *91*, pp. 249–265.

Hanson, Samuel G., Andrei Shleifer, Jeremy Stein, and Robert W. Vishny, "Banks as Patient Fixed-Income Investors," *Journal of Financial Economics*, 2015, *117*, pp. 449–469.

He, Zhiguo and Arvind Krishnamurthy, "Intermediary Asset Pricing," *American Economic Review*, 2013, *103* (2), pp. 732–770.

_, Bryan Kelly, and Asaf Manela, "Intermediary asset pricing: New evidence from many asset classes," *Journal of Financial Economics*, 2017, *126* (1), pp. 1–35.

Koijen, Ralph S. J. and Motohiro Yogo, "The Cost of Financial Frictions for Life Insurers," *American Economic Review*, 2015, *105* (1), pp. 445–475.

Krishnamurthy, Arvind and Annette Vissing-Jorgensen, "The Aggregate Demand for Treasury Debt," *Journal of Political Economy*, 2012, *120* (2), pp. 233–267.

Lenel, Moritz, "Safe Assets, Collateralized Lending and Monetary Policy," 2018. Working Paper, Princeton.

Moreira, Alan and Alexi Savov, "The Macroeconomics of Shadow Banking," *The Journal of Finance*, 2017, *72* (6), pp. 2381–2432.

Muir, Tyler, "Financial Crises and Risk Premia," *Quarterly Journal of Economics*, 2017, *132*, pp. 765–809.

Nagel, Stefan, "The Liquidity Premium of Near-Money Assets," *The Quarterly Journal of Economics*, 2016, *131* (4), pp. 1927–1971.

Patinkin, Don, *Money, Interest, and Prices: An Integration of Monetary and Value Theory*, Row, Peterson, 1956.

Piazzesi, Monika, "Bond Yields and the Federal Reserve," *Journal of Political Economy*, 2005, *113*, pp. 311–344.

__ **and Martin Schneider**, "Equilibrium Yield Curves," in Kenneth Rogoff Daron Acemoglu and Michael Woodford, eds., *Kenneth Rogoff Daron Acemoglu and Michael Woodford, eds.*, NBER Macroeconomics Annual 2006 Cambridge MA: MIT Press 2007, pp. pp. 389–442.

__ **and** __ , "Payments, Credit and Asset Prices," 2018. Working Paper, Stanford.

__ **, Ciaran Rogers, and Martin Schneider**, "Money and Banking in a New Keynesian Model," 2019. Working Paper, Stanford.

Reis, Ricardo, "Funding Quantitative Easing to Target Inflation," Technical Report 1626, Centre for Macroeconomics (CFM) 2016.

Teles, Pedro and Ruilin Zhou, "A stable money demand: Looking for the right monetary aggregate," *Economic Perspectives*, 2005, (Q I), pp. 50–63.

Tobin, James, "Commercial Banks as Creators of 'Money'," Technical Report 159, Cowles Foundation for Research in Economics, Yale University 1963.

Venkateswaran, Venky and Randall Wright, "Pledgability and Liquidity: A New Monetarist Model of Financial and Macroeconomic Activity," *NBER Macroeconomics Annual*, 2014, *28* (1), pp. 227–270.

Whitesell, William, "Interest rate corridors and reserves," *Journal of Monetary Economics*, 2006, *53* (6), pp. 1177–1195.

Williamson, Stephen D, "Liquidity, Monetary Policy, and the Financial Crisis: A New Monetarist Approach," *American Economic Review*, 2012, *102* (6), pp. 2570–2605.

Williamson, Stephen D., "Scarce Collateral, the Term Premium, and Quantitative Easing," *Journal of Economic Theory*, 2016, *164*, pp. 136–165.

Woodford, Michael, *Quantitative Easing and Financial Stability*, Santiago, Central Bank of Chile, 2016.

A Measuring the shadow rate

This section provides more details about our measure of the shadow rate. To obtain the shadow rate, we evaluate equation (9) in Gurkaynak, Sack and Wright (2007) at maturity 1/4 for their estimated parameter values. More precisely, our baseline computes the shadow rate as

$$f_t(0.25, 0) = \beta_0 + \beta_1 \exp(-0.25/\tau_1) + \beta_2(0.25/\tau_1)\exp(-0.25/\tau_1) + \beta_3(0.25/\tau_2)\exp(-0.25/\tau_2), \tag{30}$$

based on the six parameters β_0, β_1, β_2, β_3, τ_1 and τ_2 estimated by Gurkaynak et al. (2007). The estimated forward curve produces noisy estimates of the very short end of the yield curve. Our baseline therefore simply uses the three month forward rate to proxy for the three month yield. This approximation becomes exact as maturity tends to zero.

As an alternative approach, we have performed a first order Taylor expansion of the instantaneous forward curve around three months and computed the three month yield as the integral of this approximating forward curve from zero to three months. This approach uses also the derivative of the estimated forward curve, but leads to a shadow spread that is very similar to the baseline. We conclude that the overall effect of information contained in the curvature of the forward curve is not particularly important for the estimate of the shadow rate. We thus favor the baseline for its simplicity.

In the main text, we discuss other alternative measures; they are highly correlated with our baseline.

B Functional form derivations

This section derives the closed form solutions for leverage ℓ_t and the safe bond portfolio share $\alpha_t = \alpha_t^M + \alpha_t^B$. For the following, we define the weighted nominal rate on the portfolio of reserves and short bonds as

$$i_t^{MB} = \frac{\alpha_t^M}{\alpha_t} i_t^M + \frac{\alpha_t^B}{\alpha_t} i_t^B. \tag{31}$$

We start from the bank's Euler equations for the safe bond and the risky bond for the case in which the bank holds both of these assets:

$$\begin{aligned} E_t \left[M_{t+1} \left(1 - k\left(\tilde{\ell}_{t+1}\right) + k'\left(\tilde{\ell}_{t+1}\right)\tilde{\ell}_{t+1}\right)\left(1 + r_{t+1}^L\right)\right] &= 1, & (32)\\ E_t \left[M_{t+1}^{\$} \left(1 - k\left(\tilde{\ell}_{t+1}\right) + k'\left(\tilde{\ell}_{t+1}\right)\tilde{\ell}_{t+1}\right)\right]\left(1 + i_t^B\right) &= 1 & (33) \end{aligned}$$

We use our decomposition of ex-post leverage $\tilde{\ell}_{t+1}$ into ex-ante leverage

$$\ell_t = (1 + i_t^D) D_t / A_t \tag{34}$$

and the nominal risky return

$$1 + r_{t+1}^{\alpha,\$} = (1 + r_{t+1}^{\alpha}) P_{t+1} / P_t, \tag{35}$$

so that

$$\tilde{\ell}_{t+1} = \frac{\ell_t}{1 + r_{t+1}^{\alpha,\$}}. \tag{36}$$

Given the functional form assumption, we rewrite the Euler equations for the risky and safe bonds as

$$\begin{aligned} b(\gamma-1)\ell_t^{\gamma} &= b\bar{k}E_t\left[M_{t+1}^{\$}(1+r_{t+1}^{\alpha,\$})^{-\gamma}(1+r_{t+1}^{L,\$})\right]^{-1}, & (37)\\ 1 &= (1+i_t^B)\left(\frac{1-b\bar{k}}{1+i_t^S} + b(\gamma-1)\ell_t^{\gamma}E_t\left[M_{t+1}^{\$}(1+r_{t+1}^{\alpha,\$})^{-\gamma}\right]\right). & (38) \end{aligned}$$

Substituting out ℓ_t, we combine both equations to find

$$(1+i_t^B)\left(\frac{1-b\bar{k}}{1+i_t^S} + b\bar{k}\frac{E_t\left[M_{t+1}^{\$}(1+r_{t+1}^{\alpha,\$})^{-\gamma}\right]}{E_t\left[M_{t+1}^{\$}(1+r_{t+1}^{\alpha,\$})^{-\gamma}(1+r_{t+1}^{L,\$})\right]}\right) = 1. \tag{39}$$

To solve for the safe portfolio share α_t in closed form we use the usual small return approximation $1 + r_{t+1} \approx \exp(r_{t+1})$ and assume that the nominal return on the risky asset $\exp(r_{t+1}^{L,\$})$ and the household's nominal pricing kernel are jointly log-normal:

$$\begin{aligned} M_{t+1}^{\$} &= \exp\left(-i_t^S - \frac{1}{2}\lambda_t^{\top}\lambda_t - \lambda_t^{\top}\epsilon_{t+1}\right), & (40)\\ \exp(r_{t+1}^{L,\$}) &= \exp\left(\mu_t + \eta_t^{\top}\epsilon_{t+1}\right), & (41) \end{aligned}$$

for some standard normal vector ϵ. Following Campbell and Viceira (1999) we approximate the log portfolio return as

$$r_{t+1}^{\alpha,\$} \approx \alpha_t i_t^{MB} + (1-\alpha_t) r_{t+1}^{L} + \frac{1}{2}\alpha_t(1-\alpha_t)\sigma_t^2, \tag{42}$$

where we define $\sigma_t^2 = \eta_t^{\top}\eta_t$ as the risky bond's return variance.

For use in the following, we note that from $1 = E[\exp(M^{\$}_{t+1})\exp(r^{L,\$}_{t+1})]$ we find that

$$i^S_t = \mu_t - \lambda_t^\top \eta_t + \frac{1}{2}\sigma_t^2. \tag{43}$$

The conditional moments in the Euler equations can then be computed as

$$\begin{aligned}
&E_t\left[M^{\$}_{t+1}\exp(-\gamma r^{\alpha,\$}_{t+1} + r^{L,\$}_{t+1})\right] \\
=&E_t\left[\exp(-i^S_t - \frac{1}{2}\lambda_t^\top\lambda_t - \lambda_t^\top \epsilon_{t+1} - \gamma\alpha_t i^{MB}_t + (1-\gamma(1-\alpha_t))r^L_{t+1} - \gamma\frac{1}{2}\alpha_t(1-\alpha_t)\sigma_t^2)\right] \\
=&\exp(-\gamma(1-\alpha_t)\sigma_t^2 + \frac{1}{2}\gamma^2(1-\alpha_t)^2\sigma_t^2 + \gamma(1-\alpha)\eta_t^\top\lambda_t - \gamma\alpha_t i^{MB}_t - \gamma(1-\alpha_t)\mu_t - \gamma\frac{1}{2}\alpha_t(1-\alpha_t)\sigma_t^2)
\end{aligned}$$

and

$$\begin{aligned}
&E_t\left[M^{\$}_{t+1}\exp(-\gamma r^{\alpha,\$}_{t+1})\right] \\
=&E_t\left[\exp(-i^S_t - \frac{1}{2}\lambda_t^\top\lambda_t - (\lambda_t^\top + \gamma(1-\alpha_t)\eta_t^\top)\epsilon_{t+1} - \gamma\alpha_t i^{MB}_t - \gamma(1-\alpha_t)\mu_t - \gamma\frac{1}{2}\alpha_t(1-\alpha_t)\sigma_t^2)\right] \\
=&\exp(-i^S_t + \frac{1}{2}\gamma^2(1-\alpha_t)^2\sigma_t^2 + \gamma(1-\alpha_t)\lambda_t^\top\eta_t - \gamma\alpha_t i^{MB}_t - \gamma(1-\alpha_t)\mu_t - \gamma\frac{1}{2}\alpha_t(1-\alpha_t)\sigma_t^2)
\end{aligned}$$

so that

$$\frac{E_t\left[M^{\$}_{t+1}\exp(-\gamma r^{\alpha,\$}_{t+1} + r^{L,\$}_{t+1})\right]}{E_t\left[M^{\$}_{t+1}\exp(-\gamma r^{\alpha,\$}_{t+1})\right]} = \exp(i^S_t - \gamma(1-\alpha_t)\sigma_t^2). \tag{44}$$

We plug the above into equation (39) to find

$$\exp(i^B_t)\left((1-b\bar{k})\exp(-i^S_t) + b\bar{k}\exp(-i^S_t)\exp(\gamma(1-\alpha_t))\sigma_t^2)\right) \;=\; 1 \tag{45}$$

so that we can solve for the safe asset share α_t as

$$\alpha_t \;=\; 1 - \frac{1}{\gamma\sigma_t^2}\log\left(1 + \frac{i^S_t - i^B_t}{b\bar{k}}\right). \tag{46}$$

A higher return variance σ_t^2 of the risky bond and more curvature γ in the bank's asset management cost function increases the safe portfolio share. A higher shadow spread $i^S_t - i^B_t$ lowers the safe portfolio share.

We then rearrange the risky bond Euler equation to solve for leverage

$$\ell_t \;=\; \left(\frac{\bar{k}}{\gamma-1}\right)^{1/\gamma} \exp\Big(-\frac{1}{2}\gamma(1-\alpha_t)^2\sigma_t^2 + \alpha_t i_t^{MB} + (1-\alpha_t)(\mu_t - \eta_t^\top \lambda_t + \frac{1}{2}\sigma_t^2) + \frac{1}{2}(1-\alpha_t^2)\sigma_t^2\Big)$$

which yields

$$\ell_t \;=\; \left(\frac{\bar{k}}{\gamma-1}\right)^{1/\gamma} \exp\left(i_t^S - \alpha_t(i_t^S - i_t^{MB}) + \frac{1}{2}(1-\alpha_t^2)\sigma_t^2\right) \exp\left(-\frac{1}{2}\gamma(1-\alpha_t)^2\sigma_t^2\right).$$

Plugging in our result for the portfolio share from above, we find that

$$\ell_t \;=\; \left(\frac{\bar{k}}{\gamma-1}\right)^{1/\gamma} \exp\left(i_t^S - \alpha_t(i_t^S - i_t^{MB}) + \frac{1}{2}(1-\alpha_t^2)\sigma_t^2\right) \exp\left(-\frac{1}{2}\frac{1}{\gamma\sigma_t^2}\log\left(1 + \frac{i_t^S - i_t^B}{b\bar{k}}\right)^2\right).$$