NBER WORKING PAPER SERIES

QUANTITATIVE TIGHTENING WITH SLOW-MOVING CAPITAL

Zhengyang Jiang
Jialu Sun

Working Paper 32757
http://www.nber.org/papers/w32757

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2024

We thank Hengjie Ai, YiLi Chien, Anna Cieslak, Nicolas Crouzet, Xiang Fang, Paul Fontanier, Nicolae Gârleanu, Bryan Hardy, Arvind Krishnamurthy, Wenhao Li, Konstantin Milbradt, Gabor Pinter, Jacopo Ponticelli, Ishita Sen, and participants at the Yale Junior Macro-Finance Conference, Northwestern Kellogg, Wharton Valuation Workshop, Western Finance Association, and the Society for Economic Dynamics for discussions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2024 by Zhengyang Jiang and Jialu Sun. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Quantitative Tightening with Slow-Moving Capital
Zhengyang Jiang and Jialu Sun
NBER Working Paper No. 32757
July 2024
JEL No. E5,G12

ABSTRACT

We document shifts in investor composition during quantitative tightening, which suggest that investors adjust their portfolios at different speeds. To understand its implications for bond valuation, we develop a general equilibrium model which highlights the dynamic interaction between heterogeneous investors. In the model, long-term investors have higher risk-taking capacity, but face a portfolio adjustment cost; liquidity traders have lower risk-taking capacity, but can trade freely. Our model predicts a novel overshooting pattern: when the central bank unwinds its bond purchase, slow adjustment by long-term investors requires liquidity traders to absorb the imbalance, who demand a higher risk premium that creates excessive bond price decline and volatility in the short run. As a result, quantitative tightening is not simply a symmetric reversal of quantitative easing.

Zhengyang Jiang
Kellogg School of Management
Northwestern University
2211 Campus Drive
Evanston, IL 60208
and NBER
zhengyang.jiang@kellogg.northwestern.edu

Jialu Sun
Kellogg School of Management
jialu.sun@kellogg.northwestern.edu

1 Introduction

Central banks in many advanced countries are unwinding their bond holdings purchased during phases of quantitative easing (QE). These Treasury sales, commonly known as quantitative tightening (QT), are expected to lower bond prices and raise long-term interest rates. Can this happen in a controlled manner symmetric to the QE phase? Would financial frictions that made QE effective backfire during QT and make it ineffective or even destructive?

In this paper, we show that QT could be more than QE in reverse mode, because these policy actions shift the composition of the marginal investors in the Treasury market. We begin by documenting some new facts about investor composition during the past two QT episodes in 2017–19 and 2022–23. We classify typical financial intermediaries—brokers and dealers, banks, hedge funds—as the *liquidity traders* in this market, and we refer to the remaining market participants as the *long-term bond investors*. Consistent with the notion that liquidity traders are more attentive to the market, we find that they adjust their portfolio positions much more aggressively than domestic and foreign bond investors. In response to the Fed's sales of Treasury bonds, the liquidity traders expanded their Treasury holdings by 60% to 80% relative to their initial positions, whereas other investors responded by much less. As a result, the Treasury debt sold by the Fed was disproportionately absorbed by the liquidity traders.

Does this change in investor composition matter for bond valuation and the effectiveness of QT? Indeed, QT was accompanied by a negative return on the long-term Treasury bond portfolio. For example, from November 2021 to October 2023, the Bloomberg US Long Treasury Bond Index declined by 40%, an unusually large loss in the Treasury market which is traditionally regarded as safe and stable. However, this decline in Treasury prices could be driven by other factors such as economic fundamentals and monetary policy rates, which makes it difficult to interpret.

Motivated by these observations, we develop a general equilibrium model to study how the interaction between these two investor classes affects bond valuation. We assume that long-term bond investors have a higher risk-taking capacity, which makes them the more efficient holders of the Treasury bonds. However, they face a portfolio adjustment cost, which prevents them from

adjusting to their target portfolio instantly after shocks. In comparison, liquidity traders have a lower risk-taking capacity, but they can adjust their positions freely without transaction costs. As a result, they will be the first responders to fundamental or monetary shocks, and they charge a risk premium for providing this service.

Our model highlights a new, dynamic perspective for understanding the effects of quantitative easing and tightening. For simplicity, let us consider a transitory supply shock in period 1 that is similar to QE, which is fully reversed in period 2 in a manner similar to QT. We plot the outstanding quantity of the bond available to the public in Figure (1), Panel q, which declines in period 1 due to QE and reverts to its pre-shock level in period 2 due to QT. The reversion in period 2 is anticipated when QE happens in period 1.

In the absence of portfolio frictions, the bond price goes up slightly in period 1 as the central bank purchases the bond, and it reverts very close to the pre-shock level in period 2 as the central bank exits the market. In this case, the bond price dynamics exhibit a simple inverted-V-shape, as represented by the blue line in Figure (1), Panel p. Because both types of investors trade in the bond market without friction, the aggregate bond demand is relatively elastic and the required

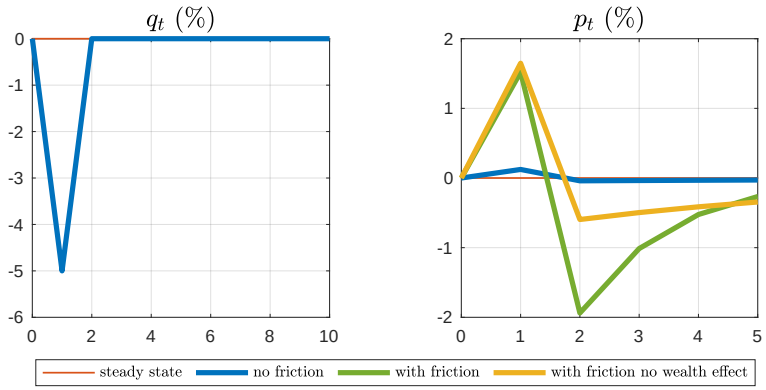


Figure 1: Price Responses to Bond Supply Shock

Notes: The left panel plots a transitory shock to the bond supply, which represents the central bank buying the bond in period 1 and selling it back in period 2. The right panel plots the bond price response. We consider three cases: without portfolio frictions, with portfolio frictions, and with portfolio frictions but keeping the long-term investors' wealth fixed.

price adjustment in response to the quantity change is small.

In the presence of portfolio frictions, this transitory supply shock could backfire and generate a novel overshooting pattern. In period 1, the bond demand becomes more inelastic: when the supply shock arrives, as the bond investors are reluctant to deviate from their prior portfolio shares in significant ways, they hold onto their bonds and crowd out the liquidity traders. Consequently, the liquidity traders hold a smaller quantity of bonds and demand a lower required return, which results in a higher bond price in period 1. Consistent with the findings in the literature, this inelastic demand curve allows QE to be effective in supporting the bond price.

However, when period 2 comes and the central bank unwinds its bond portfolio, the bond investors are starting from a lower bond position below their steady state. Their slow adjustment requires the liquidity traders to absorb the bonds offloaded by the central bank. Since the liquidity traders have a lower risk-taking capacity, they demand a higher expected return to hold more bonds, leading to a lower bond price which drops below the pre-shock level. In our calibration, the price drop in the QT period 2 is even greater than the price increase in the QE period 1, because it prices in the sum of higher expected returns required by the liquidity traders in all future periods in which they need to absorb the excess bond positions. In this way, QT generates a more complicated S-shape response in the bond price dynamics, as represented by the green line in Figure (1), Panel p, and produces excessive price decline and volatility in the QT phase.

We can further decompose this overshooting effect in the bond price into two components. First, the direct effect is driven by the long-term bond investors' slow adjustment to their target portfolio share, which lowers bond demand following QT and reduces the bond price. Second, since the bond investors' portfolio is highly exposed to the bond price, the overshooting effect is amplified by an additional wealth effect: as the bond price drops following QT, the long-term bond investors' wealth declines and reduces their risk-taking capacity. This wealth effect creates a negative feedback loop that further depresses the bond price. The yellow line in Figure (1), Panel p plots the counterfactual bond price dynamics which holds the long-term bond investors' wealth constant and thereby mutes the wealth effect. Under our calibration, the wealth effect accounts for

a large share of the post-QT bond price decline.

Quantitatively, our calibrated model implies that selling 7% of the outstanding Treasury debt as in the 2022–23 QT episode leads to a price decline of 5%. Du et al. (2024) focus on the bond yield responses during short announcement windows, and report that the cumulative announcement effect of the U.S. QT in 2022–23 is about 21 basis points on the 10-year bond yield and 19 basis points on the 30-year bond yield. These estimates correspond to roughly 2.1% and 5.7% decline in bond prices, respectively, which is consistent with our model's implication. That said, it is still possible that QT may have been partially anticipated and priced in by forward-looking investors. As a result, part of the price decline may have happened before the announcement windows, which contributed to the dramatic bond price decline in this period.

Moreover, the dynamic interaction between the two groups of investors also shapes the bond price response to cash flow shocks, which capture fiscal or fundamental shocks to the bond's real pay-off via either inflation or outright default. In our model, a positive cash flow shock is doubly good news because it not only raises the bond's pay-off but also lowers its discount rate. The discount rate effect is again due to the shock raising the bond investors' wealth, which allows them to take a greater bond position and crowd out the more risk-averse liquidity traders in this market. As the bond investors' wealth reverts to its steady-state level slowly over time, this discount rate effect can last longer than the underlying cash flow shock.

Our model also generates other realistic features of bond pricing dynamics. For example, the yield spread between the long-term bond and the short rate predicts both future interest rates and bond risk premia. The predictability of risk premium is much stronger in the presence of portfolio frictions, as the cash flow and the supply shocks generate greater and more persistent effects on the long-term bond's required return in this case. Moreover, similar to models of financial intermediation that emphasize the importance of the financiers' net worth, in our model the bond investors' wealth emerges as a measure of the aggregate risk-taking capacity, which is negatively associated with the equilibrium bond risk premium.

In reality, central bank bond purchases and sales are (i) persistent and (ii) often announced well

in advance. Is our result robust to these scenarios? In the first case, we examine a more persistent bond supply shock: after the central bank makes an unanticipated bond purchase today, it unwinds its position gradually over time. We find that a more gradual quantitative tightening policy can maintain the positive effect on the bond price upon bond purchase, while it alleviates the negative overshooting effect during the tightening phase by giving the long-term bond investors more time to adjust their portfolios. This result explains why central bank bond sales need to be slow and steady. However, this is not a free lunch: when the supply shock is persistent, bond investors experience a prolonged decline in wealth because the central bank's bond holding persistently depresses the return on the bond, which could reduce their risk-taking capacity if they also trade in other markets such as the corporate bonds and ABS markets.

In the second case, we assume that the central bank announces a bond purchase that is scheduled to occur in a later period. We find that the bond price rises immediately upon receiving this news, as market participants anticipate the forthcoming supply shock and its positive impact on the bond price. As a result, forward guidance in the form of announcements of future bond purchases can proactively boost the bond price. However, as long as the bond purchase is quickly unwound, advance announcements do not alleviate the overshooting effect. This is because, despite the investors' anticipation of the event, they are unable to adjust their portfolios gradually in advance, because they need to accommodate the bond purchase in the QE phase in the first place.

In conclusion, we document changes in investor composition during the QT episodes in 2017–19 and 2022–23, which are accompanied by negative bond returns. Motivated by these empirical observations, we develop a general equilibrium model to study how investor heterogeneity and portfolio frictions generate a rich dynamic pattern for bond prices and portfolio holdings. Our model highlights a novel overshooting effect resulting from the dynamic interaction between different investor classes, which makes QT not merely QE in reverse mode. Our model remains simple and solvable as a linear equation system with forward-looking variables, which makes it easy to incorporate our mechanism into a DSGE framework and explore how the portfolio friction interacts with the real economy. We leave this path for future work.

Literature. This paper contributes to our understanding of the effects of unconventional monetary policies. Building on the "portfolio rebalancing" channel that captures the direct effect of bond purchases on bond pricing, we develop a model to understand how investor heterogeneity and portfolio frictions shape the price dynamics (Tobin, 1958, 1969; Vissing-Jorgensen and Krishnamurthy, 2011; Christensen and Rudebusch, 2012; Greenwood and Vayanos, 2014; Greenwood, Hanson, and Vayanos, 2016; Koijen, Koulischer, Nguyen, and Yogo, 2021; Vayanos and Vila, 2021).[1] The most related paper is Greenwood, Hanson, and Liao (2018), who study how portfolio frictions transmit the effect of quantitative easing to other asset classes. Building on this literature, we focus on how portfolio frictions break the symmetry between QE and QT and generate non-monotonic dynamics over time.[2]

On the empirical side, Du, Forbes, and Luzzetti (2024) conduct a comprehensive empirical analysis of QT across developed countries. Consistent with our results, they also document that non-banks play an important role in absorbing the Treasury bonds sold by central banks. Moreover, D'Amico and Seida (2024) find that the Treasury yield sensitivity to QT surprises is larger than that to QE surprises, which is also consistent with our model's prediction. Du, Hébert, and Li (2023) also mention that dealers and hedge funds absorbed a large quantity of Treasury during QT1 in 2017–19.

Our model builds on heterogeneous-agent asset pricing models, which emphasize the heterogeneity in risk-taking capacity and in the access to financial markets (Chien, Cole, and Lustig, 2011; He and Krishnamurthy, 2013; Brunnermeier and Sannikov, 2014; Gabaix and Maggiori, 2015; Brunnermeier and Sannikov, 2016; Eisfeldt, Lustig, and Zhang, 2017; Haddad and Muir, 2021; Kekre, Lenel, and Mainardi, 2022). The most related paper is Chien et al. (2012), which examines how infrequent portfolio adjustment magnifies volatility. Our model studies the dynamic

[1] Neely (2011); Bauer and Neely (2014); Greenwood, Hanson, Stein, and Sunderam (2020); Gourinchas, Ray, and Vayanos (2022); Jiang, Krishnamurthy, Lustig, and Sun (2021) study how the effects of quantitative easing on bond risk premia and convenience yields also impact the exchange rate dynamics.

[2] Moreover, quantitative easing also impacts financial conditions by placing additional reserves on banks' balance sheets (Diamond, Jiang, and Ma, 2020; Copeland, Duffie, and Yang, 2021; Acharya, Chauhan, Rajan, and Steffen, 2022). The key message in our paper echoes this literature—the shrinkage of the central bank balance sheet may not be an entirely benign process.

interaction between agents facing different frictions and risk-bearing capacity, which generates a novel overshooting pattern in bond price as the shock dissipates in the medium run. Our work is also related to the empirical literature on investor demand in the safe asset market, which also emphasizes the role of investor heterogeneity (Koijen and Yogo, 2020; Jiang et al., 2022a,b; Fang et al., 2022; Coppola, 2022; Eren et al., 2023; Zhou, 2023; Breckenfelder and De Falco, 2023).

Finally, we build on the literature on slow-moving capital (Mitchell et al., 2007; Brunnermeier and Nagel, 2008; Calvet et al., 2009; Duffie, 2010; Bilias et al., 2010; Ameriks and Zeldes, 2011), which, in our model, is generated by the portfolio adjustment cost. Gârleanu and Pedersen (2013, 2016) examine this portfolio adjustment friction in a partial equilibrium portfolio choice model.[3] We embed the implied portfolio choice decision in a general equilibrium model with heterogeneous agents to study the dynamic interaction between constrained and unconstrained agents.

2 Investor Composition during Quantitative Tightening

We first document some new stylized facts about the change in investor composition during quantitative tightening. These results highlight the relevance of studying the heterogeneity in the investors' portfolio responses as a key to understand the pricing dynamics in the Treasury market, which we will use to motivate our heterogeneous-agent bond-pricing model in the subsequent section.

2.1 Data and Methodology

Investor Composition. We obtain data on the investor composition in the Treasury market from the Flow of Funds statistics, which breaks down the holdings of the U.S. government bills, notes and bonds by investor sectors such as pension funds, insurance companies, households, brokers

[3] Also see Duffie and Sun (1990); Heaton and Lucas (1996); Grinold (2006); Abel et al. (2007); Vayanos and Woolley (2013); Alvarez et al. (2012); Abel et al. (2013); Bacchetta et al. (2023). In the international context, Bacchetta and van Wincoop (2010); Bacchetta and van Wincoop (2021); Bacchetta, Davenport, and Van Wincoop (2022a); Bacchetta, van Wincoop, and Young (2022b) develop general equilibrium models with portfolio frictions to resolve interest rate and exchange rate puzzles.

and dealers, money market funds, and so on.

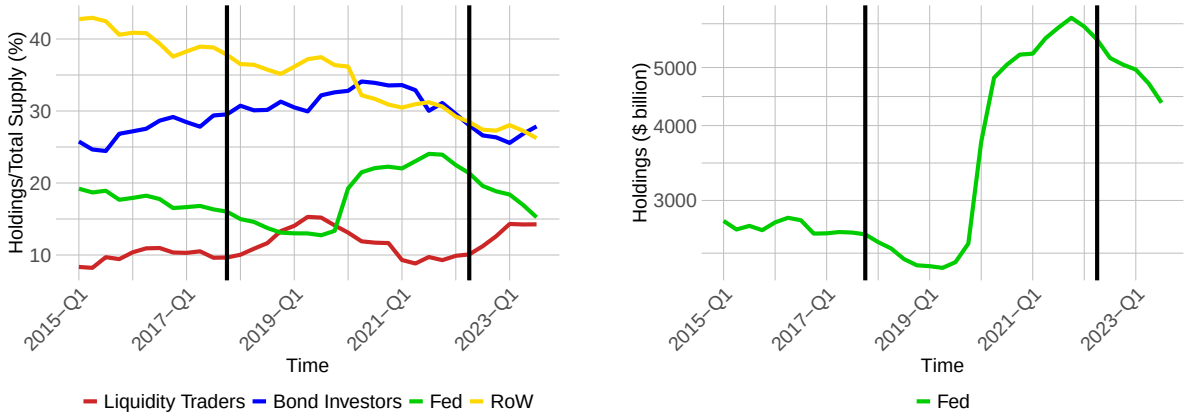
We group these investors into four categories: (i) liquidity traders, (ii) long-term bond investors, (iii) the Federal Reserve (Fed), and (iv) the rest of the world. The liquidity traders are the financial intermediaries in the Treasury market that facilitate trades and seek short-term profit opportunities in this market, which include the brokers and dealers, banks (including credit unions and US depository institutes), and households. We include the household sector in the flow of flows because it includes hedge funds, private equity, and private trusts. There is evidence that the majority of the change in Treasury holdings by the household sector during QT episodes is driven by the hedge funds. We present the details in Appendix A.3.

The long-term bond investors include all other domestic private sectors, including pension funds, insurance companies, ABS insurers, money market funds, ETFs, closed-end funds, mutual funds, government-sponsored enterprises, corporates, and holding companies. We typically think of these investors as the long-term holders of the Treasury debt who tend to trade less frequently. We do not include state and local governments, whose holdings we exclude from the total net supply. We include the Fed and the rest of the world as two separate categories.

Figure (2), Panel (a) plots the shares of Treasury holdings by these investor categories from 2015 to 2023. While the share held by foreign investors is decreasing, the shares of other investor categories are quite stable. The bond investors hold 25% to 35% of the market, the liquidity traders hold 8% to 15% of the market, and the Fed holds 12% to 25% of the market.

QT Episodes. We focus on two QT episodes and determine their beginning dates based on the FOMC meetings that first announced the start of the balance sheet normalization programs. The September 2017 FOMC meeting announced that QT1 will start in October 2017, and the May 2022 FOMC meeting announced that QT2 will start in June 2022. Hence, we assume that QT1 starts from 2017Q4, and QT2 starts from 2022Q2. We present relevant excerpts from the FOMC statements in Appendix A.2.

Figure (2), Panel (b) plots the level of the Fed's holdings in the Treasury market. The two



Panel (a) Shares Held by Investor Categories

Panel (b) Quantities Held by the Fed

Figure 2: Investor Composition in the Treasury Market

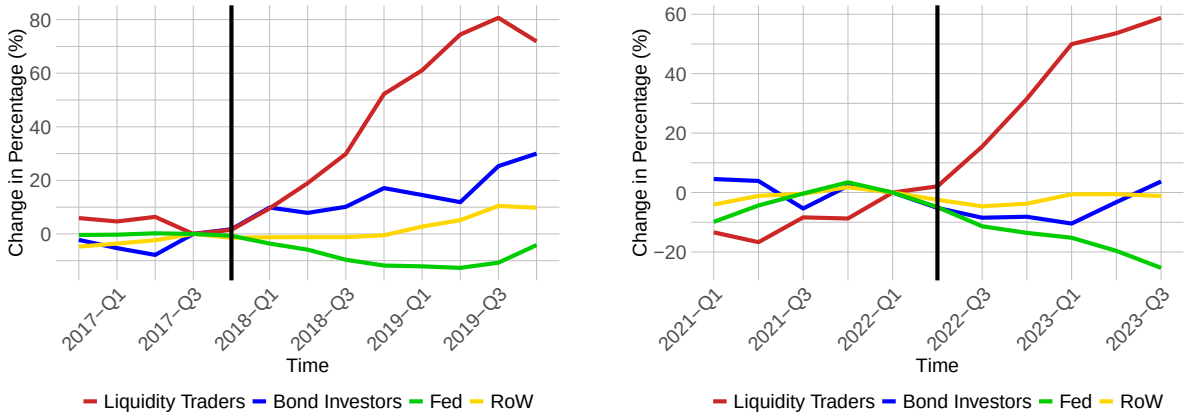
Notes: We report the share of investor holdings scaled by the aggregate supply of government debt in the left panel and the quantity of Fed's holdings (in log scale) in the right panel. The vertical black lines mark the beginning of the QT. The data are at quarterly frequency.

vertical lines mark the beginning of the two QT episodes. In QT1, the Fed's holdings declined by $285 billion from 2017Q3 to 2019Q3, or from 16% to 13% in terms of the share of the total market. In QT2, the Fed's holdings declined by $1,484 billion from 2022Q1 to 2023Q3 (our last available data point), or from 22% to 15% in terms of the share of the total market. These declines are sizable relative to the Fed's Treasury holdings.

2.2 Changes in Investor Composition during QT Episodes

Next, we zoom in on the two QT episodes and examine the change in investor composition during these episodes. In Figure (3), we plot the change in the holdings by each investor category scaled by the size of their holdings in the quarter prior to the QT episodes.

Panel (a) plots the investor composition around QT1. As the Fed reduced its Treasury holdings starting from 2017Q4, the liquidity traders increased their holdings quickly. By 2019Q3, their holdings increased by 80%. In comparison, the bond investors and the rest of the world increased their positions at a much slower pace. By 2019Q3, the bond investors' holdings increased by 30%, and the rest of the world's holdings increased only by 10%.



Panel (a): Changes Relative to 2017 Q3

Panel (b): Changes Relative to 2022 Q1

Figure 3: Investor Composition around QT Episodes

Notes: We report the change in Treasury holdings scaled by the initial quantities prior to the QT episodes. The vertical black lines mark the beginning of the QT. The data are at quarterly frequency.

Panel (b) plots the investor composition around QT2. Similarly, the reduction in the Fed's holdings was predominantly absorbed by the liquidity traders, who increased their holdings by 60% by 2023Q3. The bond investors increased their positions by less than 10%.

Together, these patterns during QT episodes suggest that the liquidity traders—brokers and dealers, banks, and hedge funds—play an important role in absorbing the Fed's sales of the Treasury bonds. Although their holdings are small relative to the holdings by the bond investors and the rest of the world, they adjust their positions quickly and increase their positions by over 50% within one year. In other words, they bear a disproportionate amount of risk and imbalance during the QT episodes we study.

In Appendix A.4, we present more granular breakdowns of the investor composition within the categories of the liquidity traders and the bond investors. Brokers and dealers, banks, and hedge funds are the major investor classes within the liquidity trader category, and they all contributed to the fast increase in the liquidity traders' holdings.

In Appendix A.5, we also conduct this analysis for the QE episodes. For the brokers and dealers who account for a major part of the liquidity traders, the patterns are symmetric: they quickly reduced their positions as the Federal Reserve bought the Treasury bonds. For the household sec-

tor, their responses were consistent with the brokers and dealers in QE2 (2020–2022), but not in QE1 (2008–2013). While the variations in the households' holdings were consistent with the available data on the hedge funds' holdings in QE2, we do not have data for QE1. It is possible that the households' holdings in QE1 were driven by other investor types.

Finally, while our analysis focuses on the investors' asset holding in the Treasury debt market (the cash market), some investors are also active in the Treasury futures market. Liang (2024) shows that, while hedge funds took long positions on Treasury debt in the cash market, they also took short positions in the Treasury futures market. On the other hand, mutual funds and other asset managers took small positions in the cash market but large and positive positions in the futures market after QT. This pattern suggests that hedge funds acted like a pass-through between the cash and futures markets. They did not take large directional bets on Treasury debt, but instead did basis trades which exploited the price discrepancy between the two markets (Glicoes et al., 2024). In contrast, mutual funds and other asset managers provided liquidity indirectly by trading in the futures market, as their cash positions on Treasury debt were subject to leverage constraints. Thus, if we jointly consider the Treasury cash and futures markets, we may need to classify mutual funds and other asset managers as liquidity traders as well. However, this modification does not affect our estimate of how much Treasury debt sold by the Fed was absorbed by the liquidity traders, since the indirect positions taken by asset managers in the futures market were already reflected in the direct positions taken by hedge funds in the cash market.

2.3 Bond Returns during QT Episodes

How do these dynamics in investor composition align with variations in the valuation of the Treasury bonds? In Figure (4), we plot the returns on Treasury bonds. We compute the returns from security-level data on CRSP, and construct return indices (i) for all Treasury bills, notes, and bonds, (ii) for bonds with maturities between 5 and 10 years, and (iii) for bonds with maturities between 10 and 30 years. The Treasury bonds with maturities between 10 and 30 years lost 3.65% in cumulative return in the one-year period following QT1, and they lost 9.14% in the one-year period

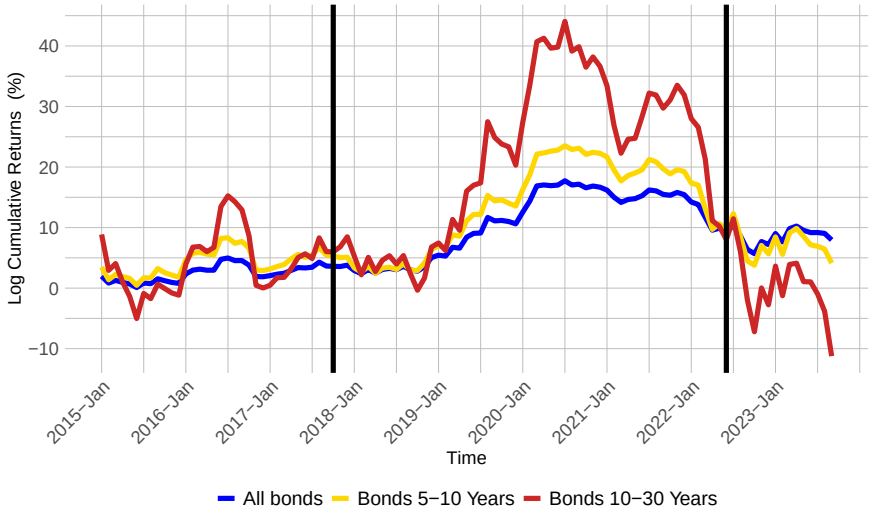


Figure 4: Cumulative Treasury Debt Returns

Notes: We plot the cumulative, cum-coupon returns of the Treasury debt in log. All Treasury includes all bills, notes, and bonds. Bonds 5–10 Years include all bonds with maturities between 5 and 10 years. Bonds 10–30 Years include all bonds with maturities between 10 and 30 years. The vertical black lines mark the beginning of the QT. The data are at monthly frequency.

following QT2. In comparison, the Treasury bills and the bonds with lower maturities were less impacted.

These patterns suggest that the shift in investor composition during the QT episodes coincided with a decline in bond price, especially during QT2. A natural question is whether the heterogeneity in investors' quantity responses contributed to the bond price decline. Empirically, it is difficult to identify the effect of QT, because demand shocks and monetary policy rates might be shifting at the same time. The QT events themselves could also contain information about other drivers of bond valuation. As a result, we turn to a theoretical approach in the next section to study how the changes in investor composition affect the valuation of the Treasury bonds.

3 Model

Our empirical observations identify a class of investors—which we refer to as the liquidity traders—who respond quickly to the Fed's sales of the Treasury bonds and expand their Treasury holdings within a short period of time. In comparison, the bond investors who usually have a longer in-

vestment horizon respond more slowly. Motivated by these observations, we develop a general equilibrium model of bond valuation with heterogeneous investors. We place an emphasis on how, in the presence of trading frictions, QT changes the composition of the marginal investors in the market and as a result affects the valuation of the Treasury bonds.

3.1 The Asset Space

We consider a simple market economy with two financial assets. One asset, referred to as *the reserve*, is a one-period bond with a risk-free interest rate i_t. This asset represents the interest-bearing reserve. As the central bank stands ready to buy or sell the reserve at the given interest rate i_t, we assume that it has a perfectly elastic supply.

The other asset, which we refer to as *the (long-term) bond*, is a perpetual bond that pays a time-varying coupon in each period. In period t, the ex-coupon bond price is denoted by p_t, and the bond pays a stochastic coupon denoted by x_t. This asset represents the government's long-term debt liabilities. While most government debt promises a fixed nominal payoff, its real value may vary and render the debt risky in real terms. The stochastic payoff could also be interpreted as capturing government default or variations in the fiscal conditions (Jiang et al., 2019).[4] For concreteness, we adopt the inflation interpretation. Then, we can think of the steady-state value $\bar{x}$ as capturing the promised nominal payoff, and the deviation $x_t - \bar{x}$ as the variations in the real cash flow due to inflation.

The cum-coupon holding period return of the bond is defined as

$$r_{t+1} = \frac{p_{t+1} + x_{t+1}}{p_t}.$$

The bond supply is denoted by q_t. We assume that the bond coupon x_t, the bond quantity q_t, and the reserve interest rate i_t are exogenously given. The bond coupon reflects the fundamental

[4]Moreover, from the perspective of foreign investors, the variations in x_t can also reflect the exchange rate risk. In this case, the foreign investors are choosing between the reserve in their own country and the perpetual bond issued by the U.S.

shocks to the government debt, the bond quantity reflects the government's issuance decision as well as the central bank's quantity-based monetary policy, and the reserve interest rate reflects the central bank's conventional monetary policy.

3.2 Investors

There are two types of investors. Both take prices as given. We assume that one type of investors is a more natural holder of risky assets because they have a lower risk aversion. However, they cannot always hold as much risky asset as they desire because they face trading frictions. The other type of investors charges a higher premium for holding risky assets, but they stand ready to absorb the excess amount of risky assets without trading frictions. For stationarity, we assume an overlapping generations (OLG) structure for both types of investors.

Bond Investors. The first type is bond investors, who allocate their portfolio between the perpetual bond and the reserve. They live for one period and inherit some wealth from the previous generation. The portfolio shares that a bond investor allocates to the perpetual bond and reserve are z_t and $1 - z_t$, respectively. Then, the investor's portfolio return in period $t + 1$ can be written as $r^p_{t+1} = i_t + z_t(r_{t+1} - i_t)$.

In period t, the investor's objective is to choose the portfolio allocation to maximize the expected excess return adjusted by risks and trading costs. The objective function is

$$\max_{z_t} \quad \mathbb{E}_t \left[r^p_{t+1} \right] - \frac{1}{2}\gamma \cdot var_t \left(r^p_{t+1} \right) - \frac{1}{2}\mu \cdot c(z_t; z_{t-1}, \bar{z})\sigma^2_t, \tag{1}$$

where γ is bond investor's risk-aversion coefficient, and $\sigma^2_t = var_t(r_{t+1})$ is the conditional variance of the bond return. This objective function can be motivated by the optimization problem of an investor who has a power utility over wealth in the next period, and faces a transaction cost that directly lowers the portfolio return. We present this derivation in Appendix B.1.

The trading cost function $c(z_t; z_{t-1}, \bar{z})$ is defined as

$$c(z_t; z_{t-1}, \bar{z}) = (1-\eta)(z_t - z_{t-1})^2 + \eta(z_t - \bar{z})^2, \tag{2}$$

which assumes that the bond investors face two types of costs. First and foremost, they are penalized for changing their portfolio weight z_t relative to the last period's value z_{t-1}. This is the key friction that generates our main results. Second, the bond investors are also penalized for deviating from the benchmark portfolio allocation $\bar{z}$, which can be motivated by the investors' desire to stay close to a benchmark portfolio. This assumption does not affect the qualitative response of the bond price to shocks, but it produces quantitatively more realistic dynamics in the medium run. We pick a small $\eta = 0.1$, so that the trading costs are mainly driven by rebalancing the portfolio weight relative to the last period's value. The trading cost is scaled by the asset return variance σ_t^2 following Gârleanu and Pedersen (2013).

Denote the bond's excess return as

$$er_{t+1} = r_{t+1} - i_t.$$

Then, the first-order condition with respect to z_t implies the following solution for the optimal portfolio weight z_t:

$$z_t = \frac{\gamma}{\gamma+\mu}\frac{\mathbb{E}_t[er_{t+1}]}{\gamma\sigma_t^2} + \frac{\mu}{\gamma+\mu}(\eta\bar{z} + (1-\eta)z_{t-1}), \tag{3}$$

which is a weighted average between $\mathbb{E}_t[er_{t+1}]/(\gamma\sigma_t^2)$, the standard mean-variance solution for a one-period investor, and $(\eta z_{t-1} + (1-\eta)\bar{z})$, the reference portfolio determined by the previous period's portfolio weight z_{t-1} and the benchmark portfolio allocation $\bar{z}$.

In Appendix C.3, we consider a more general case in which the bond investors have an infinite horizon. In this case, when they decide their portfolio allocation, they take into account not only how the current portfolio deviates from the previous position, but also how it affects the transaction

costs arising from future portfolio adjustments. This additional consideration effectively makes the investors even more reluctant to deviate from the given path of portfolio allocation, which further lowers the demand elasticity in a way similar to increasing the trading cost parameter μ. Quantitatively, our results remain robust to this extension. As such, we stick to the baseline model which has a simpler algebraic solution and more intuitive interpretation.

Finally, let w_t denote the bond investors' wealth. We assume its law of motion follows an autoregressive process:

$$w_t = (1-\phi)w^e + \phi w_{t-1} r_t^p,$$

which introduces a bequest motive to ensure stationarity in a simple way. Specifically, the bond investors of generation t receive a fixed endowment w^e and also inherit part of the wealth from the previous generation $t-1$. The previous generation $t-1$ bequeaths a fraction ϕ of their wealth to the next generation, and consumes the remaining wealth. In this way, after many generations following an initial shock, the bond investors' wealth will return to a steady-state level $\bar{w}$ in expectation.

Liquidity Traders. The second class of investors is the liquidity traders, with a constant mass of m. They also face the mean-variance trade-off over their portfolio return, but they do not pay the transaction cost that the bond investors do. These liquidity traders take a zero-cost long-short position on the bond versus the reserve. Let ℓ_t denote their holdings of the bond, and $-\ell_t$ denote their holdings of the reserve. The return on their portfolio in period $t+1$ is given by $r_{t+1}^{\ell} = \ell_t(r_{t+1} - i_t)$.

At time t, liquidity trader's objective is to choose their portfolio allocation in bond, ℓ_t, to maximize the present value of expected excess returns, penalized for risks.

$$\max_{\ell_t} \quad \mathbb{E}_t\left[r_{t+1}^{\ell}\right] - \frac{1}{2}\omega \cdot var_t\left(r_{t+1}^{\ell}\right), \tag{4}$$

where ω is the risk aversion coefficient of the liquidity traders. We assume that the liquidity traders

are more risk-averse than the bond investors, i.e., $\gamma < \omega$, so the bond has to offer a higher risk premium if it is held by the liquidity traders as opposed to the bond investors.

Solving the first-order condition with respect to ℓ_t, we obtain the following solution for their optimal portfolio choice ℓ_t:

$$\ell_t = \frac{\mathbb{E}_t[er_{t+1}]}{\omega\sigma_t^2}. \tag{5}$$

These liquidity traders play similar roles as the arbitrageurs in the preferred habitat models (Vayanos and Vila, 2021). However, unlike the preferred habitat investors in such models, the long-term bond investors in our model do not have a fixed habitat, and instead optimize their portfolio composition subject to trading frictions. As a result, our model generates richer predictions for the short-run and long-run elasticities of bond demand.

Alternatively, Gabaix and Maggiori (2015) provides a different microfoundation, in which these liquidity traders are risk-neutral but face a financial friction that is increasing in the variance of the risky asset's return. Regardless of the microfoundation, these liquidity traders absorb excess positions, both positive and negative, in the bond market from the bond investors as well as the government sector, and they charge a risk premium for providing this service. Their optimality condition implies a tight link between their portfolio position ℓ_t and the bond's discount rate $\mathbb{E}_t[er_{t+1}]$, which will be the key to understand the equilibrium outcomes as we next discuss.

The Government and the Central Bank. Besides the two types of private investors, there is a government sector who controls the aggregate supply of the bond. We model its action as an exogenous policy.

The government issues and retires the perpetual bond over time. Let q_t^{gov} denote the quantity outstanding in period t. The government's budget constraint is

$$(p_t + x_t)q_{t-1}^{gov} = s_t + p_t q_t^{gov},$$

where the left-hand side captures what the government owes and the right-hand side captures its primary surplus s_t and the proceeds from new issuances of the bond.

Moreover, the central bank can also buy and sell the perpetual bond. Let q_t^{cb} denote the quantity held by the central bank. Then, the total amount of the bond available to private investors can be written as $q_t = q_t^{gov} - q_t^{cb}$.

3.3 Market Clearing and Model Summary

The reserve has an elastic supply as the central bank stands ready to buy and sell at the given interest rate i_t. The market clearing condition for the bond is

$$w_t z_t + m\ell_t = q_t p_t.$$

The exogenous variables in this model are the bond coupon x_t, the bond quantity q_t, and the reserve interest rate i_t. We assume they all follow AR(1) processes:

$$\begin{aligned}
\log x_{t+1} &= (1-\theta_x)\log\bar{x} + \theta_x \log x_t + \sigma_x \varepsilon_{x,t+1}, \\
\log q_{t+1} &= (1-\theta_q)\log\bar{q} + \theta_q \log q_t + \sigma_q \varepsilon_{q,t+1}, \\
\log i_{t+1} &= (1-\theta_i)\log\bar{i} + \theta_i \log i_t + \sigma_i \varepsilon_{i,t+1}.
\end{aligned}$$

We log-linearize the model around the non-stochastic steady state, obtained by assuming that the exogenous innovations are zero for all periods t. In our derivation, we also simplify by assuming a constant return volatility: $\sigma_t = \bar{\sigma}$ for all t, which is consistent with the resulting log-linear solution. We use the hat notation for log deviation: $\hat{a}_t = \log(a_t/\bar{a})$ except for the portfolio shares $\hat{z}_t = z_t - \bar{z}_t$ and $\hat{\ell}_t = \ell_t - \bar{\ell}$. Then, we obtain the following result after log-linearization.

Proposition 1. *After log-linearization, we can describe the model using the following system of*

forward-looking linear equations, which includes the (linearized) bond return-price identity:

$$\hat{r}_t = \kappa_1 \hat{p}_t - \hat{p}_{t-1} + (1-\kappa_1)\hat{x}_t,$$

the bond's market clearing condition:

$$\hat{p}_t + \hat{q}_t = \frac{\bar{w}\bar{z}}{\bar{p}\bar{q}}\left(\hat{w}_t + \frac{1}{\bar{z}}\hat{z}_t\right) + \frac{m\bar{\ell}}{\bar{p}\bar{q}}\left(\frac{1}{\bar{\ell}}\hat{\ell}_t\right),$$

the law of motion for the bond investors' wealth:

$$\hat{w}_t = \bar{r}^p \phi \hat{w}_{t-1} + \bar{r}^p \phi \left(\hat{i}_{t-1} + \bar{z}\left(\hat{r}_t - \hat{i}_{t-1}\right) + \hat{z}_{t-1}(\hat{r}_t - \hat{i}_{t-1}) + \hat{z}_{t-1}(\log \bar{r} - \log \bar{i})\right),$$

the bond investors' portfolio choice:

$$\hat{z}_t = \frac{\mu}{\gamma+\mu}(1-\eta)\hat{z}_{t-1} + \frac{\gamma}{\gamma+\mu}\frac{1}{\bar{\sigma}^2\gamma}\mathbb{E}_t[\hat{r}_{t+1} - \hat{i}_t],$$

and the liquidity traders' portfolio choice:

$$\hat{\ell}_t = \frac{1}{\bar{\sigma}^2\omega}\mathbb{E}_t[\hat{r}_{t+1} - \hat{i}_t].$$

The proof is in Appendix B.3. This proposition ties the bond price to the bond's cash flows and discount rates via a standard present value formula:

$$\hat{p}_t = (1-\kappa_1)\sum_{j=1}^{\infty}\kappa_1^{j-1}\hat{x}_{t+j} - \sum_{j=1}^{\infty}\kappa_1^{j-1}\hat{r}_{t+j}, \tag{6}$$

and the economic content of this model is to connect the discount rates $\hat{r}_{t+j}$ endogenously to the two classes of investors' optimal portfolio choices. Using this proposition, we can solve the bond price dynamics around the steady state, as detailed in Appendix B.4.

4 Calibration and Results

4.1 Model Calibration

We calibrate the model annually and present our parameter values in Panel (a) of Table (1). Whenever applicable, we also report the calibration target we used in the last column, which is based on the security-level Treasury data from the Center for Research in Security Prices (CRSP). The data

Table 1: Model Parameter and Implied Steady-State Values

Panel (a) Parameters			
Interpretation	Symbol	Parameter Value	Empirical Moment
Bond Investor-Side Parameters:			
Risk-aversion coefficient	γ	2	
Steady-state wealth	$\bar{w}$	90	
Wealth replenishment rate	ϕ	0.90	
Portfolio adjustment penalty	μ	150	
Benchmark deviation penalty	η	0.1	
Liquidity Trader-Side Parameters:			
Risk-aversion coefficient	ω	6	
Mass	m	10	
Asset-Side Parameters:			
Steady-state bond quantity	$\bar{q}$	100	
Steady-state interest rate	$\bar{i}-1$	1.56%	1.56%
Steady-state coupon	$\bar{x}$	5.31%	5.31%
Bond return volatility	$\bar{\sigma}$	11.64%	11.64%
Interest rate volatility	σ_i	1.11%	1.11%
Bond quantity volatility	σ_q	5%	
Coupon volatility	σ_x	25%	
Panel (b) Implied Steady-State Values			
Bond investor's leverage	$\bar{z}$	119.03%	
Expected excess return of bonds	$\bar{er}$	3.11%	3.07%
Bond yield spread	$\bar{y}-\bar{i}$	3.20%	2.15%
Bond Sharpe ratio	$\bar{er}/\bar{\sigma}$	0.27	0.26

Notes: This table presents the parameters and implied steady-state values used in our simulations.

is monthly, and we pick the sample from 2000-January to 2023-September to better match the QE and QT episodes. We construct market value-weighted bond statistics for all bonds with date-to-maturity between 10 and 30 years to represent the long-term bond in our model. The short rate is monthly from Ken French's data library. In this sample, the average short rate is $\bar{i} - 1 = 1.56\%$, the short rate volatility $\sigma_i = 1.11\%$, the average coupon rate is $\bar{x} = 5.31\%$, and the bond return volatility $\bar{\sigma} = 11.64\%$. We use these values to directly calibrate the parameters in the model.

We set the portfolio adjustment penalty $\mu = 150$, which implies that a deviation from the benchmark portfolio share $z_{t-1} = \bar{z}$ by $\Delta z_t = \pm 10\%$ reduces the portfolio return by $\frac{1}{2}\mu(\Delta z_t)^2\bar{\sigma}^2 = 1.02\%$ per annum. While the magnitude of this friction is greater than the typical level of transaction costs in the Treasury market, we interpret the portfolio friction more broadly as the combination of various frictions that prevent the bond investors from fully rebalancing their portfolio. This parameter value also implies that the friction is not prohibitively high to eliminate the bond investors' trading activities.

For the other parameters, we pick values to jointly target empirical bond moments. Specifically, the cash flow volatility is $\sigma_x = 25\%$ and the bond quantity volatility is $\sigma_q = 5\%$. For the bond investors, their risk aversion is $\gamma = 2$. Their steady-state wealth is $\bar{w} = 90$, the wealth mean-reversion rate is $\phi = 0.90$, and the penalty is $\eta = 10\%$ on deviation from the benchmark portfolio and 90% on deviation from the last period's portfolio. For the liquidity traders, their risk aversion is $\omega = 6$, with a mass of $m = 10$.[5] These parameter values allow us to target the historical average bond excess return at 3.07% per annum, the bond Sharpe ratio at 0.26 per annum, and a bond demand elasticity of 3.1 which is close to Koijen and Yogo (2020)'s estimate using long-term debt.

We present the steady-state values implied by this set of parameter values in Panel (b) of Table (1). In the steady state, bond investors maintain a slightly levered position, holding a long position

[5]In our model, the liquidity traders' risk-taking capacity is summarized by the ratio between their mass m and their risk aversion coefficient ω. We could obtain identical results if we assume that ω is smaller and proportionally shrink the mass of the liquidity traders. Thus, the assumption that the liquidity traders are more risk-averse than the bond investors per se is not crucial. All we need is that the liquidity traders' risk-taking capacity is only a small fraction of the entire investor base, so that when the long-term bond investors face trading frictions, excess bond positions are concentrated in the hands of the liquidity traders, who require a higher risk premium as a compensation.

in bonds equal to 119.03% of their wealth and a short position in reserves equal to -19.03% of their wealth. The implied steady-state expected excess return of the bond is 3.11%, and the steady-state bond yield spread is 3.20%.[6]

4.2 Impulse Responses to Transitory Shocks

The Cash Flow Shock. First, we plot the impulse responses to a positive shock to the bond's cash flow x_t in Figure (C.5). The top panels report the dynamics of the bond valuation, including its cash flow x_t, ex-coupon price p_t, cum-coupon return r_t, and expected excess return $\mathbb{E}_t[er_{t+1}]$. The bottom panels report the dynamics of investors' holdings and wealth, including the bond investors' wealth w_t, their portfolio share in the bond z_t, their bond holding in dollar value $w_t z_t$, and the liquidity traders' holding in dollar value $m\ell_t$. In this plot, the shock is transitory: the cash flow x_t increases in period 1 and reverts to the steady-state value in period 2. We consider two variants of the model: one in which the bond investors do not face the transaction cost ($\mu = 0$), and one in which the bond investors face a transaction cost ($\mu = 150$).

We begin by noting the similarities between the two model variants. A high cash flow x_t in period 1 raises the bond's cum-coupon return r_t in period 1, which increases the bond investors' wealth w_t as they hold a considerable share of the bond in their portfolio. As the bond investors become wealthier, they hold a higher dollar amount $w_t z_t$ of the bond, even if the share of their wealth z_t allocated to the bond remains the same or, as the figure shows, declines slightly due to a lower expected bond return. Given a constant bond supply, this increase in the bond investors' holdings crowds out the liquidity traders' holdings of the bond $m\ell_t$, leading the liquidity traders to hold a smaller quantity of bond and require a lower bond risk premium for compensation in equilibrium.

In this sense, a positive cash flow shock is doubly good news for the bond market: it not only raises the cash flow, but also lowers the discount rate, which further boosts the bond valuation.

[6] For the convenience of discussion, we also introduce the bond yield y_t defined as the yield-to-maturity evaluated at the steady-state (nominal) cash flows, which solves $p_t = \sum_{h=1}^{\infty} \bar{x}/(y_t)^h$.

Conversely, a negative cash flow shock is doubly bad news, as it not only reduces the cash flow but also raises the discount rate. Moreover, beyond period 2, as the bond investors' wealth is a persistent state variable, their bond holdings $w_t z_t$ in dollar value remain high and continue to crowd out the liquidity traders' positions. As a result, the bond valuation remains high even after the direct cash-flow effect of the transitory shock disappears.

The two model variants also have important differences. If the bond investors face a transaction cost, when the bond's expected return declines in period 1, they become less willing to lower their portfolio share z_t allocated to the bond. As a result, the portfolio share z_t declines less upon the shock's arrival in the $\mu = 150$ case relative to the $\mu = 0$ case. This sluggish portfolio adjustment implies that the bond investors hold more bond $w_t z_t$ when the shock arrives, which further crowds

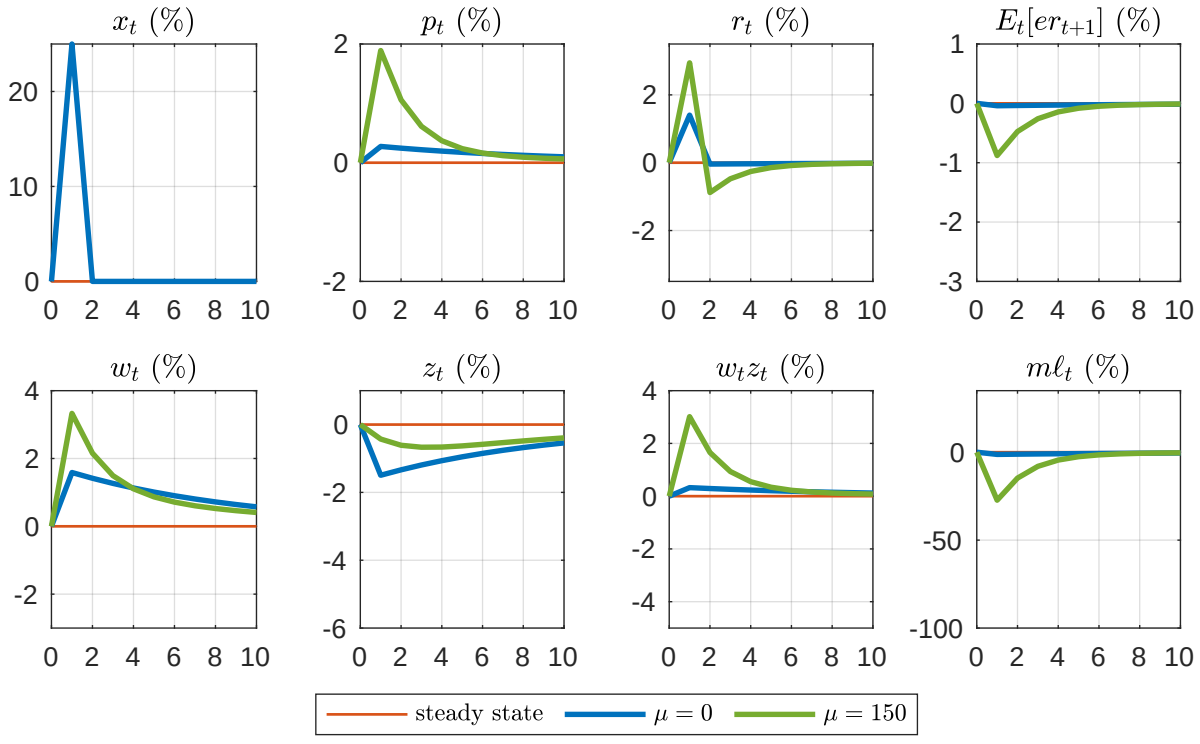


Figure 5: Impulse Response to a Transitory Cash Flow Shock

Notes: This figure reports the responses of the key state variables to a transitory shock to the bond's cash flow. We consider two cases: with and without transaction costs.

out the liquidity traders and leads to more dramatic responses in the bond price and the bond's expected return. In this way, the presence of portfolio friction makes the bond valuation more sensitive to the cash flow shock.

The Bond Quantity Shock. Next, we consider a transitory reduction in bond quantity q_t available to the public in period 1. As we discussed in the model setup, this shock can come from either a surprise reduction in government debt issuance, or bond purchase by the central bank. For the sake of discussion, we interpret this shock as central bank bond purchase. More precisely, since the bond quantity reverts to the steady-state level in period 2, this shock represents a transitory quantitative easing that is immediately unwound in the following period. We report the impulse responses in the same format in Figure (6).

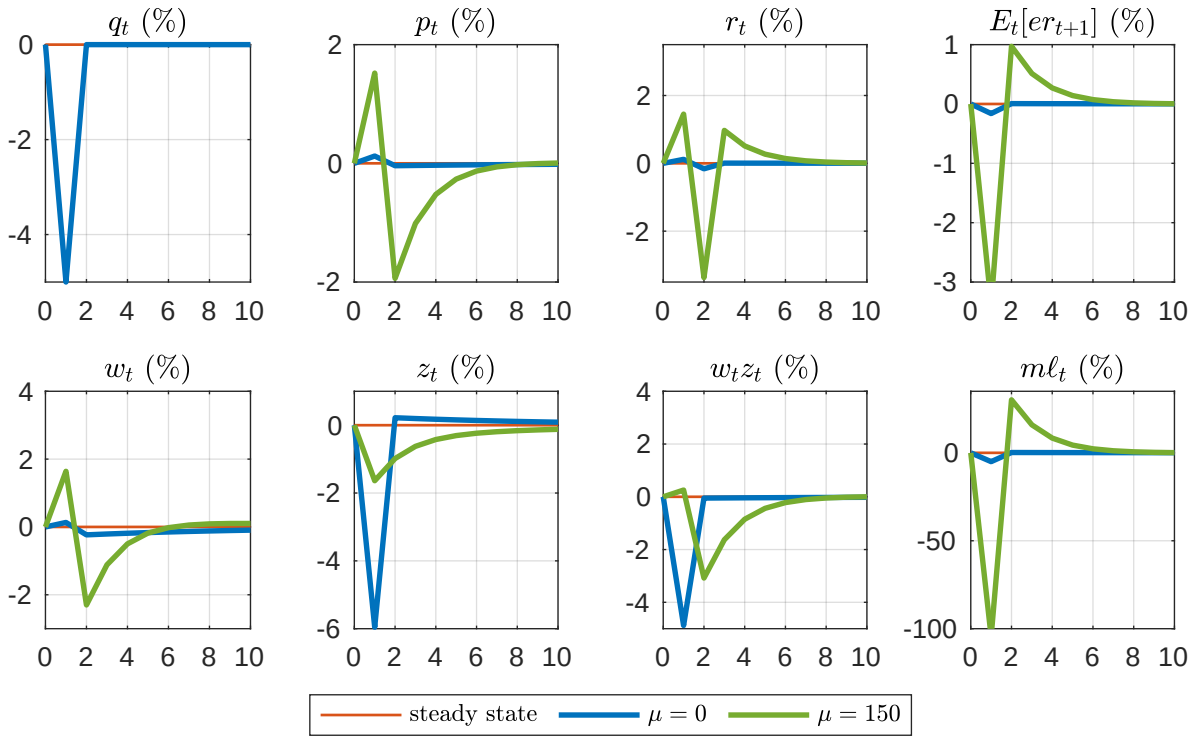


Figure 6: Impulse Response to a Transitory Supply Shock

Notes: This figure reports the responses of the key state variables to a transitory shock to the bond supply. We consider two cases: with and without transaction costs.

The frictionless case with $\mu = 0$ presents a simple story. As the bond quantity shrinks in period 1, the bond price p_t goes up, and the bond investors become more wealthy. Both the bond investors and the liquidity traders lower their holdings of the bond as the total supply shrinks, which generates a reduction in the discount rate that is consistent with the price increase. In period 2, as the supply shock dissipates, the bond price falls back and the bond return turns negative. This negative return generates a slight loss in the bond investors' wealth, which takes a few more periods to recover. In this way, the price response has a simple inverted-V shape that first goes up in the QE phase and then goes down in the QT phase.

The case with the transaction cost (i.e., $\mu = 150$) has much richer dynamics. Upon the arrival of the negative supply shock, the bond investors again have sluggish adjustment in their portfolio choice z_t and hold onto more bond, which requires the liquidity traders to sell more bond in equilibrium. As the liquidity traders require a lower expected return, the bond price p_t increases more in period 1. In other words, the portfolio friction makes the investors' demand more inelastic in the short run. Quantitatively, a 5% reduction in the bond quantity raises the bond price by about 1.5%, implying a demand elasticity of 3.3, which is consistent with the estimate Koijen and Yogo (2020) obtains in a global asset demand system.

Surprisingly, unlike the $\mu = 0$ case, the bond price does not revert to the steady-state level when the supply shock dissipates in period 2. Instead, the bond price experiences a significant drop below the steady-state level. To understand this result, we note that the quantitative easing lowers the bond investors' portfolio share z_t in period 1, which is sticky due to the portfolio friction. When the supply shock dissipates, the bond investors' bond holding z_t remains low. As a result, the liquidity traders have to hold more bonds from period 2 onwards, which raises the required discount rate and lowers the bond price below the steady-state value.

Quantitatively, this "overshooting" effect lowers the bond price by 2% below the steady state in period 2, which is even stronger than the supply shock's immediate positive effect in period 1. The key intuition is that the bond price is forward-looking: its response to the supply shock depends not only on how the shock affects the discount rate today, but also on how it affects the discount

rates in future periods. For example, in period 2, investors expect not only a higher bond discount rate $\mathbb{E}_2[er_3]$ today due to the liquidity traders' higher bond holdings, but also higher discount rates in future periods as this situation persists. As a result, the response of the period-2 bond price is greater than the response of the period-2 discount rate, which temporarily lowers the bond price below the long-run level in a fashion similar to Dornbusch (1976) in the context of exchange rate dynamics.

Moreover, this result offers a reason why it is a natural benchmark for the central bank to purchase the entire new bond issuance in crisis periods. The key is to modulate how much debt the liquidity traders have to absorb. Suppose the bond investors remain completely inactive. Then, a central bank bond purchase equal to the net issuance amount prevents the liquidity traders from absorbing additional bond positions and charging a higher discount rate. When the fundamental of the bond deteriorates and the bond investors are also reducing their bond positions, the central bank needs to purchase more than the issuance amount to absorb the additional supply from the market and prevent the bond price from going down, a pattern we observed during the Covid-19 crisis period.

Isolate the Wealth Effect. We find it useful to additionally consider a case which mutes the wealth effect, which is presented in Appendix C.1 in detail. In this case, we assume that the bond investors' wealth is fixed at the steady-state level, which allows us to isolate the direct effect of the supply shock from the amplifying effect due to the its impact on the long-term bond investors' wealth and hence their risk-taking capacity.

Consider the same transitory bond supply shock that reduces the quantity of the publicly available long-term bond. As the central bank offloads the purchased bond in the QT phase, the bond price drops and lowers the bond investors' wealth due to their portfolio exposure. This wealth effect in turn reduces the bond investors' risk-taking capacity and their demand for the bond, which further lowers the bond price and creates a negative feedback loop. We find this wealth effect to be quantitatively important in our model. As shown in Figure (C.6), it explains over half of the bond

price drop in the QT phase. Interestingly, the wealth effect is also asymmetric between the QE and QT phases: the bond price increase in the QE phase is similar in magnitude with and without the wealth effect, as it is driven by the direct effect of the supply shock. In contrast, the bond price drop in the QT phase is much larger with the wealth effect.

The importance of this wealth effect reminds us of the banking crisis in early 2023, which was accompanied by mark-to-market losses on the U.S. banks' bond portfolios due to rising interest rates. The wealth effect we highlight in this section suggests that the reduction in the long-term investors' risk-taking capacity could also be a main contributor to the long-term bond price decline in this episode, as it can make the bond price even more sensitive to shocks.

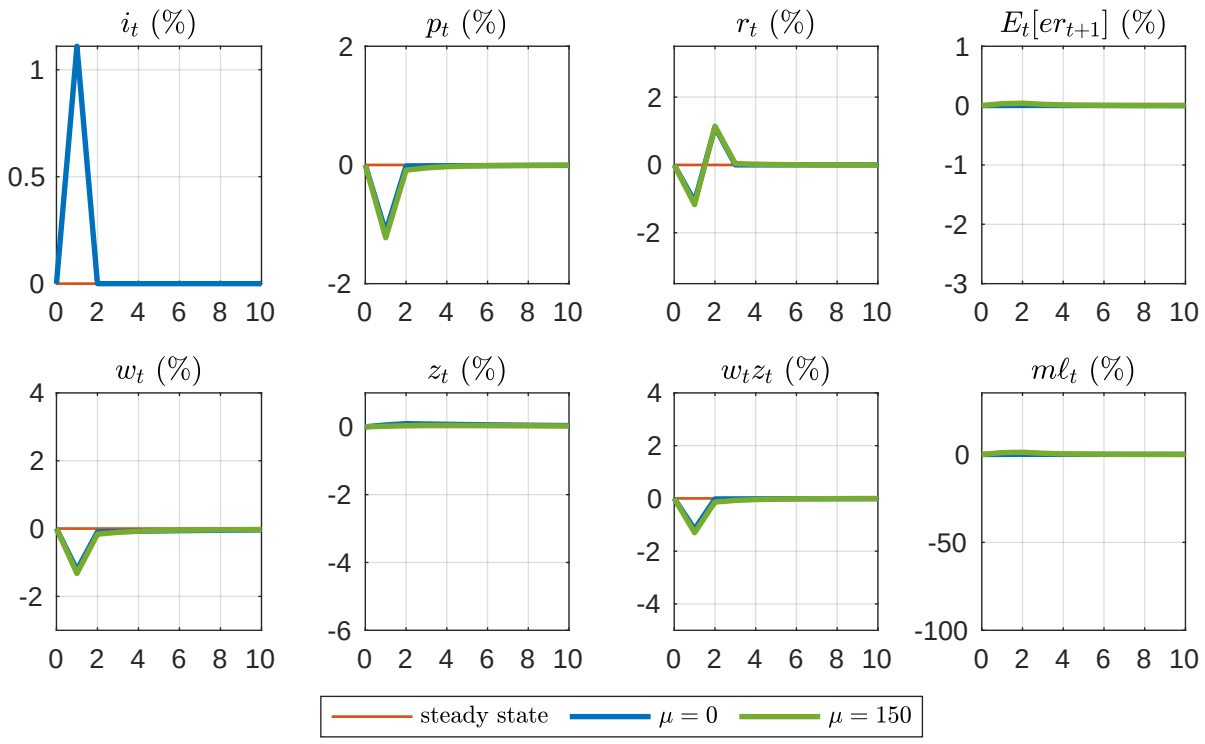


Figure 7: Impulse Response to a Transitory Interest Rate Shock

Notes: This figure reports the responses of the key state variables to a transitory shock to the interest rate. We consider two cases: with and without transaction costs.

The Interest Rate Shock. Figure (C.7) plots the impulse responses to a transitory increase in the interest rate on the reserve i_t in period 1, which returns to the steady-state value in period 2. We find that the two model variants exhibit very similar dynamics. An increase in the interest rate i_t in period 1 reduces the bond's price p_t and the cum-coupon return r_t in period 1, which in turn decreases the wealth of bond investors as well as the dollar value $w_t z_t$ of the bond asset. Since the bond investors' share of bonds does not change significantly given the transitory interest rate shock, it does not lead to a change in the position of liquidity traders, and therefore does not alter the risk premium demanded by the liquidity traders.

In period 2, as the interest rate shock dissipates, the bond price recovers and returns very close to the steady-state level. The bond return turns positive, and the bond investors' wealth also recovers. Consequently, the interest rate shock leaves a very small impact on the bond investors' wealth from period 2 onwards. This is a key difference between the response to the interest rate shock in Figure (C.7) and the response to the cash flow shock in Figure (C.5). The positive cash flow shock persistently raises bond investors' wealth after the shock dissipates. As a result, the cash flow shock has persistent effects on the equilibrium quantity of bonds held by the liquidity traders and therefore on the required return on bonds, whereas the interest rate shock does not.

4.3 Simulated Moments

We further illustrate the asset pricing properties of our model by studying its simulated moments. Our model incorporates three exogenous shocks ($\varepsilon_{x,t+1}$, $\varepsilon_{q,t+1}$, and $\varepsilon_{i,t+1}$), which impact the bond's cash flow, supply, and the reserve's interest rate, respectively. We begin by turning on only one exogenous shock at a time, and report the simulated moments in each case in Columns (1)–(3) of Table (2). In each case, we simulate the model for 100,000 periods. In this simulation, we assume that the bond's cash flow, supply, and the interest rate on the reserve are persistent with an AR(1) coefficient of 0.7.

In Column (1), we consider the bond's cash flow shock alone. This shock generates a bond return volatility of 4% per annum if the bond investors face no transaction cost, and 7% per annum

Table 2: Simulated Moments with Persistent Shocks ($\theta = 0.7$)

Moments	(1) x Shock	(2) q Shock	(3) i Shock	(4) All Shocks	(5) Data
Panel A: Model with no transaction cost $\mu = 0$*:*					
Vol(er_t) (%)	4.08	0.40	2.23	4.65	11.63
Autocorr(er_t)	-0.01	-0.16	-0.00	-0.01	0.05
Vol(i_t) (%)	0.00	0.00	1.11	1.11	1.11
β_{IE}	-0.00	-0.00	0.35	0.33	0.44
β_{RP}	0.72	8.69	0.00	0.05	1.76
Panel B: Model with transaction cost $\mu = 150$					
Vol(er_t) (%)	7.38	5.14	2.05	9.18	11.63
Autocorr(er_t)	-0.09	-0.27	-0.02	-0.14	0.05
Vol(i_t) (%)	0.00	0.00	1.11	1.11	1.11
β_{IE}	0.00	-0.00	0.34	0.27	0.44
β_{RP}	3.59	13.94	0.03	1.21	1.76

Notes: This table reports the simulated moments in the presence of each exogenous shock, including the volatility and the autocorrelation of the bond excess return, the volatility of the interest rate, and predictability of interest rate and bond excess return by the bond yield spread. We assume that the bond cash flow, bond supply, and interest rate processes are persistent with an AR(1) coefficient of 0.7. We consider two cases: with and without transaction costs.

if they face a transaction cost of $\mu = 150$. This higher return volatility is consistent with our earlier observation that the transaction cost makes the bond's return more sensitive to the cash flow shock.

In this simulated sample, we also consider two standard regressions involving the bond's yield spread, defined as $y_t - i_t$. First, we regress the change in the interest rate on the reserve on the yield spread:

$$i_{t+1} - i_t = \alpha_{IE} + \beta_{IE}(y_t - i_t) + u_{IE,t+1}. \tag{7}$$

This regression captures the extent to which the long-term bond yield prices in the expected path of future short-term interest rates. We refer to the regression coefficient β_{IE} as the *interest rate expectations* (IE) coefficient. Under the expectation hypothesis, β_{IE} should be positive. However, in this case with no variation in the short rate, we find that β_{IE} is zero.

Moreover, we regress the bond's excess return on the yield spread:

$$er_{t+1} = r_{t+1} - i_t = \alpha_{RP} + \beta_{RP}(y_t - i_t) + u_{RP,t+1}. \tag{8}$$

This regression captures the extent to which the long-term bond yield forecasts the bond's future excess return (Fama and Bliss, 1987). We refer to the regression coefficient β_{RP} as the *return predictability* (RP) coefficient. Consistent with the finding in the literature, the yield spread captures the bond's risk premium, and we find a positive β_{RP}. Moreover, in the presence of the transaction cost, β_{RP} is higher, implying a greater time variation in the bond's risk premium.

In Column (2), we consider the bond's supply shock alone. In this case, the transaction cost raises the bond return volatility from 0.4% to 5% per annum, which demonstrates the importance of the transaction cost in amplifying the bond's return volatility. The supply shock also generates a very large RP coefficient, suggesting that the central bank's bond supply policy can be very effective in influencing the bond's pricing and risk premium.

In Column (3), we consider the interest rate on the reserve shock alone. Consistent with the impulse response plots we saw in the last section, the interest rate shock only has a very weak effect on the bond's risk premium. As we discussed above, this is because the interest rate shock does not generate a strong and persistent wealth effect. On the other hand, the interest rate shock generates variations in the short rate, which is priced in by the long-term bond yield and generates a positive IE coefficient.

Finally, in Column (4), we consider all three shocks together. In the presence of transaction costs, the bond return volatility is 9% per annum, which is closer to the real bond return volatility of 11% than in the case without transaction costs. In Column (5), using U.S. Treasury data from January 2000 to September 2023 and focusing on long-term bonds with maturities above 10 years, we find that the IE coefficient is 0.44 and the RP coefficient is 1.76. With transaction costs, our model generates a positive IE coefficient of 0.27 and a positive RP coefficient of 1.21, but the model without transaction costs fails to match the RP coefficient.

In the case with all shocks turned on, we also examine the relationship between the bond's risk premium and the bond investors' risk-bearing capacity as measured by their wealth. Our model shares the feature with many financial intermediation models that the net worth of the more efficient holders of the risky asset—which are the bond investors in our model—emerge as an important state variable that measures the aggregate risk-bearing capacity. When their net worth declines, their risk-bearing capacity becomes more limited, which leads to a higher risk premium on the risky asset (Adrian et al., 2014; He et al., 2017).

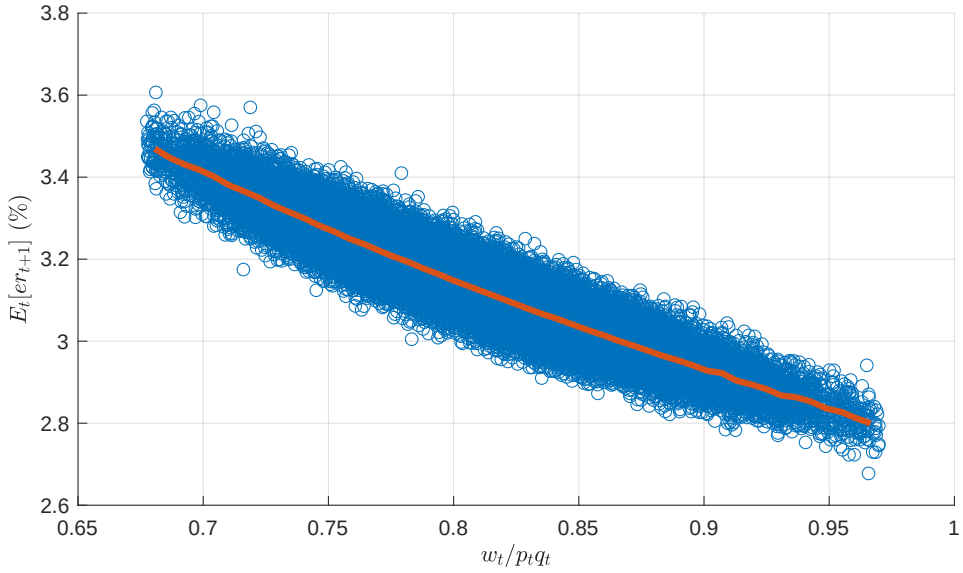


Figure 8: Bond Expected Excess Return v.s. Bond Investor Wealth over Bond Market Size.

Notes: The figure reports the distribution of the bond's expected excess return $\mathbb{E}_t[er_{t+1}]$ in relation to the bond investors' wealth divided by the bond market size, i.e., $w_t/(p_tq_t)$, under the presence of portfolio adjustment frictions.

To show this point, Figure (8) reports the distribution of the bond's expected excess return $\mathbb{E}_t[er_{t+1}]$ as a function of the bond investors' wealth-to-bond market cap ratio, i.e., $w_t/(p_tq_t)$. We use the bond market cap to normalize the bond investors' wealth because the bond supply is time-varying due to the central bank's purchases and sales. We find a strongly negative relationship, which confirms the intuition that the bond's risk premium is higher when the bond investors' risk-bearing capacity is more limited.

5 Debt Supply Management

Our bond pricing model sheds light on the effects of debt supply shocks including quantitative easing and tightening. In this section, we consider two forms of debt supply management, which are (1) a surprise central bank bond purchase that unwinds slowly over time, and (2) an announcement of central bank bond purchase that takes place in a later period. These two additional scenarios shed light on the effects of the central bank's gradualism and forward guidance, respectively.

5.1 Gradual Tightening in the Presence of Slow-Moving Capital

Quantitative easing has become an important tool to support the bond price and lower the long-term interest rate amidst recessions and crises. Our results above suggest that a transitory quantitative easing that unwinds too quickly generates an overshooting effect that worsens the price drop once the supply shock dissipates. In this section, we discuss whether a more gradual tightening strategy generates different effects.

We consider a simple AR(1) family of bond supply shocks, which corresponds to different θ_q coefficients in the specification of the exogenous supply process, reproduced below,

$$\log q_{t+1} = (1-\theta_q)\log \bar{q} + \theta_q \log q_t + \sigma_q \varepsilon_{q,t+1}.$$

We choose $\theta_q = \{0, 0.5, 0.80, 0.90, 0.95\}$, which implies half-lives of $\{0, 1, 3.1, 6.6, 13.5\}$ years. Figure (9) reports the impulse responses under transaction costs (i.e., $\mu = 150$). We find that, as we make the bond supply shock more persistent, the increase in period-1 bond price becomes stronger, and the overshooting effect in subsequent periods becomes weaker. Both patterns can be explained by the following intuition: as the central bank quits quantitative easing more gradually, the bond investors have more time to adjust their portfolio holdings and hence behave more similarly to the frictionless case. As a result, the liquidity traders are not required to absorb more bonds from period 2 onwards, leading to a persistent decrease in the bond's discount rate which raises the bond price across all periods. Under our calibration, when the persistence parameter is

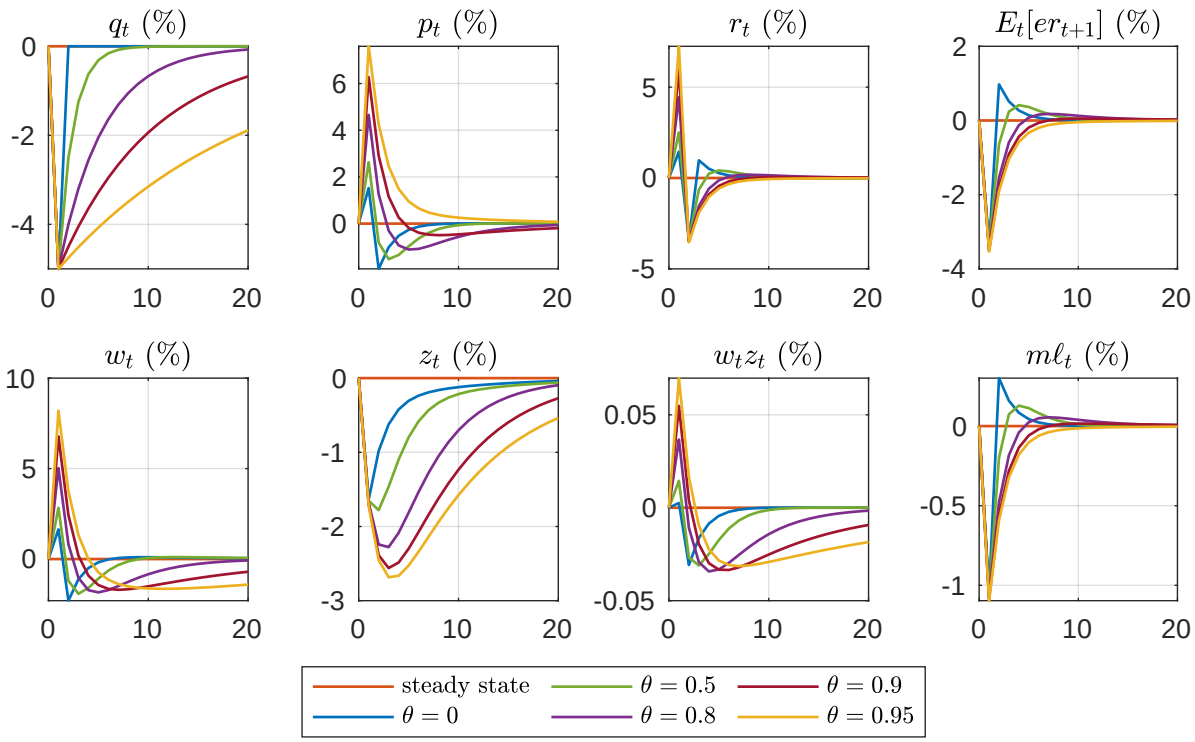


Figure 9: Impulse Response to Bond Supply Shocks with Different Persistence

Notes: This figure reports the responses of the key state variables to a family of bond supply shocks. These shocks are characterized by different AR(1) coefficients for the exogenous bond supply process.

$\theta_q = 0.95$, which implies a half-life of 13.5 years, the bond price never drops below its steady-state level, implying that the quantitative easing only has positive effects on the bond price.

Does the Fed conduct QT slow enough to be consistent with this case? Our empirical analysis suggests that the liquidity traders' bond positions increased dramatically and much more than other investors after QT announcements, which is inconsistent with a persistent decline in their bond positions as predicted by our model with more persistent QT. As such, QT in the U.S. may have been conducted faster than the long-term bond investors' portfolio adjustment speed, which could have generated the overshooting effect.

Moreover, we also note that, while a persistent supply shock keeps the bond price above the steady-state value throughout the QT episode, it is not a free lunch: it also lowers the bond in-

vestors' wealth w_t for an extended period of time. This is because this shock supports bond price by persistently lowering the realized returns from holding the bond. Since the bond investors hold the bond with a leverage, their wealth is persistently depressed. As such, if they are also active marginal investors in other asset markets such as the corporate bond market, then, this wealth effect could reduce their risk-taking capacity and raise the risk premium in other asset markets.

5.2 Impulse Responses to Shocks to Expectations

We also consider a shock to the expectation of the bond supply, which could be interpreted as a form of forward guidance. To make the figures easier to read, we assume that the news arrives in period 1 which announces a change in the bond supply in period 5. We report the impulse responses in Figure (10). For simplicity, we only include the bond quantity q_t, ex-coupon price p_t, expected excess return $\mathbb{E}_t[er_{t+1}]$, and the bond investors' portfolio share in the bond z_t. We report the full impulse responses in Appendix C.2.

In period 5, when the shock is realized, the bond quantity declines by 5% and, as the investors have downward-sloping demand curves, the bond price goes up. In anticipation of the higher price in period 5, the bond price also increases once the news arrives in period 1. In other words,

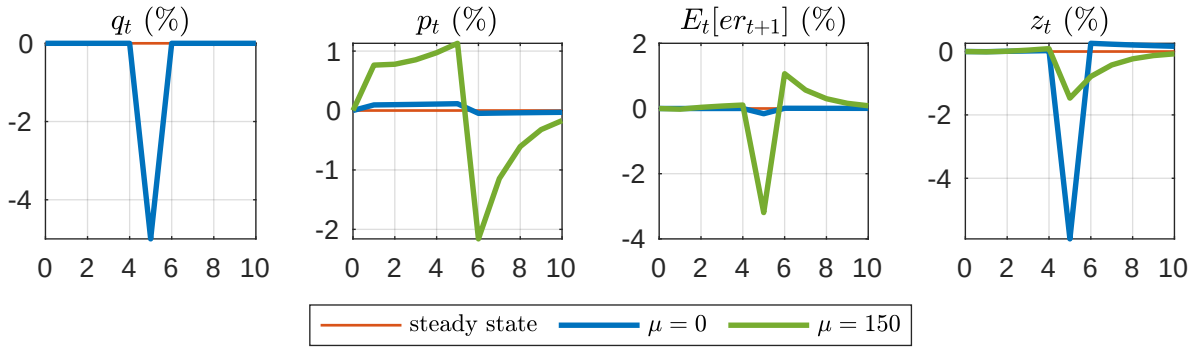


Figure 10: Impulse Response to an Expected Bond Supply Shock

Notes: This figure reports the responses of the key state variables to a transitory shock in the expectation of bond supply. We examine two cases: with and without transaction costs.

expectation of future central bank bond purchases can also support the bond price up front.

This expected quantity shock also generates the overshooting pattern that we discussed in the benchmark case in Figure (6). In the presence of portfolio friction (i.e., $\mu = 150$), as the bond quantity reverts to the steady-state level in period 6, the bond investors' portfolio share z_t remains low. As a result, the liquidity traders have to absorb additional bond shares and require a higher discount rate, leading to a subsequent drop in the bond price. In this case, even though the bond investors have prior knowledge of the scheduled central bank bond sale, they cannot adjust their portfolio gradually towards the desired position in advance. This is because the required adjustment in QT period 6 is relative to the change in the portfolio share in QE period 5. Before the QE happens and lowers the bond investors' portfolio share, their portfolio share stays close to the desired level and has no reason to adjust.

6 Conclusion

This paper develops a simple framework for understanding bond price dynamics in the presence of slow-moving capital. It highlights a novel overshooting effect that negatively impacts the long-term bond price when quantitative tightening is conducted too fast. As a result, due to the portfolio friction, quantitative tightening is not simply a reversal of quantitative easing. Given the central role of the long-term bond market in determining other assets' discount rates, these dynamic effects could have far-reaching implications for asset prices and real outcomes.

References

Andrew B Abel, Janice C Eberly, and Stavros Panageas. Optimal inattention to the stock market. American economic review, 97(2):244–249, 2007.

Andrew B Abel, Janice C Eberly, and Stavros Panageas. Optimal inattention to the stock market with information costs and transactions costs. Econometrica, 81(4):1455–1481, 2013.

Viral V Acharya, Rahul Singh Chauhan, Raghuram G Rajan, and Sascha Steffen. Liquidity dependence: Why shrinking central bank balance sheets is an uphill task. Available at SSRN, 2022.

Tobias Adrian, Erkko Etula, and Tyler Muir. Financial intermediaries and the cross-section of asset returns. The Journal of Finance, 69(6):2557–2596, 2014.

Fernando Alvarez, Luigi Guiso, and Francesco Lippi. Durable consumption and asset management with transaction and observation costs. American Economic Review, 102(5):2272–2300, 2012.

John Ameriks and Stephen P Zeldes. How do household portfolio shares vary with age? SSRN, 2011.

Philippe Bacchetta and Eric van Wincoop. Infrequent portfolio decisions: A solution to the forward discount puzzle. American Economic Review, 100(3):870–904, June 2010.

Philippe Bacchetta and Eric van Wincoop. Puzzling exchange rate dynamics and delayed portfolio adjustment. Journal of International Economics, 131:103460, 2021.

Philippe Bacchetta, Margaret Davenport, and Eric Van Wincoop. Can sticky portfolios explain international capital flows and asset prices? Journal of International Economics, 136:103583, 2022a.

Philippe Bacchetta, Eric van Wincoop, and Eric R Young. Infrequent random portfolio decisions in an open economy model. Review of Economic Studies, 90(3):1125–1154, 08 2022b.

Philippe Bacchetta, Simon Tieche, and Eric van Wincoop. International portfolio choice with frictions: Evidence from mutual funds. Review of Financial Studies, 04 2023.

Ayelen Banegas, Phillip J Monin, and Lubomir Petrasek. Sizing hedge funds' treasury market activities and holdings. 2021.

Michael D Bauer and Christopher J Neely. International channels of the fed's unconventional monetary policy. Journal of International Money and Finance, 44:24–46, 2014.

Yannis Bilias, Dimitris Georgarakos, and Michael Haliassos. Portfolio inertia and stock market fluctuations. Journal of Money, Credit and Banking, 42(4):715–742, 2010.

Johannes Breckenfelder and Veronica De Falco. Investor heterogeneity and large-scale asset purchases. Technical report, 2023.

Markus K Brunnermeier and Stefan Nagel. Do wealth fluctuations generate time-varying risk aversion? micro-evidence on individuals' asset allocation. American Economic Review, 98(3): 713–736, 2008.

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

Markus K Brunnermeier and Yuliy Sannikov. The i theory of money. Technical report, National Bureau of Economic Research, 2016.

Laurent E Calvet, John Y Campbell, and Paolo Sodini. Fight or flight? portfolio rebalancing by individual investors. The Quarterly journal of economics, 124(1):301–348, 2009.

YiLi Chien, Harold Cole, and Hanno Lustig. A multiplier approach to understanding the macro implications of household finance. The Review of Economic Studies, 78(1):199–234, 2011.

YiLi Chien, Harold Cole, and Hanno Lustig. Is the volatility of the market price of risk due to intermittent portfolio rebalancing? American Economic Review, 102(6):2859–2896, 2012.

Jens H. E. Christensen and Glenn D. Rudebusch. The response of interest rates to us and uk quantitative easing. The Economic Journal, 122(564):F385–F414, 10 2012.

Adam Copeland, Darrell Duffie, and Yilin Yang. Reserves were not so ample after all. Technical report, National Bureau of Economic Research, 2021.

Antonio Coppola. In safe hands: The financial and real impact of investor composition over the credit cycle. 2022.

Stefania D'Amico and Tim Seida. Unexpected supply effects of quantitative easing and tightening. The Economic Journal, 134(658):579–613, 2024.

William Diamond, Zhengyang Jiang, and Yiming Ma. The reserve supply channel of unconventional monetary policy. Jacobs Levy Equity Management Center for Quantitative Financial Research Paper, 2020.

Rudiger Dornbusch. Expectations and exchange rate dynamics. Journal of Political Economy, 84 (6):1161–1176, 1976.

Wenxin Du, Benjamin Hébert, and Wenhao Li. Intermediary balance sheets and the treasury yield curve. Journal of Financial Economics, 150(3):103722, 2023.

Wenxin Du, Kristin Forbes, and Matthew N Luzzetti. Quantitative tightening around the globe: What have we learned? Technical report, National Bureau of Economic Research, 2024.

Darrell Duffie. Presidential address: Asset price dynamics with slow-moving capital. Journal of Finance, 65(4):1237–1267, 2010.

Darrell Duffie and Tong-sheng Sun. Transactions costs and portfolio choice in a discrete-continuous-time setting. Journal of Economic dynamics and control, 14(1):35–51, 1990.

Andrea L Eisfeldt, Hanno Lustig, and Lei Zhang. Complex asset markets. Technical report, National Bureau of Economic Research, 2017.

Egemen Eren, Andreas Schrimpf, and Fan Dora Xia. The demand for government debt. Available at SSRN 4466154, 2023.

Eugene F. Fama and Robert R. Bliss. The information in long-maturity forward rates. The American Economic Review, 77(4):680–692, 1987.

Xiang Fang, Bryan Hardy, and Karen K Lewis. Who holds sovereign debt and why it matters. Technical report, National Bureau of Economic Research, 2022.

Xavier Gabaix and Matteo Maggiori. International liquidity and exchange rate dynamics. The Quarterly Journal of Economics, 130(3):1369–1420, 2015.

Nicolae Gârleanu and Lasse Heje Pedersen. Dynamic trading with predictable returns and transaction costs. Journal of Finance, 68(6):2309–2340, 2013.

Nicolae Gârleanu and Lasse Heje Pedersen. Dynamic portfolio choice with frictions. Journal of Economic Theory, 165:487–516, 2016.

Jonathan Glicoes, Benjamin Iorio, Phillip Monin, and Lubomir Petrasek. Quantifying treasury cash-futures basis trades. 2024.

Pierre-Olivier Gourinchas, Walker D Ray, and Dimitri Vayanos. A preferred-habitat model of term premia, exchange rates, and monetary policy spillovers. Technical report, National Bureau of Economic Research, 2022.

Robin Greenwood and Dimitri Vayanos. Bond Supply and Excess Bond Returns. Review of Financial Studies, 27(3):663–713, 01 2014.

Robin Greenwood, Samuel G Hanson, and Dimitri Vayanos. Forward guidance in the yield curve: short rates versus bond supply. Series on Central Banking Analysis and Economic Policies No. 24, 2016.

Robin Greenwood, Samuel G Hanson, and Gordon Y Liao. Asset price dynamics in partially segmented markets. Review of Financial Studies, 31(9):3307–3343, 2018.

Robin Greenwood, Samuel G Hanson, Jeremy C Stein, and Adi Sunderam. A quantity-driven theory of term premia and exchange rates. Technical report, National Bureau of Economic Research, 2020.

Richard Grinold. A dynamic model of portfolio management. Journal of Investment Management, 4(2):5, 2006.

Valentin Haddad and Tyler Muir. Do intermediaries matter for aggregate asset prices? Journal of Finance, 76(6):2719–2761, 2021.

Zhiguo He and Arvind Krishnamurthy. Intermediary asset pricing. American Economic Review, 103(2):732–770, 2013.

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

John Heaton and Deborah J Lucas. Evaluating the effects of incomplete markets on risk sharing and asset pricing. Journal of Political Economy, 104(3):443–487, 1996.

Zhengyang Jiang, Hanno Lustig, Stijn Van Nieuwerburgh, and Mindy Z Xiaolan. The us public debt valuation puzzle. Technical report, National Bureau of Economic Research, 2019.

Zhengyang Jiang, Arvind Krishnamurthy, Hanno N Lustig, and Jialu Sun. Beyond incomplete spanning: Convenience yields and exchange rate disconnect. 2021.

Zhengyang Jiang, Robert J Richmond, and Tony Zhang. A portfolio approach to global imbalances. Technical report, National Bureau of Economic Research, 2022a.

Zhengyang Jiang, Robert J Richmond, and Tony Zhang. Understanding the strength of the dollar. Technical report, National Bureau of Economic Research, 2022b.

Rohan Kekre, Moritz Lenel, and Federico Mainardi. Monetary policy, segmentation, and the term structure. 2022.

Ralph SJ Koijen and Motohiro Yogo. Exchange rates and asset prices in a global demand system. Technical report, National Bureau of Economic Research, 2020.

Ralph SJ Koijen, François Koulischer, Benoît Nguyen, and Motohiro Yogo. Inspecting the mechanism of quantitative easing in the euro area. Journal of Financial Economics, 140(1):1–20, 2021.

Nellie Liang. Opening remarks at New York Fed-ECB workshop on nonbank financial institutions. https://www.newyorkfed.org/research/conference/2024/ECB-FRBNY-NBFI-workshop, 2024. Accessed: 2024-07-09.

Mark Mitchell, Lasse Heje Pedersen, and Todd Pulvino. Slow moving capital. American Economic Review, 97(2):215–220, 2007.

Christopher J Neely. The large-scale asset purchases had large international effects. 2011.

James Tobin. Liquidity preference as behavior towards risk. Review of Economic Studies, 25(2): 65–86, 1958.

James Tobin. A general equilibrium approach to monetary theory. Journal of Money, Credit and Banking, 1(1):15–29, 1969.

Dimitri Vayanos and Jean-Luc Vila. A preferred-habitat model of the term structure of interest rates. Econometrica, 89(1):77–112, 2021.

Dimitri Vayanos and Paul Woolley. An institutional theory of momentum and reversal. Review of Financial Studies, 26(5):1087–1145, 2013.

Annette Vissing-Jorgensen and Arvind Krishnamurthy. The effects of quantitative easing on interest rates: Channels and implications for policy. Brookings Papers on Economic Activity, 43(2): 215–287, 2011.

Haonan Zhou. The fickle and the stable: Global financial cycle transmission via heterogeneous investors. Available at SSRN 4616182, 2023.

Appendix

A Data

A.1 Data Sources

Investor composition in the Treasury market is obtained from Federal Reserve Statistical Release, Tables L210: https://fred.stlouisfed.org/release/tables?rid=52&eid=809048.

Treasury bond return data are obtained from CRSP via Wharton Research Data Services: https://wrds-www.wharton.upenn.edu/.

FOMC transcripts are obtained from the Fed website: https://www.federalreserve.gov/mediacenter/files/FOMCpresconf20170920.pdf, https://www.federalreserve.gov/mediacenter/files/FOMCpresconf20220504.pdf.

A.2 FOMC Announcements of QT

2017 September FOMC Chair Yellen: *"... We also decided that in October we will begin the balance sheet normalization program that we outlined in June. This program will reduce our securities holdings in a gradual and predictable manner ... As I noted, the Committee announced today that it will begin its balance sheet normalization program in October. This program, which was described in the June addendum to our Policy Normalization Principles and Plans, will gradually decrease our reinvestments of proceeds from maturing Treasury securities and principal payments from agency securities. As a result, our balance sheet will decline gradually and predictably. For October through December, the decline in our securities holdings will be capped at $6 billion per month for Treasuries and $4 billion per month for agencies. These caps will gradually rise over the course of the following year to maximums of $30 billion per month for Treasuries and $20 billion per month for agency securities and will remain in place through the process of normalizing the size of our balance sheet. By limiting the volume of securities that private investors will have to absorb as we reduce our holdings, the caps should guard against outsized moves in interest rates*

and other potential market strains."

2019 May FOMC Chair Powell: *"Since October 2017, we have been allowing our asset holdings to decline by not reinvesting all of the payments we receive as securities matured or were prepaid. Today we announced that we intend to slow the runoff of our assets starting in May and to cease runoff entirely in September of this year."*

2019 July FOMC Chair Powell: *"We also decided to conclude the runoff of our securities portfolio in August rather than in September, as previously planned. And I'll discuss the thinking behind today's interest rate reduction and then turn to the path forward."*

Donna Borak: *" ... And could you also elaborate a little bit further on why the Fed decided to speed up its balance sheet runoff two months earlier today?"*

Chair Powell: *" ... In terms of the balance sheet, that was really just a matter of simplicity and consistency. Really nothing more to it than that."*

2019 October FOMC Chair Powell: *" ... In the transition to the efficient and effective level of reserves in this regime, we slowed the gradual decline in our balance sheet in March and we stopped it in July."*

2022 May FOMC Chair Powell: *"In addition, we are beginning the process of significantly reducing the size of our balance sheet... We also decided to begin the process of reducing the size of our balance sheet, which will play an important role in firming the stance of monetary policy.... With regard to our balance sheet, we also issued our specific plans for reducing our securities holdings. Consistent with the principles we issued in January, we intend to significantly reduce the size of our balance sheet over time in a predictable manner by allowing the principal payments from our securities holdings to roll off the balance sheet, up to monthly cap amounts. For Treasury securities, the cap will be $30 billion per month for three months and will then increase to $60 billion per month. The decline in holdings of Treasury securities under this monthly cap*

will include Treasury coupon securities and, to the extent that coupon securities are less than the monthly cap, Treasury bills. For agency mortgage-backed securities, the cap will be $17.5 billion per month for three months and will then increase to $35 billion per month. At the current level of mortgage rates, the actual pace of agency MBS runoff would likely be less than this monthly cap amount. Our balance sheet decisions are guided by our maximum- employment and price-stability goals. "

Michael McKee:*"The balance sheet: Why did you decide to wait until June 1st to begin letting securities roll off and not start – immediately start in the middle of this month, say, and do you have another - a newer or a better estimate for the monetary policy impact of letting the balance sheet decline?"*

Chair Powell: *"So why June 1? It was just pick a date, you know, and that happens to be – that happened to be the date that we picked; there was nothing magic about it. You know, it's not going to have any macroeconomic significance over time. We just picked that. Sometimes we publish these calendars on the first day of the month and that's what we're doing. I wouldn't read anything into it.*

In terms of the effect, I mean, I would just stress how uncertain the effect is of shrinking the balance sheet. You know, we run these models and everyone does in this field and make estimates of what will be the-how do you measure, you know, a certain quantum of balance sheet shrinkage compared to quantitative easing? And you know, these are very uncertain. I really can't be any clearer. There won't be any clearer-you know, people estimate that, broadly on the path we're on-and this will be taken probably too seriously but sort of one quarter percent-one rate increase over the course of a year at this pace but, I would just say, with very wide uncertainty bands, very wide. We don't really know. There are other estimates that are much smaller than that, by the way, and some of you may have read about that. That's kind of a mainstream estimate."

A.3 Hedge Fund Holdings in the Households Sector

Banegas et al. (2021) mention that, in Financial Accounts of the U.S., the "household sector in-

cludes domestic hedge funds." To validate this assertion, we conduct a comparison between the Treasury holdings of households and those of the hedge funds from Banegas et al. (2021). As shown in Figure (A.1), hedge funds both long and short U.S. Treasurys. Their short positions in the Treasury market have been quite stable, whereas their long positions increased by about $600 billion from 2017Q3 to 2019Q2. In comparison, Figure (A.2) shows that the household sector's Treasury holdings increased by $860 billion during the same period. As a result, we conclude that the hedge funds were the primary driver of the increase in the household sector's Treasury holdings.

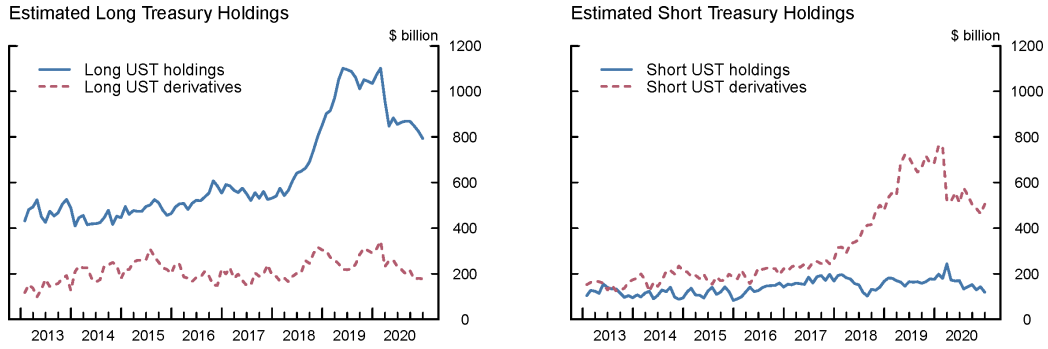


Figure A.1: Qualifying Hedge Funds' US Treasury Holdings

Notes: This figure refers to Figure 6 of Banegas et al. (2021), which reports the estimated long and short Treasury holdings and derivatives positions of qualifying hedge funds.

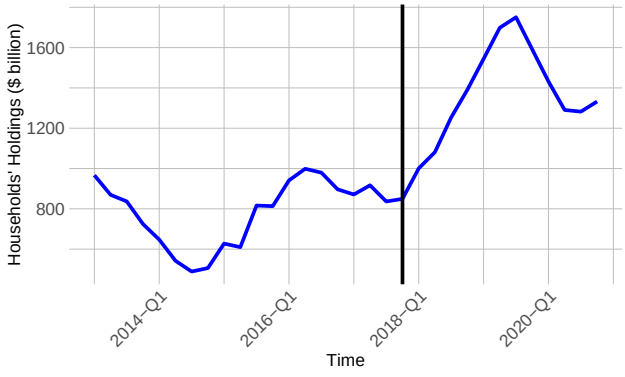


Figure A.2: Households' US Treasury Holdings

Notes: This figure reports the level of households' treasury holdings. The vertical black line marks the beginning of the first QT.

A.4 More Granular Breakdown of Investor Composition

Figure (A.3) plots the breakdown of liquidity traders' holdings scaled by the total supply. Households and US depository institutes hold the most Treasury within liquidity traders, followed by brokers and dealers.

Figure (A.4) illustrates the distribution of bond investors' holdings scaled by the total supply. Among bond investors, federal government retirement funds, mutual funds, and MMFs hold the largest shares of Treasuries, followed by private pension funds and local government retirement funds.

Figure (A.5) illustrates the breakdown of changes in liquidity traders' holdings around QT. During QT1, US banks exhibited the most significant reaction, although their absolute holdings were relatively small. Households, foreign banks, and brokers and dealers also remarkably increased their Treasury holdings. During QT2, brokers and dealers and households experienced the most significant increases in their holdings, while US banks had the most substantial increase within the first two quarters.

Figure (A.6) illustrates the breakdown of changes in bond investors' holdings as a percentage around QT. During QT1, holding companies, ABS issuers, and ETFs experienced the most significant increases in their Treasury holdings relative to their initial positions. However, the absolute sizes of their holdings were relatively small.

During QT2, bond investors responded quite diversely. Similar to QT1, holding companies had fast increases in their Treasury holdings within the first year, whereas property insurance companies, non-financial corporates, and ETFs also had notable increases by over or around 40% by 2023Q3.

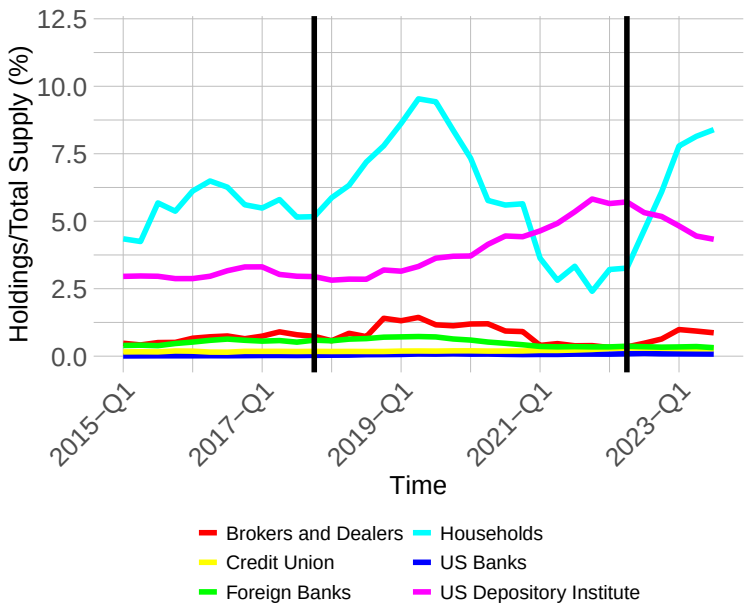


Figure A.3: Shares Held by Liquidity Traders

Notes: We report the share of investor holdings scaled by the aggregate supply of government debt for liquidity traders. The vertical black lines mark the beginning of the QT. The data are at quarterly frequency.

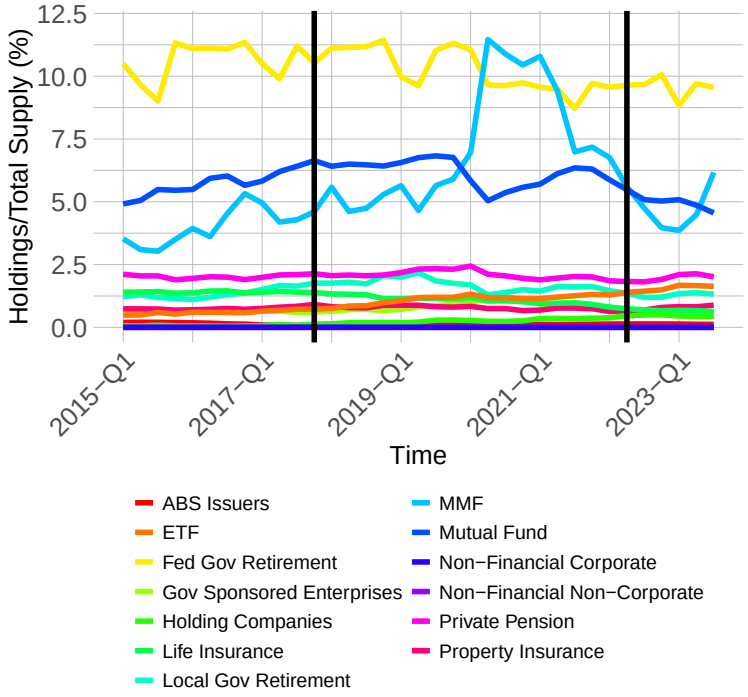


Figure A.4: Shares Held by Bond Investors

Notes: We report the share of investor holdings scaled by the aggregate supply of government debt for bond investors. The vertical black lines mark the beginning of the QT. The data are at quarterly frequency.

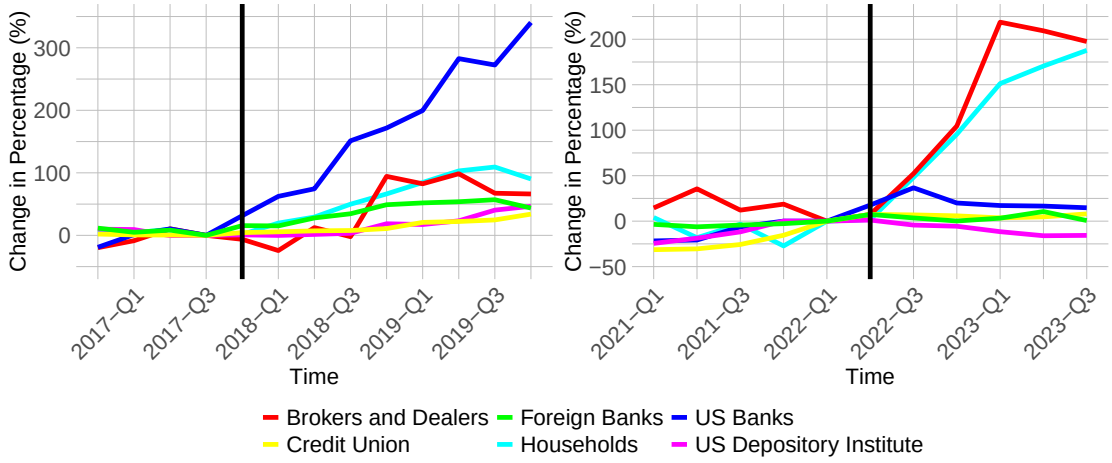


Figure A.5: Liquidity Traders Composition around QT Episodes

Notes: We report the breakdown of the changes in liquidity traders' Treasury holdings scaled by the initial quantities prior to the QT episodes. The left panel plots changes relative to 2017 Q3, and the right panel plots changes relative to 2022 Q1. The vertical black lines mark the beginning of the QT. The data are at quarterly frequency.

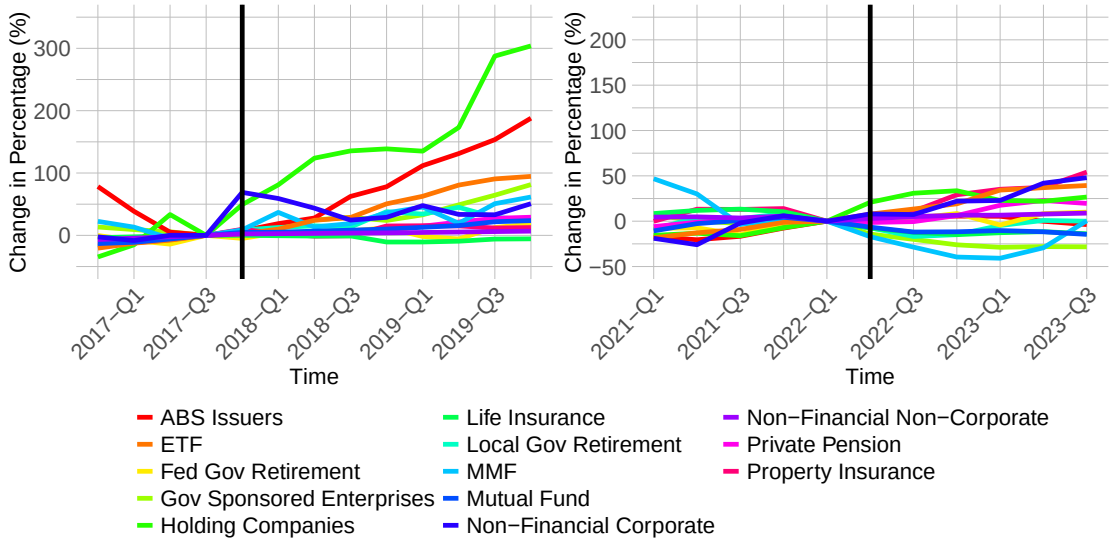


Figure A.6: Bond Investors Composition around QT Episodes

Notes: We report the breakdown of the changes in bond investors' Treasury holdings scaled by the initial quantities prior to the QT episodes. The left panel plots changes relative to 2017 Q3, and the right panel plots changes relative to 2022 Q1. The vertical black lines mark the beginning of the QT. The data are at quarterly frequency.

A.5 Investor Composition during QE

We also examine the change in investor composition during two QE episodes. The first one (QE1) spans from November 2008 to June 2013. On November 25, 2008, Ben Bernanke announced the initiation of a program aimed at purchasing the direct obligations of housing-related government-sponsored enterprises. Later, on June 19, 2013, Bernanke announced a gradual reduction, commonly referred to as "tapering," of some of the Fed's QE policies, contingent upon continued positive economic data. Within this time frame, there were three rounds of QE, commencing in November 2008, November 2010, and September 2012.

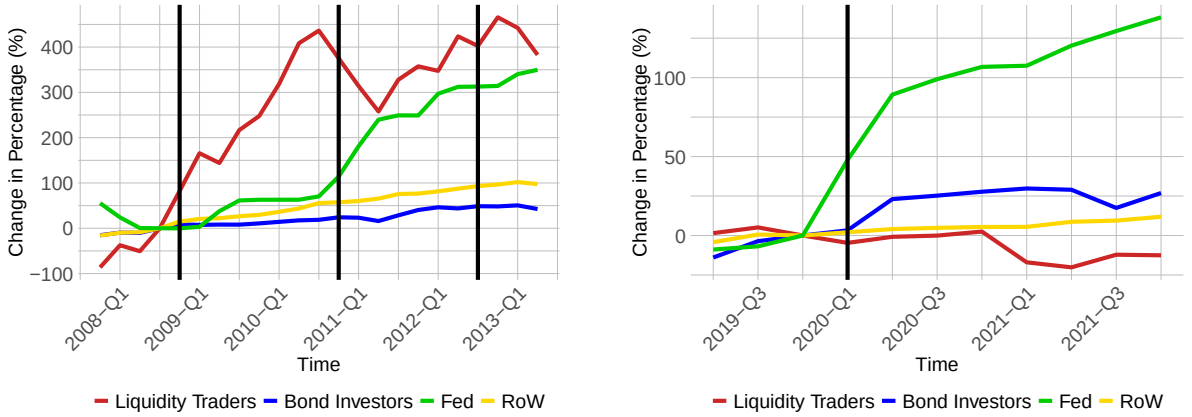
The second one (QE2) spans from March 2020 to May 2022, which precedes QT2 which we study in the main text.

Figure (A.7) plots the changes in the holdings by each investor category scaled by the size of their holdings in the quarter prior to the QE episodes. In both episodes, the Federal Reserve notably increased its Treasury holdings. However, while the liquidity traders lowered their holdings during QE2, they increased their holdings during QE1. To understand this difference, we further break down the holdings among different types of liquidity traders in Figure (A.8). The brokers and dealers lowered their holdings during both QE episodes.

However, the domestic household sector increased their holdings while the Fed was also buying bonds during QE1, whereas they decreased their holdings during QE2. While we do not have further breakdown of the transactions by the household sector, it is possible that the bond purchases during QE1 were not made by the hedge funds, but by other investor types categorized under the household sector which might be best not classified as liquidity traders. In comparison, as we discussed in Section A.3, data on the hedge funds' Treasury holdings during QE2 are available, and they are consistent with the household sector's holdings in this episode.

Overall, the brokers and dealers acted in a way consistent with our view of the liquidity traders, who absorbed the imbalances created by the Federal Reserve's purchases and sales. The household sector had different reactions in the two episodes, which might be due to different types of investors within this sector. It is also important to note that there was another major seller of the Treasury

bonds during both QE episodes: the government issuer itself. As such, the Fed's purchase of the Treasury bonds during QE do not necessarily mean that the private investors have to lower their holdings.



Panel (a): Changes Relative to 2008 Q4

Panel (b): Changes Relative to 2020 Q1

Figure A.7: Investor Composition around QE Episodes

Notes: We report the change in Treasury holdings scaled by the initial quantities prior to the QE episodes. The vertical black lines mark the beginning of the QE. The data are at quarterly frequency.

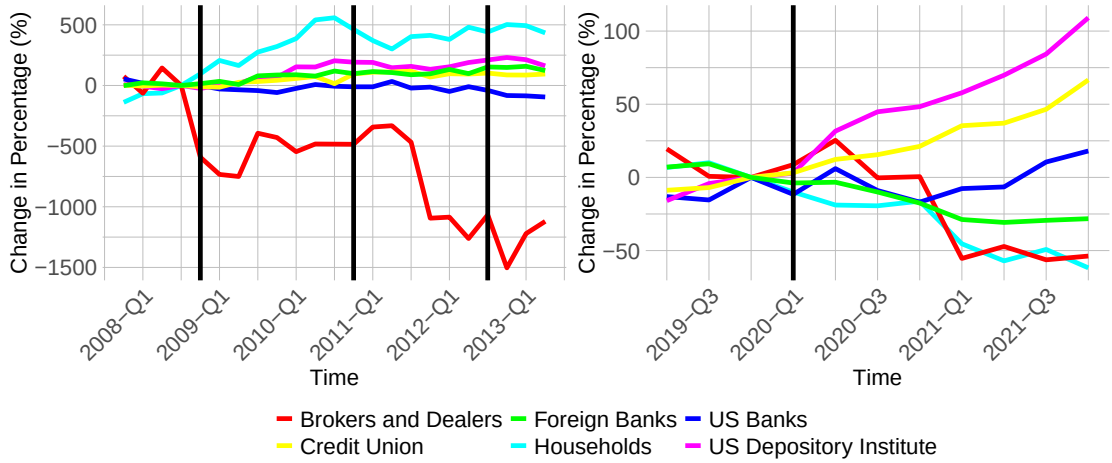


Figure A.8: Liquidity Traders Composition around QE Episodes

Notes: We report the change in Treasury holdings scaled by the initial quantities prior to the QE episodes. The vertical black lines mark the beginning of the QE. The data are at quarterly frequency.

B Proof

B.1 Bond Investor's Portfolio Choice Problem

Assuming bond investors have power utility over their wealth with a positive risk-aversion coefficient $\gamma \neq 1$, and that the bond cum-coupon holding period return r_t follows a log-normal distribution. In the absence of transaction costs and other financial frictions, the portfolio choice problem for bond investors can be formulated as follows:

$$\max_{z_t} \quad \mathbb{E}_t \left[\frac{w_{t+1}^{1-\gamma}}{1-\gamma} \right] \qquad \text{s.t.} \qquad w_{t+1} = w_t r_{t+1}^p$$

where $r_{t+1}^p = i_t + z_t(r_{t+1} - i_t)$. The above problem is equivalent to

$$\max_{z_t} \quad \mathbb{E}_t \left[\log r_{t+1}^p \right] + \frac{1}{2}(1-\gamma) var_t(\log(r_{t+1}^p)).$$

For the portfolio return $r_{t+1}^p = i_t + z_t(r_{t+1} - i_t)$, we have the approximation that

$$\begin{aligned}
\log(r_{t+1}^p) - \log(i_t) &= \log(i_t + z_t(r_{t+1} - i_t)) - \log(i_t) \\
&= \log(1 + z_t(r_{t+1}/i_t - 1)) \\
&= \log\left(1 + z_t\left(\exp\left(\log(r_{t+1}) - \log(i_t)\right) - 1\right)\right) \\
&\approx z_t\left(\log(r_{t+1}) - \log(i_t)\right) + \frac{1}{2} z_t\left(1 - z_t\right)\left(\log(r_{t+1}) - \log(i_t)\right)^2 \\
&\approx z_t\left(\log(r_{t+1}) - \log(i_t)\right) + \frac{1}{2} z_t\left(1 - z_t\right) var_t(\log(r_{t+1}))
\end{aligned}$$

where the approximate equality comes from a second-order Taylor approximation of the function $f(x) = \log(1 + z(\exp(x) - 1))$ around $x = 0$, using the fact that for short time intervals and log-normally distributed returns, $(\log(r_{t+1}) - \log(i_t))^2 \approx var_t(\log(r_{t+1}))$. Therefore, r_{t+1}^p is also

approximately log-normal, and

$$var_t(\log(r^p_{t+1})) \approx z_t^2 var_t(\log(r_{t+1})).$$

By the approximate log-normality of r^p_{t+1}, we have

$$\mathbb{E}_t\left[\log r^p_{t+1}\right] + \frac{1}{2} var_t(\log(r^p_{t+1})) = \log \mathbb{E}_t\left[r^p_{t+1}\right]$$

Therefore, bond investor's problem is equivalent to

$$\max_{z_t} \quad \log \mathbb{E}_t\left[r^p_{t+1}\right] - \frac{1}{2}\gamma var_t(\log(r^p_{t+1})).$$

When introducing portfolio adjustment frictions, we consider the bond investor's problem as follows:

$$\max_{z_t} \quad \mathbb{E}_t\left[\frac{w_{t+1}^{1-\gamma}}{1-\gamma}\right] \qquad \text{s.t.} \quad w_{t+1} = w_t r^p_{t+1} \exp\left(-\frac{1}{2}\mu \cdot c(z_t; z_{t-1}, \bar{z}) var_t(\log(r_{t+1}))\right)$$

where the function $c(z_t; z_{t-1}, \bar{z})$ captures the portfolio adjustment cost. Then, by the log-normality of r_{t+1}, bond investor's problem is equivalent to:

$$\max_{z_t} \quad \log \mathbb{E}_t\left[r^p_{t+1}\right] - \frac{1}{2}\gamma var_t(\log(r^p_{t+1})) - \frac{1}{2}\mu \cdot c(z_t; z_{t-1}, \bar{z}) var_t(\log(r_{t+1})),$$

which is consistent with the bond investor's problem in the main text after log-linearization.

B.2 Portfolio Return Approximation

Taking the logarithm of the bond's cum-coupon holding period return, we obtain:

$$\begin{aligned}\log r_{t+1} &= \log\left(p_{t+1} + x_{t+1}\right) - \log\left(p_t\right) \\ &= \log p_{t+1} - \log p_t + \log\left(1 + x_{t+1}/p_{t+1}\right)\end{aligned}$$

Define $f(z) = \log(1 + \exp(z))$. Then, we can linearize $\log\left(1 + x_{t+1}/p_{t+1}\right)$ as follows.

$$\begin{aligned}\log\left(1 + x_{t+1}/p_{t+1}\right) &= f\left(\log x_{t+1} - \log p_{t+1}\right) \\ &\approx f(\log\bar{x} - \log\bar{p}) + f'(\log\bar{x} - \log\bar{p})\left(\log x_{t+1} - \log p_{t+1} - (\log\bar{x} - \log\bar{p})\right) \\ &\approx f(\log\bar{x} - \log\bar{p}) + \frac{\exp(\log\bar{x} - \log\bar{p})}{1 + \exp(\log\bar{x} - \log\bar{p})}\left(\log x_{t+1} - \log p_{t+1} - (\log\bar{x} - \log\bar{p})\right)\end{aligned}$$

where $(\log\bar{x} - \log\bar{p})$ is the steady-state value of $\log x_{t+1} - \log p_{t+1}$.

Therefore, the resulting first-order approximation for the log return is

$$\log r_{t+1} \approx \kappa_0 + \kappa_1 \log p_{t+1} + (1 - \kappa_1)\log x_{t+1} - \log p_t, \tag{B.1}$$

where

$$\kappa_1 = \frac{1}{1 + \exp(\log\bar{x} - \log\bar{p})}, \tag{B.2}$$

$$\kappa_0 = -\log(\kappa_1) - (1 - \kappa_1)\log(1/\kappa_1 - 1). \tag{B.3}$$

Next, we take the logarithm of the portfolio return $r_t^p = i_{t-1} + z_{t-1}\left(r_t - i_{t-1}\right)$ and rearrange.

$$\begin{aligned}\log r_t^p - \log i_{t-1} &= \log\left(1 + z_{t-1}\left(\frac{r_t}{i_{t-1}} - 1\right)\right) \\ &= \log\left(1 + z_{t-1}\left(\exp\left(\log r_t - \log i_{t-1}\right) - 1\right)\right).\end{aligned}$$

Denote $g(z,x) = \log(1 + z(\exp(x) - 1))$. We linearize $g(z,x)$ around $z = \bar{z}$ and $x = 0$.

$$\begin{aligned} g(z,x) &= g(\bar{z},0) + g_x(\bar{z},0)x + g_{xx}(\bar{z},0)x^2 + g_z(\bar{z},0)(z-\bar{z}) + g_{zz}(\bar{z},0)(z-\bar{z})^2 \\ &\quad + g_{zx}(\bar{z},0)(x-0)(z-\bar{z}) + o\left(||(z-\bar{z},x)||^2\right) \\ &= \bar{z}x + \bar{z}(1-\bar{z})x^2 + (z-\bar{z})x + o\left(||(z-\bar{z},x)||^2\right). \end{aligned}$$

Therefore, linearizing the portfolio return around the portfolio share equal to $\bar{z}$ and the excess return equal to zero, we have

$$\begin{aligned} \log r_t^p - \log i_{t-1} &= g(z_{t-1}, \log r_t - \log i_{t-1}) \\ &= \bar{z}\left(\log r_t - \log i_{t-1}\right) + \bar{z}\left(\bar{z}-1\right)\left(\log r_t - \log i_{t-1}\right)^2 \\ &\quad + \hat{z}_{t-1}(\log r_t - \log i_{t-1}) + o\left(||(\hat{z}_{t-1}, \log r_t - \log i_{t-1})||^2\right) \end{aligned}$$

Similarly, for the steady-state, we have

$$\begin{aligned} \log \bar{r}^p - \log \bar{i} &= g(\bar{z}, \log \bar{r} - \log \bar{i}) \\ &= \bar{z}\left(\log \bar{r} - \log \bar{i}\right) + \bar{z}\left(\bar{z}-1\right)\left(\log \bar{r} - \log \bar{i}\right)^2 \\ &\quad + o\left(||(0, \log \bar{r} - \log \bar{i})||^2\right) \end{aligned}$$

Finally, by taking the difference between them, we have:

$$\begin{aligned} \hat{r}_t^p - \hat{i}_{t-1} \approx \bar{z}\left(\hat{r}_t - \hat{i}_{t-1}\right) + \hat{z}_{t-1}(\hat{r}_t - \hat{i}_{t-1}) + \hat{z}_{t-1}(\log \bar{r} - \log \bar{i}) \\ + \underbrace{\bar{z}\left(\bar{z}-1\right)\left(\left(\log r_t - \log i_{t-1}\right)^2 - \left(\log \bar{r} - \log \bar{i}\right)^2\right)}_{\text{Quadratic Term}} \end{aligned} \tag{B.4}$$

B.3 Proof of Proposition 1

First, we log-linearize the Campbell-Shiller decomposition of bond return in (B.1) around the steady-state values.

$$\hat{r}_{t+1} \approx \kappa_1 \hat{p}_{t+1} - \hat{p}_t + (1-\kappa_1)\hat{x}_{t+1} \tag{B.5}$$

where

$$\kappa_1 = \frac{1}{1+\exp(\log \bar{x} - \log \bar{p})},$$
$$\kappa_0 = -\log(\kappa_1) - (1-\kappa_1)\log(1/\kappa_1 - 1).$$

Second, we log-linearize the market clearing condition around the steady-state values.

$$\hat{p}_t + \hat{q}_t = \frac{\bar{w}\bar{z}}{\bar{p}\bar{q}}(\hat{w}_t + \frac{1}{\bar{z}}\hat{z}_t) + \frac{m\bar{\ell}}{\bar{p}\bar{q}}(\frac{1}{\bar{\ell}}\hat{\ell}_t) \tag{B.6}$$

Third, we log-linearize the law of motion of wealth around the steady-state values.

$$\hat{w}_t \approx \bar{r}^p(\phi\hat{w}_{t-1} + \phi\hat{r}_{p,t})$$

Then, by (B.4),

$$\begin{aligned}\hat{w}_t \approx & \bar{r}^p\phi\hat{w}_{t-1} + \bar{r}^p\phi\left(\hat{i}_{t-1} + \bar{z}\left(\hat{r}_t - \hat{i}_{t-1}\right) + \hat{z}_{t-1}(\hat{r}_t - \hat{i}_{t-1}) + \hat{z}_{t-1}(\log\bar{r} - \log\bar{i})\right) \\ & + \bar{r}^p\phi\underbrace{\left(\bar{z}\left(\bar{z}-1\right)\left(\left(\log r_t - \log i_{t-1}\right)^2 - \left(\log\bar{r} - \log\bar{i}\right)^2\right)\right)}_{\text{Quadratic Term}}\end{aligned} \tag{B.7}$$

Next, we linearize the optimal portfolio choice of the bond investor around the steady-state

values. The first-order condition of the bond investor's optimization problem is as follows:

$$\mathbb{E}_t[r_{t+1} - i_t] - \gamma z_t \sigma_t^2 - \mu \left(z_t - (\eta \bar{z} + (1-\eta) z_{t-1}) \right) \sigma_t^2 = 0. \tag{B.8}$$

Let $\bar{\sigma}$ and $\bar{er}$ represent the steady-state values of bond return volatility and excess return, respectively. We linearize the first-order condition around the steady-state values of the variables.

$$\mathbb{E}_t \left[er_{t+1} - \gamma \hat{z}_t \bar{\sigma}^2 - \gamma \bar{z} \sigma_t^2 - \mu \left(\hat{z}_t - (1-\eta) \hat{z}_{t-1} \right) \bar{\sigma}^2 \right] \approx 0$$

Rearranging the terms, we obtain

$$\hat{z}_t = \frac{\mu}{(\gamma + \mu)} (1-\eta) \hat{z}_{t-1} + \frac{1}{\bar{\sigma}^2 (\gamma + \mu)} \mathbb{E}_t[er_{t+1}] - \frac{1}{\bar{\sigma}^2 (\gamma + \mu)} \gamma \bar{z} \sigma_t^2 \tag{B.9}$$

In the steady-state, we have $\hat{z}_t = \hat{z}_{t-1} = 0$ and $\mathbb{E}_t[er_{t+1}] = \bar{er} = \gamma \bar{z} \bar{\sigma}^2$, so this equation is satisfied. Assuming that $\sigma_t = \bar{\sigma}$, we obtain

$$\hat{z}_t = \frac{\mu}{\gamma + \mu} (1-\eta) \hat{z}_{t-1} + \frac{\gamma}{\gamma + \mu} \frac{1}{\bar{\sigma}^2 \gamma} \mathbb{E}_t[\hat{r}_{t+1} - \hat{i}_t]. \tag{B.10}$$

Finally, let's linearize the portfolio choice of liquidity traders:

$$\hat{\ell}_t \;\; = \;\; \frac{1}{\bar{\sigma}^2 \omega} \mathbb{E}_t[\hat{r}_{t+1} - \hat{i}_t].$$

B.4 The Steady State

We assume that the steady-state values of dividend $\bar{x}$, bond quantity $\bar{q}$, bond investor's wealth $\bar{w}$, reserve rate $\bar{i}$, and bond return volatility σ are exogenously given. We need to solve for the

steady-state values of portfolio shares $\bar{z}$ and $\bar{\ell}$, bond return $\bar{r}$, and bond price $\bar{p}$ such that:

$$\begin{aligned}
\bar{r} &= 1 + \frac{\bar{x}}{\bar{p}} \\
\bar{p}\bar{q} &= \bar{w}\bar{z} + m\bar{\ell} \\
\bar{z} &= \frac{\bar{r} - \bar{i}}{\gamma\bar{\sigma}^2} \\
\bar{\ell} &= \frac{\bar{r} - \bar{i}}{\omega\bar{\sigma}^2}.
\end{aligned}$$

Notice that this is a system of quadratic equations with two groups of solutions as follows:

$$\begin{aligned}
\bar{z} &= \frac{(1-\bar{i})}{2\gamma\bar{\sigma}^2} \pm \sqrt{\left(\frac{1-\bar{i}}{2\gamma\bar{\sigma}^2}\right)^2 + \frac{\bar{x}\bar{q}}{\left(\bar{w} + m\frac{\gamma}{\omega}\right)\gamma\bar{\sigma}^2}} \\
\bar{\ell} &= \frac{\gamma}{\omega}\bar{z} \\
\bar{p} &= \frac{\bar{w}\bar{z} + m\bar{\ell}}{\bar{q}} \\
\bar{r} &= 1 + \frac{\bar{x}}{\bar{p}}
\end{aligned}$$

One set of solutions corresponds to a positive holding share $\bar{z}$ and a positive bond return $\bar{r}$, while the other set of solutions corresponds to a negative holding share $\bar{z}$ and a negative bond return $\bar{r}$. We select the positive solution.

C Additional Theoretical Results

C.1 Isolate the Wealth Effects

In our setting with transaction costs ($\mu > 0$), we consider an additional case in which the bond investors' wealth is held constant at the steady-state level. This exercise allows use to decompose the bond price dynamics into two components: the direct effect from the shock and the wealth effect due to the responses in the bond investors' wealth. Technically, we simply assume that bond investors of each generation t receive a fixed endowment $\bar{w}$ and consume all their proceeds, so that the deviation in the wealth of bond investors is $\hat{w}_t = 0$.

The Cash Flow Shock. We decompose the impulse responses to a positive shock to the bond's cash flow, x_t, in Figure (C.1). When bond investors' wealth remains constant, the transitory cash flow only causes the bond return to jump in period 1, and nothing else. In other words, the discount rate effect we found in the main text is entirely due to the effect of the cash flow shock on the bond investors' wealth.

The Bond Quantity Shock. Next, we decompose the impulse responses to a transitory reduction in bond quantity q_t available to the public in period 1 in Figure (C.6). The impacts of the quantity shock on bond price and portfolio positions in period 1 are similar with and without the wealth effect. However, when the quantity supply shock unwinds in period 2, if the bond investors' wealth are replenished every period, the financial distress experienced by bond investors due to their wealth decline from the drop in bond price is alleviated. In this case, bond investors have more wealth to hold more bond positions, and liquidity traders need to absorb a lower quantity of bonds sold by bond investors. This leads to a lower premium demanded by liquidity traders and hence a smaller price drop in period 2, relative to the case with time-varying wealth.

The Interest Rate Shock. In Figure (C.3), we decompose the impulse responses to a transitory interest rate shock. When the bond investors' wealth responds to portfolio returns, an increase in

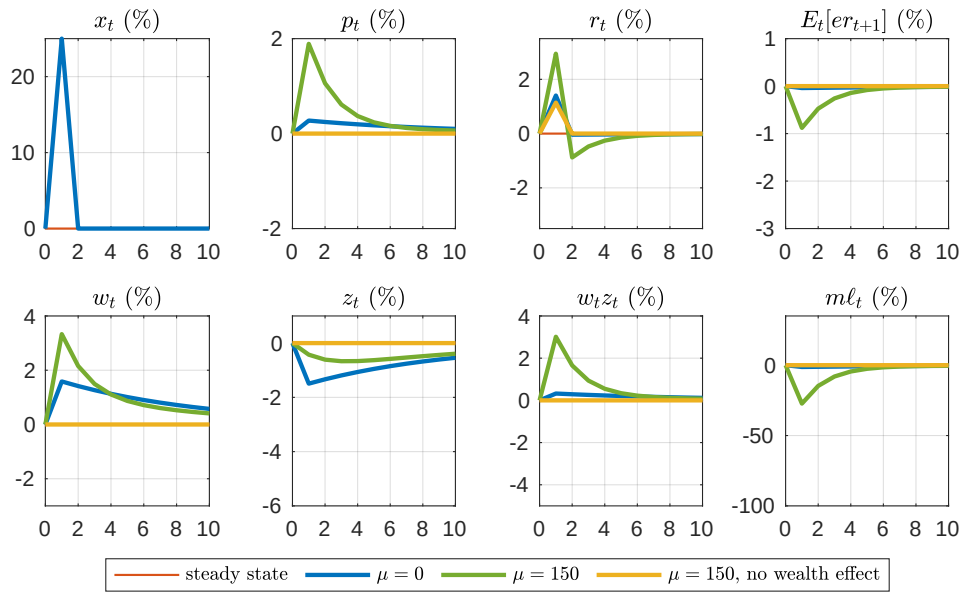


Figure C.1: Decomposed Impulse Response to a Transitory Cash Flow Shock

Notes: This figure reports the decomposed effects in the responses of the key state variables to a transitory shock to the bond's cash flow. We consider three cases: with transaction costs, without transaction costs, and with transaction costs but no wealth effect.

the interest rate i_t in period 1 reduces the bond price p_t and thereby decreases the bond investors' wealth. So, if bond investors maintain a constant level of wealth, they can adjust their portfolios to hold more bonds compared to the scenario with time-varying wealth. In this case, upon the interest rate shock, liquidity traders need to sell some bonds to meet the increased demand from bond investors. Consequently, they ask for a smaller bond risk premium in period 1, which mitigates the price drop due to the direct effect of the increased interest rate.

In period 2, as the interest rate shock dissipates, the bond price rebounds and returns very close to the steady-state level. Similarly, the wealth of bond investors also reverts to approximately the steady-state level, and the dynamics of state variables are similar both with and without the bond investors' wealth effect.

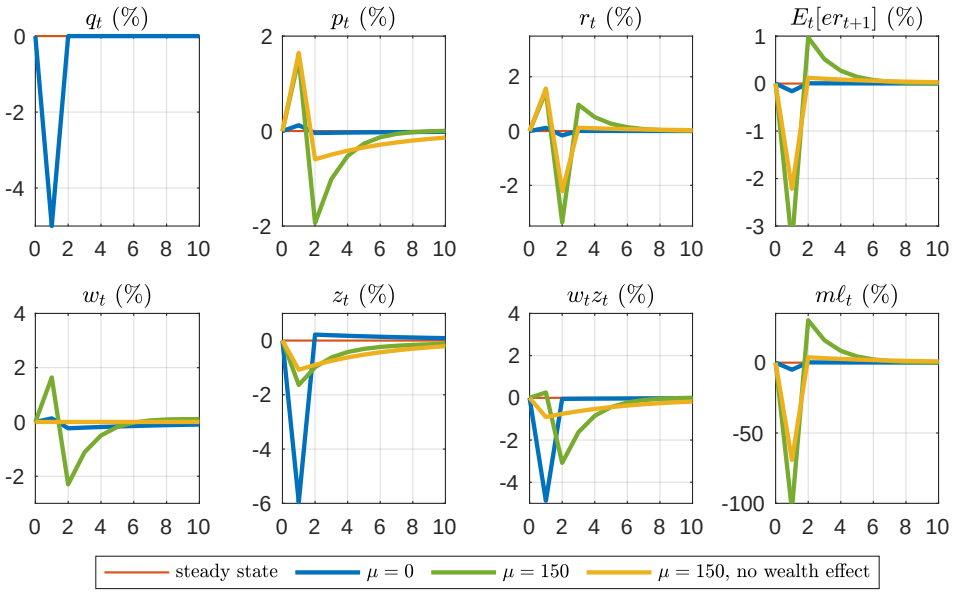


Figure C.2: Decomposed Impulse Response to a Transitory Supply Shock

Notes: This figure reports the decomposed effects in the responses of the key state variables to a transitory shock to the bond quantity. We consider three cases: with transaction costs, without transaction costs, and with transaction costs but no wealth effect.

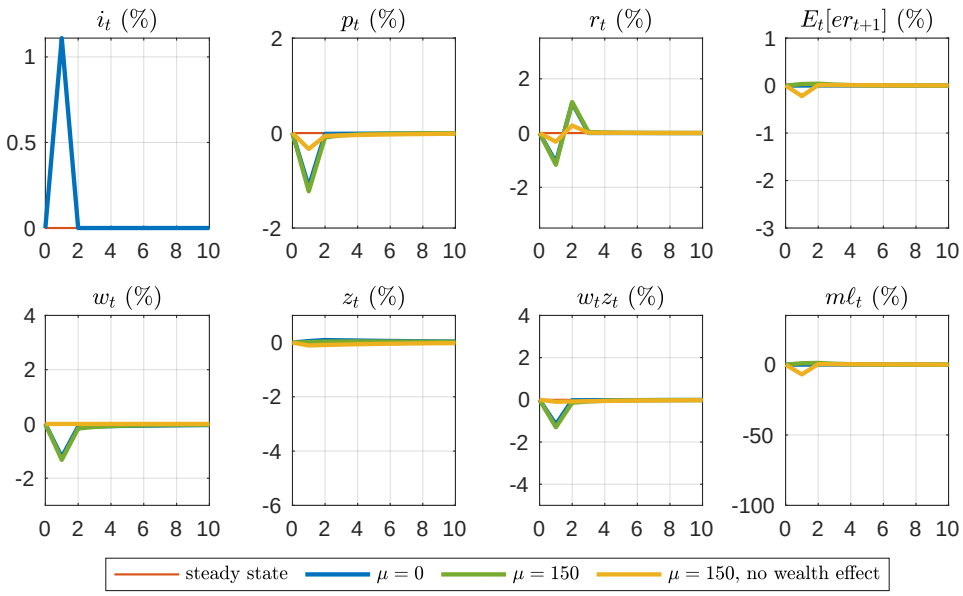


Figure C.3: Decomposed Impulse Response to a Transitory Interest Rate Shock

Notes: This figure reports the decomposed effects in the responses of the key state variables to a transitory shock to the interest rate. We consider three cases: with transaction costs, without transaction costs, and with transaction costs but no wealth effect.

C.2 Impulse Responses to Shocks to Expectations

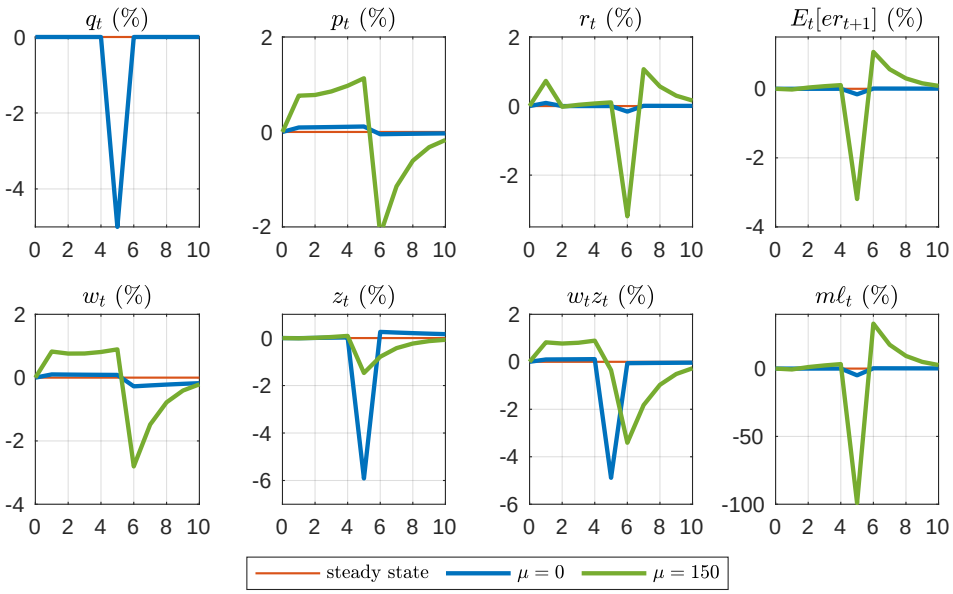


Figure C.4: Impulse Response to an Expected Bond Quantity Shock

Notes: This figure reports the responses of all the key state variables to a transitory shock in the expectation of bond supply. We examine two cases: with and without transaction costs.

C.3 Model Extension with Long-Horizon Bond Investors

In this appendix, we consider a model extension in which the bond investors have an infinite horizon. In this case, when the bond investors make portfolio decisions, they take into account not only how the current portfolio deviates from the previous position, but also how it affects the transaction costs arising from future portfolio adjustments.

We follow the setting in Bacchetta et al. (2023). Focusing on the first-order dynamics, we additionally abstract away from the variance shock by assuming a constant return variance $\bar{\sigma}^2 = var_t(r_{t+s})$ for any t and $s = 1, 2, \ldots$. With an infinite investment horizon, the bond investor's portfolio choice problem is

$$\sum_{s=0}^{\infty} \beta^s \left(\mathbb{E}_t \left[r^p_{t+s+1} \right] - \frac{1}{2} \gamma \, var_t \left(r^p_{t+s+1} \right) - \frac{1}{2} \mu \cdot c \left(z_{t+s}; z_{t+s-1}, \bar{z} \right) \bar{\sigma}^2 \right). \tag{C.1}$$

The trading cost function $c\left(z_t; z_{t-1}, \bar{z}\right)$ is the same as in the main text:

$$c\left(z_t; z_{t-1}, \bar{z}\right) = (1-\eta)\left(z_t - z_{t-1}\right)^2 + \eta \left(z_t - \bar{z}\right)^2.$$

Since the portfolio return is $r^p_{t+s+1} = i_{t+s} + z_{t+s}(r_{t+s+1} - i_{t+s})$, we have

$$var_t \left(r^p_{t+s+1} \right) = var_t \left(z_{t+s} r_{t+s+1} \right) = z^2_{t+s} \, var_t \left(r_{t+s+1} \right) = z^2_{t+s} \bar{\sigma}^2,$$

and the bond investor's portfolio choice problem becomes

$$\sum_{s=0}^{\infty} \beta^s \left(\mathbb{E}_t \left[i_{t+s} + z_{t+s} \left(r_{t+s+1} - i_{t+s} \right) - \frac{1}{2} \gamma z^2_{t+s} \bar{\sigma}^2 - \frac{1}{2} \mu \left((1-\eta) \left(z_{t+s} - z_{t+s-1} \right)^2 + \eta \left(z_{t+s} - \bar{z} \right)^2 \right) \bar{\sigma}^2 \right] \right).$$

The first-order condition with respect to z_t is

$$\mathbb{E}_t \left[er_{t+1} \right] - \gamma z_t \bar{\sigma}^2 - \mu \bar{\sigma}^2 (1-\eta) \left(z_t - z_{t-1} \right) - \mu \bar{\sigma}^2 \eta \left(z_t - \bar{z} \right) + \beta \mu \bar{\sigma}^2 (1-\eta) \left(\mathbb{E}_t [z_{t+1}] - z_t \right) = 0.$$

Rearranging the terms, we have

$$(\gamma + \mu + \beta\mu(1-\eta))\, z_t \bar{\sigma}^2 = \mathbb{E}_t\left[er_{t+1}\right] + \mu\left(\eta\bar{z} + (1-\eta)z_{t-1} + \beta(1-\eta)\mathbb{E}_t[z_{t+1}]\right)\bar{\sigma}^2,$$

which implies an optimal portfolio solution

$$z_t = \frac{\gamma}{\gamma + \mu + \beta\mu(1-\eta)}\frac{\mathbb{E}_t\left[er_{t+1}\right]}{\gamma\bar{\sigma}^2} + \frac{\mu}{\gamma + \mu + \beta\mu(1-\eta)}\left(\eta\bar{z} + (1-\eta)z_{t-1} + \beta(1-\eta)\mathbb{E}_t[z_{t+1}]\right). \tag{C.2}$$

This is the central equation that extends the optimal portfolio choice Eq. (3) in the baseline case. When the bond investors have an infinite horizon, their portfolio choice additionally depends on the expectation of future portfolio choice $\mathbb{E}_t[z_{t+1}]$, with weight $\mu\beta(1-\eta)$. For comparison, the weight on the past portfolio choice z_{t-1} is $\mu(1-\eta)$. This new term reflects the investors' desire to stay close to the target towards which they are adjusting towards in the next period, which allows them to minimize the transaction costs in a forward-looking fashion.

In the steady state, we still have

$$\bar{z} = \frac{\bar{er}}{\gamma\bar{\sigma}^2}.$$

Combining these results, we derive the following forward-looking linear equation for the bond investor's portfolio choice.

$$\begin{aligned}\hat{z}_{t+s} &= \frac{\gamma}{\gamma + \mu + \beta\mu(1-\eta)}\left[\frac{\mathbb{E}_t\left[er_{t+s+1}\right] - \bar{er}}{\gamma\bar{\sigma}^2}\right] + \frac{\mu(1-\eta)}{\gamma + \mu + \beta\mu(1-\eta)}\left(\hat{z}_{t+s-1} + \beta\mathbb{E}_t[\hat{z}_{t+s+1}]\right) \\ &\approx \frac{\gamma}{\gamma + \mu + \beta\mu(1-\eta)}\frac{\mathbb{E}_t[\hat{er}_{t+s+1}]}{\gamma\bar{\sigma}^2} + \frac{\mu(1-\eta)}{\gamma + \mu + \beta\mu(1-\eta)}\left(\hat{z}_{t+s-1} + \beta\mathbb{E}_t[\hat{z}_{t+s+1}]\right).\end{aligned}$$

For calibration, we need to specify the discount rate parameter β. We choose $\beta = 0.96$, and the results are similar within its neighborhood. In the other extreme, if we pick $\beta = 0$, then, the investors become one-period optimizers again and the results converge to baseline case in the main

text.

The Cash Flow Shock. The impulse responses to the transitory cash flow shock are similar to those in the main text, both in shape and magnitude. Given that bond investors now face an infinite-horizon problem, their portfolio shares are a weighted average of the mean-variance solution, the last-period holdings, and the expected next-period holdings, which makes their demand more inelastic. Consequently, the price response is slightly greater in this case.

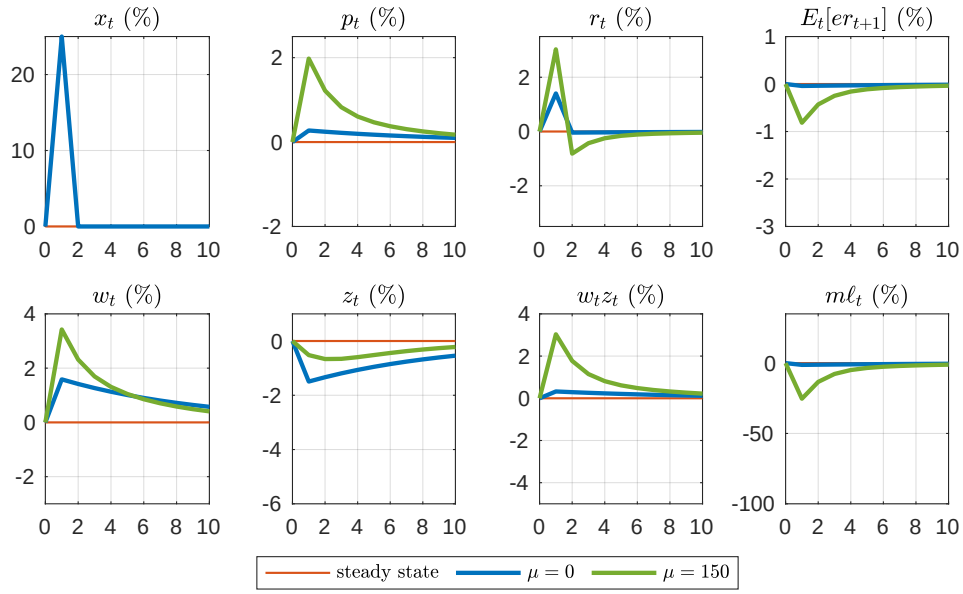


Figure C.5: Impulse Response to a Transitory Cash Flow Shock with Long-Horizon Bond Investors

Notes: This figure reports the responses of the key state variables to a transitory shock to the bond's cash flow when the bond investors face a long-horizon problem. We consider two cases: with and without transaction costs.

The Bond Quantity Shock. The impulse responses to the transitory bond quantity shock are similar to the results in the main text qualitatively, but there are some quantitative differences. First, the initial bond price increase in response to QE in period 1 is stronger, implying an even lower demand elasticity of 1.9 instead of 3.3 from the main text. Second, the price overshoot in period 2 is smaller in magnitude. The intuition is that, as the bond investors have a longer horizon,

they understand that if they lower their bond share in period 1 during the QE phase, they need to buy back more bonds and incur greater transaction costs in period 2 during the QT phase. This makes them even more reluctant to adjust their bond holdings in period 1. As a result, when QT happens in period 2, the liquidity traders need to absorb fewer bonds sold by the central bank, leading to a smaller price decline.

We find that this is similar to increasing μ in the case where the bond investors optimize over a one-period problem, which also makes bond investors put a smaller weight on the mean-variance term and a greater weight on the transaction cost term, and makes the demand more inelastic. Indeed, Eq. (C.2) shows that the infinite-horizon investors assign a higher weight ($\mu+\beta\mu(1-\eta)$) to the transaction cost term relative to the mean-variance term (γ), whereas the one-period investors assign a weight of μ to the transaction cost term. That said, the intuition about the bond price overshoot remains robust.

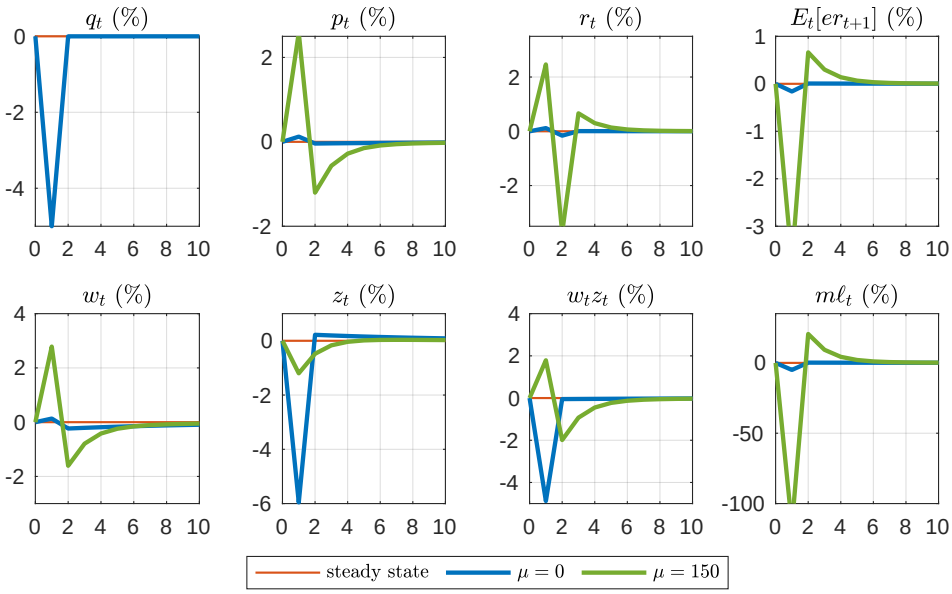


Figure C.6: Impulse Response to a Transitory Supply Shock with Long-Horizon Bond Investors

Notes: This figure reports the responses of the key state variables to a transitory shock to the bond supply when the bond investors face a long-horizon problem. We consider two cases: with and without transaction costs.

The Interest Rate Shock. The impulse responses to the transitory interest rate shock with long-horizon bond investors are almost identical to those in the main text.

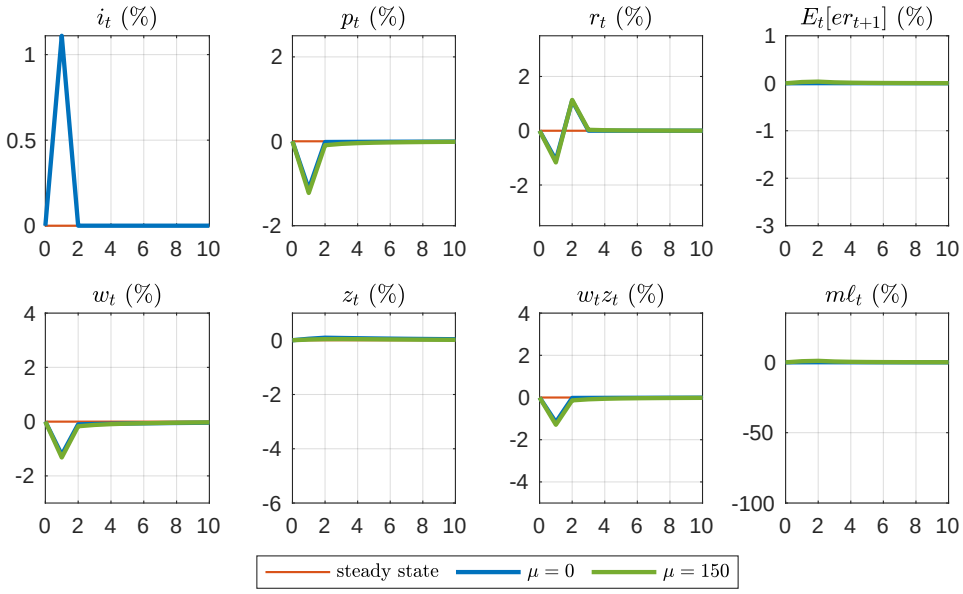


Figure C.7: Impulse Response to a Transitory Interest Rate Shock with Long-Horizon Bond Investors

Notes: This figure reports the responses of the key state variables to a transitory shock to the interest rate when the bond investors face a long-horizon problem. We consider two cases: with and without transaction costs.