ISSN 1936-5330

Research Working Papers

Rethinking Central Bank LSAPs: The Power of Market Functioning Purchases

A. Lee Smith and Victor J. Valcarcel

July 2026
RWP 26-05
http://doi.org/10.18651/RWP2026-05

FEDERAL RESERVE BANK OF KANSAS CITY

Rethinking Central Bank LSAPs: The Power of Market Functioning Purchases*

A. Lee Smith† Victor J. Valcarcel‡

July 2026

Abstract

We show that operationally similar central bank asset purchases can have markedly different effects. Combining security-level data on the Federal Reserve's duration-adjusted asset holdings with narrative event-based identification reveals that purchases made for accommodation strongly affect yields, while purchases made for market functioning primarily enhance liquidity. We advance a partial equilibrium model of an intermediated bond market that can reconcile these findings. When trading flow is orderly, large-scale asset purchases (LSAPs) operate through the expected supply of duration with large effects on yields. When trading flow is disorderly, LSAPs reduce dealer inventories, improve liquidity, and compress bid-ask spreads.

JEL Classification: E3, E4, E5

Keywords: Monetary Policy, Balance Sheet Policy, SOMA, Quantitative Easing, Primary Dealers, Market Making, Narrative Restrictions, Structural VAR, Liquidity, Intermediation.

*We are indebted to Mariia Dzholos for excellent assistance and research support. The authors thank the Center for Advancement of Data and Research in Economics (CADRE) at the Federal Reserve Bank of Kansas City (FRB-KC) for providing essential data and computational resources that contributed to the results reported in this paper. The authors also thank CADRE Research Facilitator Jacob Dice for valuable technical expertise. Valcarcel thanks FRB-KC for hosting him while a portion of this work was completed. Other acknowledgments are withheld for review. The views expressed are those of the authors and do not necessarily reflect the positions of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

†Federal Reserve Bank of Kansas City. Email: andrew.smith@kc.frb.org
‡University of Texas at Dallas. Email: victor.valcarcel@utdallas.edu

1 Introduction

The Federal Reserve has greatly expanded the use of its balance sheet as a policy tool. Large-scale asset purchases (LSAPs) have been used in some instances to alleviate strains in financial markets and, at other times, to provide monetary policy accommodation. In the Global Financial Crisis, the QE 1 LSAP program was announced on November 2008 to improve liquidity in the MBS market and ease mortgage rates. Over time, QE 1 purchases were expanded to include longer-term Treasury securities with the intent of reducing longer-term yields and providing further accommodation. Later rounds of LSAPs for accommodation — QE 2/MEP/QE 3 — would follow. In March 2020, LSAPs were again initiated to address severe market dysfunction and purchases were later maintained to provide accommodation.

Despite being operationally similar, there is growing evidence that LSAPs made for policy accommodation elicit markedly different effects than market functioning purchases. For example, the event study literature has generally found that QE 1/2/3 drove down longer-term interest rates upon announcement of the forthcoming balance sheet expansion. In sharp contrast, Vissing-Jorgensen (2021) shows that announcements of LSAPs by the Federal Reserve in the spring of 2020 had little to no effect on Treasury yields. It was not until purchases were executed and scaled up to "*the amounts needed*" to restore market functioning that they began to have discernible effects. Fleming et al. (2021) show that these effects were concentrated on bid-ask spreads and liquidity metrics in the Treasury market.

Evidence of contrasting effects between accommodation and market functioning purchases poses a challenge to the current intellectual framework for understanding LSAPs, which is largely premised on the seminal work of Vayanos and Vila (2021). The Vayanos and Vila (2021) model predicts that LSAPs reduce longer-term interest rates by shifting duration risk away from risk-averse investors, reducing the market price of duration. Through this lens, LSAPs that reduce the supply of duration should always lower interest rates.

However, in this paper we provide empirical evidence that accommodation and market functioning LSAPs both reduce the supply of duration but systematically differ in their effects on yields. We then present a theoretical model that builds on Vayanos and Vila to incorporate bond market intermediation and provides a distinct role for market functioning purchases. In our model, LSAPs sometimes reduce the price of duration and lower yields, and other times reduce order flow imbalances and compress bid-ask spreads. Our results provide a sharp distinction between accommodation and market-functioning LSAPs.

Our empirical analysis centers on a duration-adjusted measure of the Federal Reserve's bond portfolio, as par values would be insufficient to capture the duration-risk channel in Vayanos and Vila (2021). Using security-level data on the Federal Reserve's System Open Market Account (SOMA) — comprising Treasury and government agency securities — we construct a duration-adjusted SOMA portfolio. Intuitively, our duration-adjusted measure of the Fed's SOMA portfolio homogenizes the par value of each security into ten-year equivalents so that longer-duration securities are more heavily weighted. Credible identification necessitates relatively high-frequency data. Therefore, we construct our duration-adjusted SOMA measure at a weekly frequency and combine it with Treasury yields and indicators of MBS and Treasury market liquidity in a structural vector autoregressive (VAR) model.

We impose narrative restrictions to identify and trace out the dynamic effects of *accommodation* and *market functioning* interventions. Using only three narrative events — the LSAP announcement in November 2008, the announcement of the Maturity Extension Program in September 2011, and the pandemic purchases in March 2020 — drastically sharpens the identification of accommodation and market functioning shocks. Our identified accommodation LSAP shocks (henceforth, 'accommodation shocks') lead to large declines in Treasury yields on impact despite purchases occurring in the weeks and months ahead. This result is consistent with much of the event study literature that assumes that the yield effects of LSAPs are priced in to markets upon announcement. While accommodation purchases have persistent effects on yields, our estimates indicate that they have little effect on market liquidity metrics. In contrast, market functioning LSAP shocks (henceforth, 'market functioning shocks') lead to significant improvements in measures of market liquidity. Despite increasing the amount of duration risk on the Fed's balance sheet, market functioning purchases have a minimal effect on yields.

We develop a partial equilibrium model of market functioning purchases that can rationalize our empirical findings. While building on the work of Vayanos and Vila (2021) we take a major departure from their framework: we assume that the bond market is intermediated by market makers, akin to primary dealers.

In our framework, market makers act as intermediaries in the bond market, passively taking buy and sell orders. We distinguish between two environments, one where the order flow is balanced and one where it is imbalanced. These are each depicted in Figure 1 and fully characterized in Section 4. When the order flow is balanced, the inventory of the market maker is low and central bank asset purchases affect the bond market primarily through the expected supply of duration, as in Vayanos and Vila (2021). Our model predicts that, in this state of the world, LSAPs can incite large declines in yields but have little to no effect on market liquidity — this is depicted in the top panel of Figure 1. In contrast, when order flow deteriorates and inventories accumulate on market maker balance sheets, our model shows that LSAPs reduce elevated inventories, which can sharply compress bid-ask spreads and improve liquidity conditions. In this state of the world, depicted in the bottom panel of Figure 1, LSAPs have no effect on yields.

Given that, in practice, U.S. government bond markets are intermediated by dealers, our model places a central role on the state of intermediaries' balance sheets as a determining factor for the transmission of LSAPs. As a result, the effects of central bank asset purchases on the price of duration risk — to which all long-duration assets are priced by assumption — are not constant. Instead, the effects depend crucially on market makers' own balance sheets. Through this dependence, the model is able to reconcile how operationally similar asset purchases can have markedly different effects on yields and liquidity conditions.

The rest of this paper proceeds as follows. Section 2 presents our empirical analysis with its results reported in Section 3. Section 4 presents our theoretical model of bond market supply, demand, and intermediation. Section 5 outlines some lessons and implications of our model for central bank balance sheet policy. Section 6 concludes.

2 Estimating the Effects of LSAPs

We present new empirical evidence of the differential effects between the Federal Reserve's market functioning and accommodation purchases. There is a substantial body of research on the effects of quantitative easing on interest rates, but much less work on the effects of market functioning purchases.[1] To our knowledge, there has been little systematic comparison between the two types of LSAP programs on a duration-adjusted, dollar-for-dollar basis. Such a comparison requires detailed data on the Fed's security holdings. Moreover, this detailed data must be analyzed at a reasonably high frequency to achieve credible time-series identification. Specifically, given the rich historical record of the Fed's LSAP interventions, we employ an event-based narrative identification strategy. We begin with a description of the data then discuss our structural VAR model and identification strategy.

[1]See Williams (2014) or Bhattarai and Neely (2022) for surveys of this literature.

2.1 Data and the Federal Reserve's SOMA Portfolio

The New York Fed provides detailed data on the securities held in the Federal Reserve's System Open Market Account (SOMA), the account that holds the assets acquired by way of the Fed's open market operations. This data is available at a weekly frequency since 2003 and is updated every Thursday, reflecting balances as of the week-ending Wednesday.

We use this security-level data to construct a measure of the duration of the Fed's SOMA portfolio. According to the theory in Vayanos and Vila (2021), LSAPs operate primarily by changing the duration risk held by the marginal investor. Therefore, the theoretically relevant measure of the Fed's SOMA portfolio is the duration-adjusted holdings rather than the par value of the holdings. To convert the par-value of SOMA into a duration-adjusted measure, we apply the Macaulay Modified Duration formula to each Treasury security.[2] As described in Hanson (2014), the duration of an MBS is not a simple function of coupon payments and interest rates given the potential for pre-payments to respond to changes in interest rates. Therefore, we use an index of agency MBS duration available from Bloomberg to measure the duration of the Fed's MBS holdings.[3]

For both Treasuries and MBS, we calculate the duration for each security with the yield curve fixed at its position on July 30, 2014, following Greenwood et al. (2014). This permits the interpretation of movements in the duration of the SOMA portfolio as arising from changes in security holdings, rather than changes in interest rates. We take the ratio of each security's duration to the duration of the 10-yr Treasury note on July 30, 2014, effectively scaling all duration into 10-yr equivalents. Finally, to construct a measure of the duration of the SOMA portfolio, we scale the par value of each security by the the relative duration of the security and sum across securities. The resulting formula for SOMA TYE (ten-year equivalents) is expressed below:

$$\text{SOMA TYE}_t = \sum_{s=1}^{S} Par\,Value_{s,t} \frac{Duration_{s,t}}{Duration_{10yr\,TSY,t}}. \tag{1}$$

Figure 2 shows our resulting measure of the duration-adjusted SOMA portfolio, which we refer to as SOMA TYE, and compares our measure to the par-value of the SOMA portfolio. While the contours of the two SOMA measures are similar, there are some important differences as well.

[2]The Macaulay Modified Duration approximates the percentage change in the price of a fixed-income security in response to a change in interest rates.

[3]The ticker symbol we use is: LUMSMD.

Most notably for our analysis of the duration effects of the Fed's LSAP interventions is the September 2011 - December 2012 maturity extension program, or MEP. During the MEP the Fed sold about $670 billion of Treasury securities maturing in three years or less using the proceeds to purchase Treasury securities with maturities beyond six years. The objective of the MEP was to put downward pressure on longer-term interest rates by removing duration risk from investors and placing it on the Fed's balance sheet.

Importantly — given that the MEP constituted a change in the maturity composition rather than the size of the Fed's balance sheet — Figure 2 highlights that the par value of the SOMA portfolio was unaffected by this policy, while the SOMA TYE quickly grew during this period. Figure 3 shows that the MEP increased the average maturity of the Fed's Treasury portfolio from about six years to over 10 years. This episode demonstrates how our measure of SOMA TYE is better able to capture the full range of the Fed's balance sheet policies, including those that affect the size and composition of the SOMA portfolio.

In our structural VAR model, we complement our weekly SOMA TYE duration variable with three others that are also measured on a week-ending Wednesday basis: the benchmark 10-yr Treasury yield, a measure of liquidity in the Treasury market, and a measure of liquidity in the MBS market. We discuss our liquidity metrics in turn.

Figure 4a shows our measure of Treasury market liquidity, an index of aggregate U.S. government liquidity produced by Bloomberg. The index is a measure of yield deviations from a fair-value model of U.S. Treasury notes and bonds with remaining maturities of one year or greater. Lower values of the index indicate smaller pricing "errors" and therefore better liquidity in the Treasury market; whereas higher values of the index reflect worse liquidity. Figure 4a shows the index peaked exceeding a value of 20 in the aftermath of GFC. Since then, the index typically lies between values of 2 and 4, subject to local peaks in late 2015, the onset of COVID in the U.S in March 2020, and November 2022 as mounting post-COVID inflationary pressures led to large adjustments in interest rates.

Since market functioning interventions have also targeted MBS liquidity, we include a measure of MBS market liquidity in our structural VAR model. Figure 4b shows the option-adjusted spread for agency MBS also obtained from Bloomberg. The spread is measured in basis points above the Treasury yield curve and captures the compensation investors demand for holding an MBS after removing prepayment risk.[4] Agency MBS issuance by government

[4]The risk of prepayment is embedded in MBS because refinancing, home sales, and foreclosures result in

sponsored enterprises have always had an implicit — and over much of our sample, explicit — government guarantee against default. Thus, we interpret the spread as primarily representing liquidity risk in the MBS market rather than credit risk; with higher values reflecting worse liquidity. The spread reaches its highs (near 200 bps) during the GFC. Subsequently, the MBS option-adjusted spread widened towards 100 on several occasions, including the end of 2011, March 2020, and in March 2023 following the failure of several regional banks.

2.2 Structural VAR Model

We combine our data on the Fed's duration-adjusted SOMA holdings and financial market variables in a weekly-frequency structural VAR model. Our estimation sample spans August 2007 through June 2026, yielding nearly 1000 weekly observations. Even with our long high-frequency sample, identifying assumptions are needed to distinguish between accommodation shocks, market functioning shocks, and other week-to-week movements in the Fed's balance sheet.

The primary challenge is that an observed movement in the Fed's SOMA holdings could reflect a market functioning intervention, the execution of a previously announced LSAP, or a traditional open market operation to manage reserve balances. However, standard invertibility results suggest that an appropriately specified structural VAR model is capable of recovering the two shocks of interest. We specify the following n-variable VAR system of lag length p and sample size T:

$$\mathbf{y}_t'\mathbf{A}_0 = \mathbf{x}_t'\mathbf{A}_+ + \varepsilon_t', \quad \varepsilon_t \sim \mathcal{N}(0, I_n), \tag{2}$$

for $1 \leq t \leq T$. The observed variables are collected in $\mathbf{y}_t$, an $n \times 1$ vector. The block row vector $\mathbf{A}_+$ vectorizes the collection of p-many $n \times n$ autoregressive matrices such that $\mathbf{A}_+' = [\mathbf{A}_1' \cdots \mathbf{A}_p' \, \mathbf{c}']$, where $\mathbf{c}$ is a $1 \times n$ vector of intercept terms, and $\mathbf{x}_t$ stacks all the lagged endogenous terms such that $\mathbf{x}_t' = [\mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-p}', 1]$. The dimension of $\mathbf{A}_+$ is $(np+1) \times n$ and the dimension of $\mathbf{x}_t$ is $(np+1) \times 1$. The structural shocks are collected in ε_t, an $n \times 1$ vector, where first two elements correspond to the market functioning and accommodation shocks.[5]

early principal repayments. These prepayments generate uncertain cash flow timing; which, in principle, an option-adjusted spread can capture better than other simpler yield spreads.

[5]We make this assumption without loss of generality since the ordering of the variables plays no role in our identification strategy.

In general, $\mathbf{A}_0 \neq \mathbf{I}_{n\times n}$ and therefore equation (2) cannot be directly estimated to recover the structural shocks. Instead, we estimate the reduced-form VAR, which is given by:

$$\mathbf{y}_t' = \mathbf{x}_t'\mathbf{B} + \mathbf{u}_t' \quad \mathbf{u}_t \sim \mathcal{N}(0, \Sigma), \tag{3}$$

for $1 \leq t \leq T$. Comparing equations (2) and (3) reveals the mapping between the reduced-form parameters, the matrices $\mathbf{B}$ and Σ, and the structural parameters $\mathbf{A}_0$ and $\mathbf{A}_+$. In particular, $\mathbf{B} = \mathbf{A}_+\mathbf{A}_0^{-1}$ and $E[\mathbf{u}_t\mathbf{u}_t'] = \Sigma = (\mathbf{A}_0\mathbf{A}_0')^{-1}$. Similarly, $\mathbf{u}_t' = \varepsilon_t'\mathbf{A}_0^{-1}$ are the one-period ahead forecast errors. The need for identifying assumptions arises because there are more free parameters in $\mathbf{A}_0$ than can be gleaned from the data based on the mapping $\Sigma = (\mathbf{A}_0\mathbf{A}_0')^{-1}$.

We employ narrative sign restrictions to identify the structural SVAR parameters, as proposed by Antolín-Díaz and Rubio-Ramírez (2018). Narrative restrictions improve over well-known shortcomings of traditional sign restrictions. While traditional sign restrictions have been widespread in the monetary literature, they only provide set identification.[6] This often results in a large set of candidate structural VAR models that are consistent with the restrictions, increasing the potential for misidentification. Arias et al. (2019) and Wolf (2020) provide evidence of misidentification in the context of conventional monetary policy shocks. Both papers propose augmenting sign restrictions with additional identifying restrictions to narrow the identified set. Narrative restrictions can accomplish this by requiring that the SVAR align with the historical record. Antolín-Díaz and Rubio-Ramírez (2018) show that "even a single event can dramatically sharpen the inference or even alter the conclusions of SVARs only identified with traditional sign restrictions."

Narrative restrictions are particularly well-suited to our purposes because of the rich narrative account of the Fed's LSAP interventions. Intuitively, narrative sign restrictions constrain the structural parameters to ensure that the output of the structural VAR is consistent with the historical record at a handful of important events. To make matters concrete, we place narrative restrictions on the structural shocks, impulse response functions (IRF), and historical decompositions (HD), each defined below.[7]

The algorithm in Antolín-Díaz and Rubio-Ramírez (2018) proceeds by considering candidate draws of the structural VAR parameters. Let $\Theta = (\mathbf{A}_0, \mathbf{A}_+)$ collect the value of the structural parameters. Hence, for a given candidate draw of Θ and the data $(\mathbf{y}_t)$, the

[6]See Faust (1998), Canova and De Nicolo (2002), and Uhlig (2005) for seminal work on sign restrictions.
[7]See Kilian and Lütkepohl (2017) for a comprehensive treatment.

structural shocks at time t are given by:

$$\varepsilon_t'(\Theta) = \mathbf{y}_t'\mathbf{A}_0 - \mathbf{x}_t'\mathbf{A}_+ \quad \text{for } 1 \leq t \leq T. \tag{4}$$

Given a value Θ of the structural parameters — and starting from the impact matrix $\mathbf{L}_0(\Theta) = (\mathbf{A}_0^{-1})'$ — IRFs can be defined recursively by:

$$\begin{aligned} \mathbf{L}_k(\Theta) &= \sum_{\ell=1}^{k}(\mathbf{A}_\ell\mathbf{A}_0^{-1})'\mathbf{L}_{k-\ell}(\Theta), \quad \text{for } 1 \leq k \leq p, \\ \mathbf{L}_k(\Theta) &= \sum_{\ell=1}^{p}(\mathbf{A}_\ell\mathbf{A}_0^{-1})'\mathbf{L}_{k-\ell}(\Theta), \quad \text{for } p < k < \infty. \end{aligned} \tag{5}$$

where the response of the ith variable to the jth structural shock at horizon k corresponds to the element in row i and column j of the matrix $\mathbf{L}_k(\Theta)$.

Once structural shocks $\varepsilon_t(\Theta)$ are obtained from estimation of (2), HDs can be obtained by accumulating the contribution of each jth shock to the observed unexpected change in the ith variable between periods t and $t+h$ as follows:

$$H_{i,j,t,t+h}(\Theta, \varepsilon_t, \ldots, \varepsilon_{t+h}) = \sum_{\ell=0}^{h}\mathbf{e}_{i,n}'\mathbf{L}_\ell(\Theta)\mathbf{e}_{j,n}\mathbf{e}_{j,n}'\varepsilon_{t+h-\ell}, \tag{6}$$

where $\mathbf{e}_{j,n}$ is the jth column of $\mathbf{I}_n$, for $1 \leq i,j \leq n$ and for $h \geq 0$.

2.3 VAR Identifying Restrictions

We identify accommodation shocks and market functioning shocks using a combination of sign, zero, and narrative restrictions. Before we describe those restrictions, it is worth emphasizing that our investigation centers on the effects of balance sheet policy. Therefore, all of our restrictions are imposed exclusively on our SOMA TYE measure, leaving the remaining variables in our VAR unencumbered by assumptions and letting the data speak for itself on how balance sheet policy affects yields and liquidity measures.[8] We first layout our baseline sign and zero restrictions before moving on to the more informative and important narrative restrictions.[9]

[8]Regarding the orthogonality of the accommodation and market functioning shocks, we considered other identification strategies which operate on different assumptions and approaches, some of which are more restrictive. Nevertheless, other results premised on alternative identification strategies offer a consistent story and are available in Appendix B.

[9]The following set of restrictions are implemented with Algorithm 1 in Antolín-Díaz and Rubio-Ramírez (2018). A comprehensive discussion on Bayesian inference can be found in their paper.

Assumption 1: Sign Restriction: *A market functioning shock ($j = 1$)* ***increases*** *SOMA TYE on impact (i.e. in the current week).*

Assumption 2: Sign Restriction: *An accommodation shock ($j = 2$)* ***increases*** *SOMA TYE 52 weeks after the shock.*

Assumption 3: Zero Restriction: *An accommodation shock ($j = 2$) has* ***no effect*** *on SOMA TYE within the current week.*

The sign restrictions in Assumptions 1 and 2 simply normalize the qualitative nature of the shocks (e.g. an expansionary shock to the balance sheet must be a positive shock to SOMA balances) and hence play no meaningful role in orthogonalizing the two shocks.[10] We place our sign restriction on the market functioning shock on the impact response of SOMA TYE. However, the sign restriction for the accommodation shock on SOMA TYE is placed 52 weeks after the shock is realized.[11] This is because we restrict the accommodation shocks to have zero impact on SOMA TYE within the week when they are announced and, hence, the sign normalization must be placed on a future horizon. As we discuss next, this zero restriction is important for disentangling these two types of LSAP shocks.

The zero restriction assumes that accommodation shocks are announced before any purchases take place. This aligns with the Fed's practice of announcing LSAPs in advance. However, we emphasize that our restriction allows for the possibility that the effects on financial markets begin to take hold well before purchases are executed. Therefore, our assumption is fully consistent with the large event-study literature that reports evidence of movements in financial markets within minutes of learning about a new LSAP program (Gagnon et al., 2011; Krishnamurthy and Vissing-Jorgensen, 2011; D'Amico et al., 2012).

In contrast, Vissing-Jorgensen (2021) provides compelling evidence that there were no meaningful announcement effects for the Fed's March 2020 market functioning purchases. Moreover, the urgency of the situation has often required that market functioning interventions be announced and implemented within one week's time, as in March of 2020, for example.[12] In this respect, the zero impact restriction on accommodation shocks helps to

[10] In fact, any candidate SVAR draw can be made to satisfy the normalization sign restrictions by scaling the appropriate column of the orthonormal rotation matrix by -1, see Arias et al. (2018).

[11] The choice of 52 weeks reflects our observation that the Fed's LSAP announcements take about 12 months on average to be fully implemented after being announced. We also experimented with horizons as short as 40 weeks and as long 104 weeks with negligible differences.

[12] See the press release issued by the FOMC on March 23: `https://www.federalreserve.gov/`

distinguish these from market functioning interventions.

However — as we will show to be consistent with our prior discussion of misidentification — this zero restriction alone is insufficient to credibly identify these two types of LSAP shocks. Therefore, we incorporate two narrative restrictions, tied to three key dates. Note that the dates listed in the narrative restrictions are the relevant *Wednesday* on which the shocks would appear in our data. We give exact dates of press releases or announcements pertaining to the narrative account in the discussion below.

Assumption 4: Narrative Sign Restriction 1: *The accommodation shock* ($j = 2$) ***was positive*** *on November 26, 2008, and September 21, 2011.*

These two dates were both consequential in the chronology of the Fed's LSAP programs. Tuesday November 25, 2008 constituted the first LSAP announcement and marked a watershed moment in the Fed's use of its balance sheet as a policy tool. The announced purchase of $500bn MBS and $100bn agency debt enlarged not only the expected size of the SOMA portfolio, but also increased its expected duration and, hence, had a pronounced effect on expected SOMA TYE.

From a narrative standpoint, this first LSAP announcement was also much less anticipated than some of the subsequent ones. Moreover, this LSAP announcement was not confounded by any newly issued forward rate guidance. In particular, the Tuesday November 25, 2008 press release was not an FOMC decision, but rather Chairman Bernanke's decision, hence there was no accompanying Committee guidance on future interest rates. This stands in sharp contrast to the December 16, 2008 or March 18, 2009 FOMC Statements — two other important dates when the Committee hinted or announced large-scale purchases of longer-term Treasuries. However, on both of these dates the FOMC also issued new forward guidance for short-term interest rates.[13]

Similar rationales underpin our choice of the second date, the MEP announcement on September 21, 2011. As previously discussed, the MEP was an explicit policy of extending SOMA duration with no expansion in the overall size of the SOMA portfolio, achieved by

newsevents/pressreleases/monetary20200323a.htm

[13]Prior research has failed to reach a consensus on the relative importance of the duration effects of asset purchases versus forward guidance and signaling effects of the FOMC's LSAP announcements on longer-term interest rates (Krishnamurthy and Vissing-Jorgensen, 2011; Woodford, 2012; Bauer and Rudebusch, 2014; Ray et al., 2024). The problem is challenging since many of the FOMC's LSAP announcements included changes in forward guidance. Swanson (2020) provides a detailed decomposition of FOMC announcements.

selling $670 billion of shorter-term Treasury securities and using the proceeds to purchase longer-term Treasury securities. In the same press release, the FOMC also announced that it would begin reinvesting principal payments from agency debt and agency MBS instead of reinvesting the proceeds across the Treasury curve, further increasing the expected duration of the SOMA portfolio.

In many respects, this policy is perhaps the cleanest treatment effect to consider when studying the duration-risk channel of Vayanos and Vila (2021). Moreover, as with the prior LSAP date, this September 21, 2011 announcement was not accompanied by revised forward guidance over future interest rates. One shortcoming of this date is that it may have been somewhat anticipated by the time it was announced. Nevertheless, D'Amico and Seida (2024) provide evidence that the MEP announcement was larger than markets expected and Krishnamurthy and Vissing-Jorgensen (2011) document a significant yield decline on the day of the announcement.[14] Finally, reports from the financial press further indicate that the Fed's duration-extension from the MEP was larger than expected.

By restricting the sign of the accommodation shock on these dates, we require the structural VAR model to be consistent with the above narrative evidence that these two LSAP announcements led to an increase in the expected duration of the Fed's SOMA portfolio. We now turn to the narrative restriction that we apply to credibly identify the market functioning shock.

Assumption 5: Narrative Sign Restriction 2: *The market functioning shock ($j = 1$)* ***was positive*** *on March 25, 2020 and it was* ***the overwhelming driver*** *of SOMA balances that week.*

This falls under a different class of narrative sign restrictions from what we have discussed thus far. Rather than restricting the sign of the structural shock or the IRF, our assumption that the market functioning shock was the overwhelming driver of SOMA TYE balances conveys information on the relative magnitude of the contribution of that shock on that date, restricting the HD. Antolín-Díaz and Rubio-Ramírez (2018) provide two versions for HD restrictions: *Type A* and *Type B*.

[14]Respondent to the FRB New York's Desk survey of primary dealers expected a $325 billion MEP with a 78% probability. However, the new agency MBS reinvestment policy, coupled with the details of the MEP, brought the combined duration effects somewhat above that average expected duration. D'Amico and Seida (2024) show that the duration surprise of the MEP announcement was substantial — over $100 billion.

In our setup, *Type A* specifies that the absolute value of the contribution of the market functioning shock to the unexpected change in SOMA TYE is larger than the absolute value of the contribution of any other structural shock on the week of March 25, 2020 — thus, *Type A*'s interpretation is ***"the most important driver."*** *Type B* specifies that the absolute value of the market functioning shock's contribution to the unexpected change in SOMA TYE is larger than the absolute sum of the absolute value of the contributions of all other structural shocks on the week of March 25, 2020 — thus, *Type B*'s interpretation is ***"the overwhelming driver."*** *Type B* is more restrictive than *Type A*, reflecting our confidence that market functioning was the predominant driver of SOMA TYE in late March 2020.[15]

While this is a strong restriction, it is well supported by the narrative evidence. In particular, the pace of Fed asset purchases during the week of March 18-March 25 was unprecedented. Vissing-Jorgensen (2021) documents the daily pace of the Fed's Treasury purchases increased to about $70 billion on March 19 and the pace of the Fed's agency MBS purchases increased to almost $40 billion on March 20, each staying near this pace through April 1. Purchases were accelerated as earlier attempts to stabilize bond markets in the prior weeks proved insufficient, including a March 15, 2020 FOMC announcement of at least $500 billion in Treasury purchases and $200 billion in MBS purchases "over coming months."

By March 23, 2020, the FOMC announced that purchases would instead continue in "*the amounts needed*" to support market functioning. The increase in SOMA TYE the week of March 25 was nearly $200 billion, the largest ever before that week and only surpassed since then by the following week's purchases totaling $230 billion. This evidence supports our assumption that the market functioning shock was the overwhelming driver of the increase in SOMA balances that week.

3 Empirical Results

This section presents results from our two identification strategies: the baseline sign and zero specification — based on Assumptions 1-3 above — and the narrative sign restriction scheme — based on the additional Assumptions 4 and 5 above. Figure 5 shows the posterior distributions of our two shocks of interest around the three dates that we claim are key for

[15] Antolín-Díaz and Rubio-Ramírez (2018) explain that whether Type A or Type B is more restrictive depends on whether the contribution of a particular shock to the unexpected change in a variable is to be "larger" or "smaller." While, for our case, *Type B* is more restrictive, we find little difference in the results if we instead use the weaker Type A restriction.

distinguishing them as fundamentally different shocks. Figure 6 shows the impulse responses to a market functioning shock and Figure 7 shows responses to an accommodation shock. It is worth reiterating that both shocks are identified by restricting SOMA TYE exclusively. We impose no restrictions on the remaining variables in the system, letting the data speak freely on any yield and liquidity dynamics.

3.1 Are the Dynamics Around the Three Key Events Informative?

We start by examining the implications of the baseline specification for the week around November 26, 2008. The light gray histogram in panel (a) of Figure 5 displays the posterior distribution of the accommodation shock during that week. More than 50% of the probability mass is negative implying a high probability of a *reduction* in expected future SOMA TYE despite the announcement that launched the first ever LSAP program that took place that same week.

The counterfactual interpretation of the structural VAR model without narrative restrictions demonstrates the appeal of directly imposing narrative restrictions in our identifying assumptions. Otherwise, Assumptions 1-3 fail to rule out structural parameters that contradict clear evidence of an expansionary LSAP shock occurring in the week of November 26, 2008.

The baseline restrictions also lead to contradictory evidence for the second key date we consider in September 2011. Figure 5(b) shows approximately equal probabilities of our identified accommodation shock being expansionary or contractionary on September 21, 2011. The light gray histogram in panel (b) of this figure shows the posterior distribution is evenly bisected between a positive and a negative shock to expected SOMA duration during that month. Thus, the baseline result seems largely inconclusive as to whether the September 21, 2011 MEP announcement increased expected future SOMA TYE. This ambiguity is difficult to reconcile with the narrative evidence that the MEP announcement, while somewhat anticipated, drove a rally in longer-dated Treasuries.

Specifically, the September 21, 2011 Wall Street Journal article titled *In a Twist, Long-Dated Bonds Soar* noted that the "surprise Wednesday came from the details of the buying, with the Fed planning to buy a bigger-than-expected 29% on Treasurys maturing in 20 to 30 years." This narrative evidence indicates that the duration-adjusted size of the MEP was larger than expected, consistent with the evidence in D'Amico and Seida (2024) based on market surveys, and hence driving an increase in expected SOMA TYE over the next year.

Altogether — as displayed by the light gray histograms in Figure 5(a) and Figure 5(b) — the baseline specification seems inconsistent with the historical record of the Fed's LSAPs. Buttressed by our narrative evidence, Assumption 4 alone drives the posterior densities to display a strictly positive probability mass, as highlighted by the dark blue histograms in Figure 5(a) and Figure 5(b). This single additional assumption sharpens the characteristic of our identified shocks as a strictly positive shock to SOMA TYE around the first ever LSAP program as well as the MEP episode.

Finally, we inspect the posterior density of the market functioning shock on March 25, 2020 both without and with the narrative restrictions. The light gray histogram in Figure 5(c) suggests that without narrative restrictions, a majority of the probability mass lands on positive values for the market functioning shock on March 25, 2020. However, much of the posterior density falls into the negative territory or implies implausibly small increases in market functioning purchases. Recall that this week saw the largest single one week SOMA TYE increase on record up to that date. The unprecedented pace of Fed purchases that began on March 19 was explicitly targeted at restoring market functioning.

> *"The Committee directs the Desk to increase the System Open Market Account holdings of Treasury securities and agency mortgage-backed securities (MBS) in the amounts needed to support the smooth functioning of markets for Treasury securities and agency MBS."*
>
> March 23, 2020 FOMC Statement.[16]

Based on this narrative evidence, our Assumption 5 above drives the blue histogram in Figure 5(c) to a concentrated and overwhelmingly positive probability mass. The posterior density of the market functioning shock on the week of March 25, 2020 under the narrative restrictions implies a nearly 10 standard deviation market functioning purchase shock, which is congruent with the historic balance sheet expansion for the purpose of market functioning that week.

[16] https://www.federalreserve.gov/newsevents/pressreleases/monetary20200323a.htm

3.2 Responses to Accommodation and Market Functioning Shocks

We now turn to IRFs from the two specifications we consider: the baseline sign and zero restrictions and the specification with additional narrative restrictions. Figure 6 and Figure 7 show responses to our identified market functioning and accommodation shocks, respectively. For the baseline identification (Assumptions 1-3), each figure shows a dashed line corresponding with the point-wise median IRFs enveloped by a light shaded gray area that represents the 68 percent (point-wise) highest posterior density (HPD) credible sets. The point-wise median responses corresponding to the narrative restrictions (Assumptions 1-5) are shown in the solid blue lines — along with the darker blue shaded HPD credible sets.

Our narrative restrictions drastically reshape and sharpen our estimated impulse response functions to LSAP shocks. While Figure 5 shows how our additional narrative restrictions (Assumptions 4 and 5 above) hone the SOMA TYE innovations, it is worth highlighting these restrictions bind on a very small share of our sample. Our assumptions impose restrictions on the sign of the accommodation shock (Assumption 4) for two dates and the sign and contribution of the market functioning shock (Assumption 5) for a single date. Altogether, these restrict the parameter space for a sum total of three observations out of 985 weeks in our sample. Therefore, there is no econometric guarantee that these three dates could offer any enhancement over our baseline identification in a better understanding of the effects of these two shocks. Nevertheless, we find our narrative restrictions are instrumental in providing a clear picture of the effects of LSAPs. In fact, without the narrative assumptions, Figures 6 and 7 reveal that it would be difficult to make any inference about the effects of LSAPs, let alone the differential effects of market functioning and accommodation purchases.

3.2.1 Market Functioning Shocks

Shocks to market functioning lead to a rapid improvement in market liquidity alongside the expansion in SOMA TYE. Figure 6 shows that market functioning purchases increase SOMA duration-adjusted balances on impact, then ramp up over the following weeks, peaking at a 2% increase in SOMA TYE about three months after the shock. The accelerating pace of purchases lead to sharp reductions in both the MBS option-adjusted spread over Treasuries as well as in the Bloomberg Treasury market liquidity index.

The MBS option-adjusted spread (OAS) proxies the premium that investors require to hold an agency MBS security instead of a Treasury security with comparable cash flows.

Given that there is no meaningful default risk in either security, we interpret the spread as largely reflecting a liquidity premium. The impulse responses suggest that the market functioning purchases lead to reductions in this premium, and hence improvements in MBS market liquidity.[17] The declines in the Bloomberg Treasury market liquidity index indicate that market-functioning LSAPs lead to similar improvements in Treasury market liquidity.[18] The improvements in MBS and Treasury market liquidity largely mirror the contours of SOMA purchases. Peak liquidity improvements occur in the first three months, as SOMA balances peak, and then gradually diminish when SOMA balances normalize.

Despite an increase in duration-adjusted SOMA balances, the market functioning shock has no discernible effect on long-term interest rates in the first year. The 10-year Treasury yield falls slightly on impact as the market functioning purchases increase, though the effect is small at just one basis point and the credible set includes zero, suggesting no statistically significant decline in longer-term Treasury yields for the first year. After the first year, once the market has returned to functioning nominally, a lagged effect of this shock serves to reduce the 10-year Treasury yield at longer horizons. However, the effects of market functioning shocks remain concentrated on market liquidity metrics.

3.2.2 Accommodation Shocks

Accommodation LSAP shocks lead to immediate and persistent declines in longer-term Treasury yields alongside a significantly delayed increase in duration-adjusted SOMA balances. Figure 7 shows that accommodation shocks have no immediate impact on SOMA TYE balances, per our zero restriction (Assumption 3). However, the SOMA TYE response remains statistically at zero between weeks two to 13 post shock, which is not by assumption. Since our zero restriction only binds on impact (week 0), the dynamics of the responses are driven by the data. Over time, SOMA TYE balances increase by about 1% and are highly persistent. We interpret our estimates as capturing the common Fed practice of announcing accommodation purchases several weeks or even months before they are implemented.

Despite the delayed rise in SOMA TYE, the response of longer-term interest rates to the accommodation shock is immediate. The 10-year Treasury yield declines by about 10 basis points on impact, with the decline well-identified according to the width of the posterior

[17] The MBS OAS spread could also reflect model uncertainty associated with the forecast of mortgage pre-payments. However, it is not obvious how LSAPs would reduce this uncertainty, hence we attribute the declines primarily to improvements in MBS market liquidity.

[18] Recall, lower values of the index correspond to better liquidity in the Treasury market and vice versa.

credible sets. The response of the 10-year Treasury yield is highest on impact, though the yield remains depressed for years according to the impulse response which is negative and statistically significant at all horizons. It is remarkable how much the narrative restrictions dramatically sharpen the estimated response of the 10-year Treasury yield. Overall, despite using a fundamentally different empirical strategy that is focused on duration-adjusted balances, our estimated SOMA and yield dynamics to an accommodation shock are consistent with the QE event study literature. This literature has shown that markets price in changes in Fed balance sheet policy to asset prices well before any purchases are implemented.

Our estimates of the effects of accommodation shocks reveal little to no impact on market liquidity. The response of the MBS option-adjusted spread is not statistically significant. The response of the Bloomberg Treasury market liquidity index is similarly muted and shows no statistically significant change for the first year or so. As SOMA balances persist, there is some gradual improvement in Treasury market liquidity, as well as in MBS market liquidity, though the improvements in MBS market liquidity are marginal.

3.2.3 Shock Comparison

Despite market functioning and accommodation shocks both increasing duration-adjusted SOMA balances, they have markedly different effects on longer-term interest rates. This finding challenges the standard Vayanos and Vila (2021) model, which predicts that shifting duration risk on to the central bank's balance sheet should always lower interest rates. However, we find that — unlike accommodation LSAPs — market functioning purchases are estimated to have economically insignificant effects on longer-term Treasury yields. The effects of accommodation shocks dominate the response of the 10-year Treasury yield at all horizons relative to market functioning shocks.

Nevertheless, market functioning purchases are not ineffectual. To the contrary, we find that market functioning shocks have powerful effects on market liquidity — effects that are far less pronounced under accommodation purchases. These empirical conclusions are largely insensitive to alternative identification strategies (see Appendix B) and substitute measures of liquidity or long-term yields. Together, we take these findings as evidence that market functioning and accommodation LSAPs transmit through fundamentally different channels.

4 A Model of Market-Functioning LSAPs

To conceptually understand how operationally similar asset purchases can lead to markedly different dynamics for yields and liquidity conditions, we advance a partial equilibrium model of an intermediated bond market. Our model provides a distinct role for market functioning purchases in addition to the duration risk channel of Vayanos and Vila (2009, 2021). We build on the discrete time version of the Vayanos and Vila framework, as refined in Greenwood et al. (2024), all of which we henceforth refer to as VV. We briefly sketch out the key elements of the VV framework — and where we depart from it — before developing the model and analyzing the comparative statics in more detail.

In the standard VV model, heterogeneous investors trade default-free discount bonds in fixed supply of various maturities $\tau = 1, 2, ..., T$.[19] Two types of investors trade bonds directly with one another: a *preferred habitat agent* and an *arbitrageur*. The preferred habitat agent (henceforth **PH**) is a specialized investor trading along particular sections of the term structure, which includes the central bank in the VV setup. Arbitrageurs are the marginal investor absorbing total bond supply net of the PH investor.[20] Arbitrageurs are risk averse, leading to a downward-sloping demand for bond duration.

In any given period t, the arbitrageur is subject to duration risk from the τ-period bonds ($Z_t^{(\tau)}$) it holds in its portfolio, risk that increases monotonically with maturity $\tau = 1, 2, ..., T$. The price of duration risk — denoted by $\lambda_{z,t}(Z_t^{(\tau)})$ — reflects both the arbitrageur's risk aversion as well as the amount of duration risk that they hold. Changes in duration supply are, therefore, reflected in changes in $\lambda_{z,t}(Z_t^{(\tau)})$, which is the fulcrum for how yields adjust and the bond market clears.

The VV framework has become the leading model for understanding how central bank LSAPs transmit to the bond market. When a central bank purchases treasuries ($F_{b,t}^{(\tau)}$) from arbitrageurs, it removes duration risk from the arbitrageur's portfolio, thereby reducing the market price of duration. Hence, the impact of central bank purchases $F_{b,t}^{(\tau)}$ transmits solely through $\lambda_{z,t}(Z_t^{(\tau)})$. Based on this single $\lambda_{z,t}(Z_t^{(\tau)})$ risk factor, bond returns are perfectly

[19] All τ-maturity bonds have a par value of one.

[20] We build on the VV model where arbitrageurs derive utility from the market value of their holdings. Hence, prices are implicit in the market value of the arbitrageur's portfolio. However, we use par value as the unit of account because the comparative statics of interest in our model focus on the central bank's balance sheet operations which, in practice, are measured at par value. Appendix A outlines a mechanism for reconciling the market value and par value of the arbitrageur's bond holdings.

correlated across maturities.[21] Therefore, an important implication of the VV model is that an LSAP operation targeting a specific maturity will nevertheless reduce the total dollar duration held by the arbitrageur, and therefore lower the price of short rate risk used to price all long-duration assets. This means the effects of LSAPs are not merely confined to the targeted maturity but rather reduce yields across all maturities — a global term structure effect.[22]

The VV model is foundational for understanding LSAPs operating through a duration risk channel under normal trading flows. However, it is silent on central bank purchases when financial markets are under duress. The model we propose retains many desirable properties of the VV framework, but provides an explicit role for market functioning purchases.

We depart from the assumption made by VV of direct trading between **PH** and arbitrageur. Instead, we introduce a third agent that intermediates trades between the **PH** investor (including the central bank) and the arbitrageur. We designate this intermediary as the market maker (henceforth, **MM**), akin to a primary dealer. We assume the **MM** passively accepts buy/sell orders. Since the **MM** constitutes a ready buyer/seller, it supplies immediacy of liquidity services and prices these services according to a bid-ask spread. Building on work by Shen and Starr (2002), we assume the costs to the market maker of providing liquidity services increase in a convex fashion with the inventory of securities ($N_t^{(\tau)}$) that the **MM** must absorb. Assuming competitive markets, these inventory costs are recouped through the bid-ask spread that **MM**s charge.

Intermediation generates endogenous, state-dependent effects of LSAPs and introduces a role for market functioning purchases. When markets are operating in an orderly fashion, LSAPs affect bond yields through the market price of duration risk, $\lambda_{z,t}(Z_t^{(\tau)})$. However, when markets are under duress, LSAPs will support market functioning, reduce bid-ask spreads, improve liquidity, and will have no effect on the market price of duration risk.

[21] The ratio between a given τ-bond duration risk to that bond's sensitivity and the short-rate risk factor must be the same for all bonds, lest a window of opportunity opens for the arbitrageur to construct riskless arbitrage portfolios. VV argue that — while absence of arbitrage alone is not enough to guarantee constancy of $\lambda_{z,t}(Z_t^{(\tau)}) \equiv \lambda$ — "*the constant supply assumption implies the arbitrageur's bond positions are constant*" (Greenwood et al., 2024, p. 10).

[22] The majority of the SOMA portfolio is composed of U.S. treasuries, but there are also significant balances of mortgage-backed securities (MBS), and agency debt. We use the term "securities" broadly so we refer to the net supply of bonds/securities/treasuries interchangeably.

Figure 1 describes the LSAP mechanism where central bank purchases ($F_{b,t}^{(\tau)}$) expand the central bank's balance sheet ($\Delta CB_t^{(\tau)} > 0$). Under normal conditions, orderly trade flows and liquid markets keep **MM**'s beginning of period t inventories ($N_{t-1}^{(\tau+1)}$) relatively low and stable. Figure 1(a) illustrates the typical LSAP transmission under a low-inventory environment for the **MM**. Conversely, market dysfunction and one-way trading flows will lead to an unwanted security backlog and a high-inventory environment for the **MM**, which is shown in Figure 1(b). As stressed by Fleming et al. (2021), order-flow imbalances arising from the "dash for cash" event led to dysfunctional market conditions in the spring of 2020.

Figure 1(a) shows that when market flow is orderly, the **MM** absorbs little inventory and the market operates under relatively liquid conditions. In this case, the **MM** acts as a passive go-between whereby the central bank purchase removes duration from the market. This trade flow is facilitated by **MM** taking a short position to cover $F_{b,t}^{(\tau)} > 0$, selling that τ-duration to the central bank, and then replenishing its own inventory from the arbitrageur, which drains duration-adjusted balances from the arbitrageur's portfolio one-for-one ($\Delta Z_t^{(\tau)} < 0$). Our analysis below shows that — under a low-inventory environment — the removal of duration risk has an indeterminate effect on bid-ask spreads, while lowering the short-rate risk factor $\Delta\lambda_{z,t}(Z_t^{(\tau)}) < 0$, which ultimately reduces yields. As shown in our empirical analysis, the response of liquidity metrics to an accommodation shock is generally small, has an ambiguous sign depending on the metric, and is statistically insignificant. Meanwhile, the yield effects we estimate from an accommodation LSAP are negative and statistically significant.

In stress episodes where order flow has become imbalanced and the **MM** has accumulated a backlog of securities — driving up bid-ask spreads — the model operates in the high-inventory environment depicted in Figure 1(b). In this environment, the **MM** can supply the central bank's purchase $F_{b,t}^{(\tau)} > 0$ from its own inventory without replenishing it with purchases from the arbitrageur. Thus, LSAPs can reinstate balance to order flow, reducing dealer inventories, and thereby lower bid-ask spreads. Because arbitrageurs' portfolios are insulated from the purchase, $\Delta\lambda_{z,t}(Z_t^{(\tau)}) = 0$, leaving bond yields unaffected. These effects broadly match what we estimate from a market functioning shock.

Therefore, depending on order flow balance or imbalance, our model captures the traditional duration-risk channel of asset purchases outlined in Figure 1(a) and a novel market functioning channel described in Figure 1(b).

We next describe the behavior of our three agents: arbitrageurs, market makers (**MM**), and preferred habitat investors (**PH**). We distinguish the central bank (henceforth, **CB**) from other preferred habitat investors collectively grouped under **PH**.[23] We sketch out the core equations to derive the comparative statics leaving the full derivation for Appendix A.[24]

4.1 Arbitrageurs

The log price of the τ-period bond at time t is denoted by

$$\log(P_t^{(\tau)}) = p_t^{(\tau)} = -\tau y_t^{(\tau)}, \tag{7}$$

where $y_t^{(\tau)}$ is the yield on default-free discount bonds of maturities $\tau = 1, 2, \ldots, T$. The short rate r_t is the one-period bond ($r_t = y_t^{(1)}$), evolving according to an AR(1) process:

$$r_{t+1} = \bar{r} + \rho_r(r_t - \bar{r}) + \sigma_r \varepsilon_{r,t+1}, \tag{8}$$

where $\rho_r \in (0,1)$, $\bar{r}$, and $\sigma_r > 0$ are constants, and $\varepsilon_{r,t+1}$ is normally distributed with $E_t[\varepsilon_{r,t+1}] = 0$ and $\mathrm{Var}_t[\varepsilon_{r,t+1}] = 1$. We assume a constant bond supply making the short rate the sole source of uncertainty in bond pricing.

Arbitrageurs participate in a perfectly competitive market forming a continuum of securities investors with unit mass that hold $Z_t^{(\tau)}$ positions of τ-period bonds. At time t, they choose a portfolio of these securities to trade off next period's mean and variance of their wealth with coefficient of risk aversion $a \geq 0$. A full solution of this problem — which is available in Appendix A — shows the yield for maturity τ can be expressed as:

$$y_t^{(\tau)} = \frac{1}{\tau}\left[A_r^{(\tau)}(r_t - \bar{r}) + C_t^{(\tau)}\right], \tag{9}$$

where $A_r^{(\tau)}$ denotes the sensitivity of bond prices to the short rate factor r_t,[25] and $C_t^{(\tau)}$[26] is independent of r_t, but dependent on the equilibrium price of short-rate risk $\lambda_{z,t}(Z_t^{(\tau)})$ — a time-varying endogenous variable, which evolves according to:

$$\Delta C_t^{(\tau)} = \left[\sum_{j=1}^{\tau} A_r^{(j-1)}\right] \Delta\lambda_{z,t}(Z_t^{(\tau)}). \tag{10}$$

[23] These may include pension funds, sovereign-wealth funds, other index funds, as well as foreign central banks and other institutional investors who may typically demand long-duration bonds.

[24] We keep to the VV notation to the extent possible.

[25] $A_r^{(\tau)}$ is a constant that depends only on maturity τ and other structural parameters, which we specify fully in Appendix A.

[26] $C_t^{(\tau)}$ captures the term structure effects, which stem from mean reversion, volatility, and risk.

The parameter $\lambda_{z,t}(Z_t^{(\tau)})$ in equation (10) is the price of short-rate risk, which is a function of the supply of τ-period bonds ($Z_t^{(\tau)}$) arbitrageurs absorb in equilibrium. This short-rate factor is given by:

$$\Delta\lambda_{z,t}(Z_t^{(\tau)}) \equiv \frac{a\sigma_r^2}{\Phi} \sum_{\tau=2}^{T} \Delta Z_t^{(\tau)} A_r^{(\tau-1)}, \tag{11}$$

where $\Phi \geq 1$ is a constant function of structural parameters, all of which are non-negative under standard assumptions.[27]

This key $\Delta\lambda_{z,t}(Z_t^{(\tau)})$ parameter determines how net bond supply affects bond yields according to:

$$\Delta y_t^{(\tau)} \;=\; \frac{a\sigma_r^2}{\tau\Phi}\left[\sum_{j=1}^{\tau} A_r^{(j-1)}\right] \cdot \sum_{\tau=2}^{T} \Delta Z_t^{(\tau)}\, A_r^{(\tau-1)}, \tag{12}$$

where all parameters in (12) are nonnegative. This state-dependent equilibrium yield condition (12) shows that a reduction in $Z_t^{(\tau)}$ lowers the τ-bond yield through the amount of duration risk arbitrageurs hold.[28] Since $\tau^{-1}\sum_{j=1}^{\tau} A_r^{(j-1)}$ in (12) increases in τ, long-term bonds are more responsive than short-term bonds with more substantial term structure effects at higher maturities.[29]

4.2 Market Makers

As intermediaries, **MM**s provide liquidity services by ensuring the immediate ability to transact. **MM**s operate in competitive markets and earn zero expected profits in equilibrium — a condition that stems from competitive market forces, not from any optimization strategy by the **MM**. Accordingly, we assume **MM**s do not actively maximize profits in their market making activities.[30] To facilitate this immediacy, we further assume: *i*) the **MM** has a desired non-negative inventory lower bound of τ-period bonds ($\bar{N}^{(\tau)} > 0$) that is

[27]We provide additional discussion on how Φ affects the transmission from $\Delta Z_t^{(\tau)}$ to $\Delta\lambda_{z,t}(Z_t^{(\tau)})$ in Appendix A.

[28]More specifically, a decrease in $Z_t^{(\tau)}$ lowers yields because it removes $\left(\sum_{\tau=2}^{T} \Delta Z_t^{(\tau)} A_r^{(\tau-1)}\right)$ duration from arbitrageurs' portfolio, which lowers the price of short-rate risk by $\Delta\lambda_{z,t}(Z_t^{(\tau)})$ according to equation (11). The comparative statics in Section 4.4 apply equation (12) endogenously, recovering state-dependent yield responses.

[29]A reduction in the available market supply removes duration risk from the balance sheet of the arbitrageurs. The resulting decline in $\lambda_{z,t}(Z_t^{(\tau)})$ lowers yields at all maturities through the risk compensation arbitrageurs require.

[30]This distinguishes our **MM** from profit-maximizing dealers and other types of **MM**s who — due to more aggressive risk management practices of their trading desks over those of more passive intermediators — may have tighter upper and lower bounds on their day-to-day positions.

constant period-to-period and provides float, and *ii*) the **MM** may carry significant positions overnight and across periods.[31]

At any period t, the **MM** provides $V_{b,t}^{(\tau)}$ units to **PH** buyers and purchases $V_{s,t}^{(\tau)}$ units from **PH** sellers of τ-period bonds for $\tau = 1, 2, \ldots, T$. The **MM** must also fill central bank purchases $F_{b,t}^{(\tau)}$ for monetary policy purposes. Since the **MM** intermediates passively, the volumes of long ($V_{b,t}^{(\tau)}$) and short ($V_{s,t}^{(\tau)}$) orders are stochastic and independent. We assume balanced flows — on average — so that $V_{b,t}^{(\tau)}$ and $V_{s,t}^{(\tau)}$ have identical means $V^{(\tau)}$ and variances σ_V^2.[32] The **MM**'s τ-period net bond position evolves according to:

$$N_t^{(\tau)} = N_{t-1}^{(\tau+1)} - V_{b,t}^{(\tau)} + V_{s,t}^{(\tau)} + Q_{s,t}^{(\tau)} - F_{b,t}^{(\tau)}, \tag{13}$$

where $N_t^{(\tau)}$ is the **MM**'s end-of-period inventory of τ-period bonds and $N_{t-1}^{(\tau+1)}$ is an endogenous state variable taken as given in period t. We impose $N_t^{(\tau)} \geq \bar{N}^{(\tau)} \geq 0$, for all t and τ. The **MM** purchases $Q_{s,t}^{(\tau)} \geq 0$ bonds from the arbitrageur to satisfy this constraint. Of all securities in **MM**'s order book, only $Q_{s,t}^{(\tau)}$ are chosen by the **MM** — all others flowing exogenously from **PH** investors and the central bank.

We assume that liquidity services are costly to provide. Following Shen and Starr (2002), the **MM**'s costs are quadratic in end-of-period inventory I_t, which is defined as:

$$I_t = b \sum_{\tau=1}^{T} \left[P_t^{(\tau)} N_t^{(\tau)} \right]^2. \tag{14}$$

The term b captures regulatory, compliance, and financing costs. Intuitively, convex costs could materialize if rising inventories are financed at increasing cost through more internal capital or rising leverage. Empirically, Duffie et al. (2023) documents that in the U.S. Treasury market, bid-ask spreads are convex in primary dealer inventory balances, consistent

[31] For example, Adrian et al. (2020) study profit-seeking principal trading firms (PTFs) who actively trade intraday but must always close out the day flat. In principle, there may also be upper bounds on inventory positions of market making. Li et al. (2026) argue limits on inventory positions are often imposed on **MM**'s trading desk by the dealer's risk management team, which effectively curbs their risk absorption capacity. They contend these limits have to be predetermined and difficult to change, often requiring multiple levels of approval. Bräuning and Stein (2026) also consider firm-specific trading desk value-at-risk limits as well as the more aggregate supplementary leverage ratio constraints on the inventory carrying capacity of **MM**s. They estimate these limits may cost **MM**s as high as 9% of dealer's profit margins.

[32] Unlike the stochastic **PH** order flows $V_{b,t}^{(\tau)}$ and $V_{s,t}^{(\tau)}$, central bank purchases $F_{b,t}^{(\tau)}$ are treated as deterministic policy interventions, known to the **MM** at the time of execution.

with our quadratic specification. The **MM**'s net cash position evolves according to:

$$M_t = \sum_{\tau=1}^{T}(1 + S_t^{(\tau)})P_t^{(\tau)}V_{b,t}^{(\tau)} - \sum_{\tau=1}^{T}(1 - S_t^{(\tau)})P_t^{(\tau)}V_{s,t}^{(\tau)} + \sum_{\tau=1}^{T}(1 + S_t^{(\tau)})P_t^{(\tau)}F_{b,t}^{(\tau)} - \sum_{\tau=1}^{T}(1 - S_t^{(\tau)})P_t^{(\tau)}Q_{s,t}^{(\tau)} + M_{t-1} - I_t, \tag{15}$$

where $S_t^{(\tau)}$ is the bid-ask spread of the τ-period bond. Competitive market making implies that the **MM**'s expected net asset position is stable period-to-period as follows:

$$M_t + \sum_{\tau=1}^{T} P_t^{(\tau)}N_t^{(\tau)} = M_{t-1} + \sum_{\tau=1}^{T-1} P_{t-1}^{(\tau+1)}N_{t-1}^{(\tau+1)}. \tag{16}$$

Substituting equations (13), (14), and (15) into the zero profit condition in equation (16) pins down the bid-ask spread as a function of the **MM**s' inventory position, order flow, and bond prices:[33]

$$S_t^{*(\tau)}P_t^{(\tau)}\left(\mathbb{E}_t[V_{b,t}^{(\tau)}] + \mathbb{E}_t[V_{s,t}^{(\tau)}] + F_{b,t}^{(\tau)} + Q_{s,t}^{(\tau)}\right) = b\left[P_t^{(\tau)}N_t^{(\tau)}\right]^2 - N_{t-1}^{(\tau+1)}\left(P_t^{(\tau)} - P_{t-1}^{(\tau+1)}\right). \tag{17}$$

The left-hand side of (17) is intermediation revenue and the right-hand side is net inventory carrying cost — reflecting convex inventory costs less price appreciation on beginning-of-period positions.[34] Spreads are decreasing in order flow — $\mathbb{E}_t[V_{b,t}^{(\tau)}] + \mathbb{E}_t[V_{s,t}^{(\tau)}] + F_{b,t}^{(\tau)} + Q_{s,t}^{(\tau)}$ — as higher trading volumes allow **MM**s to spread carrying cost across more transactions, lowering the per-transaction spread. Bid-ask spreads are increasing and convex in current inventories $N_t^{(\tau)}$.

There is a tension that leads bond price changes to have an indeterminate effect on bid-ask spreads. Higher prices boost trading revenue and inventory returns — which lower spreads — but also raise the market value of current inventories and hence inventory costs, exerting upward pressures on spreads.

4.3 Market Clearing

Bond market clearing requires that the total fixed supply of bonds $\Omega^{(\tau)}$ is absorbed by preferred habitat investors ($PH_t^{(\tau)}$), the central bank ($CB_t^{(\tau)}$), market makers ($N_t^{(\tau)}$), and

[33] We assume that b is sufficiently large so that spreads are positive.

[34] Having taken $\mathbb{E}_t[\cdot]$, all terms in the expected spread condition are realized at time t, and the expectation binds only on the stochastic order flow $V_{b,t}^{(\tau)}$, and $V_{s,t}^{(\tau)}$. We assume the market maker sets the spread prior to observing realized order flow $V_{b,t}^{(\tau)}$, and $V_{s,t}^{(\tau)}$.

arbitrageurs ($Z_t^{(\tau)}$), implying that:

$$\Omega^{(\tau)} = PH_t^{(\tau)} + CB_t^{(\tau)} + N_t^{(\tau)} + Z_t^{(\tau)}, \tag{18}$$

for each $\tau = 1, ..., T$.

The stock of holdings of each buyer evolves according to:

$$PH_t^{(\tau)} = PH_{t-1}^{(\tau+1)} + V_{b,t}^{(\tau)} - V_{s,t}^{(\tau)}, \tag{19}$$

$$CB_t^{(\tau)} = CB_{t-1}^{(\tau+1)} + F_{b,t}^{(\tau)}, \tag{20}$$

$$Z_t^{(\tau)} = Z_{t-1}^{(\tau+1)} - Q_{s,t}^{(\tau)}, \tag{21}$$

and the stock of $N_t^{(\tau)}$ evolves according to equation (13).

4.4 Comparative Statics

We examine how inventory balances and order flow conditions shape the impact of central bank asset purchases specified as fixed one-time unexpected shifts in $F_{b,t}^{(\tau)}$. The model's dynamics for bond yields and spreads in response to these purchases differ qualitatively depending on whether inventories are at the constraint ($N_t^{(\tau)} = \bar{N}^{(\tau)}$) or slack ($N_t^{(\tau)} > \bar{N}^{(\tau)}$).

We consider two cases. If the **MM** holds low inventories, it replenishes the central bank purchase via a purchase from the arbitrageur ($Q_{s,t}^{(\tau)} > 0$) . If instead inventories are high, the **MM** draws down its own inventory without replenishment to meet the central banks purchase ($Q_{s,t}^{(\tau)} = 0$). Appendix A demonstrates that the **MM**'s solution to the unconstrained replenishment problem is given by:

$$Q_{s,t}^{(\tau)*} = max\left\{\bar{N}^{(\tau)} - N_{t-1}^{(\tau+1)} + F_{b,t}^{(\tau)}, 0\right\}. \tag{22}$$

4.4.1 Low Inventory — MM Replenishes Inventories: $Q_{s,t}^{(\tau)*} > 0$

When **MM**'s beginning-of-period inventory $\left(N_{t-1}^{(\tau+1)}\right)$ falls short of its lower bound $\left(\bar{N}^{(\tau)}\right)$ net of any central bank purchase $\left(F_{b,t}^{(\tau)}\right)$ such that:

$$N_{t-1}^{(\tau+1)} \leq \bar{N}^{(\tau)} + F_{b,t}^{(\tau)}, \tag{23}$$

and the unconstrained optimum satisfies $Q_{s,t}^{(\tau)*} > 0$ (see Appendix A), so the replenishment constraint binds at:

$$Q_{s,t}^{(\tau)*} = \bar{N}^{(\tau)} - N_{t-1}^{(\tau+1)} + F_{b,t}^{(\tau)} > 0. \tag{24}$$

Condition (23) holds even absent a central bank purchase ($F_{b,t}^{(\tau)} = 0$), requiring **MM** to still purchase bonds from the arbitrageur to restore its inventory to the lower bound. In addition, when $F_{b,t}^{(\tau)} > 0$, **MM** takes an intraday short position to fill the order, then covers with a larger replenishment purchase encompassing both the central bank shortfall and the inventory gap. Substituting (24) into the arbitrageur stock-flow equation (21) yields:

$$\Delta Z_t^{(\tau)} \;=\; Z_t^{(\tau)} - Z_{t-1}^{(\tau+1)} \;=\; -Q_{s,t}^{(\tau)*} \;=\; -\left(\bar{N}^{(\tau)} - N_{t-1}^{(\tau+1)} + F_{b,t}^{(\tau)}\right) < 0, \tag{25}$$

and substituting (25) into the state-dependent equilibrium yield condition (12) results in:

$$\Delta y_t^{(\tau)} = -\frac{a\sigma_r^2}{\tau\Phi}\left[\sum_{j=1}^{\tau} A_r^{(j-1)}\right] \cdot \sum_{\tau=2}^{T}\left(\bar{N}^{(\tau)} - N_{t-1}^{(\tau+1)} + F_{b,t}^{(\tau)}\right) A_r^{(\tau-1)}. \tag{26}$$

Since $A_r^{(\tau-1)} > 0$ for all $\tau \geq 2$ (see Appendix A) and $\bar{N}^{(\tau)} - N_{t-1}^{(\tau+1)} + F_{b,t}^{(\tau)} > 0$ by condition (24), every term in the sum is strictly positive. Therefore:

$$\frac{\partial y_t^{*(\tau)}}{\partial F_{b,t}^{(\tau)}} = -\frac{a\sigma_r^2}{\tau\Phi}\left[\sum_{j=1}^{\tau} A_r^{(j-1)}\right] \cdot A_r^{(\tau-1)} < 0. \tag{27}$$

The central bank purchase reduces all bond yields — even lowering yields of untargeted maturities — as arbitrageur dollar duration falls and $\lambda_{z,t}(Z_t^{(\tau)})$ declines. This is the VV duration risk channel, which is recovered here as a special case of our intermediated framework.

Under **MM**'s low-inventory environment, optimal replenishment exactly absorbs both the central bank purchase and any other residual inventory gap, implying $N_t^{(\tau)} = \bar{N}^{(\tau)}$. Appendix A shows that this condition, constraint (24), and a reformulation of (17) result in:

$$S_t^{*(\tau)} = \frac{bP_t^{(\tau)}\left[\bar{N}^{(\tau)}\right]^2 - N_{t-1}^{(\tau+1)}\left[1 - \dfrac{P_{t-1}^{(\tau+1)}}{P_t^{(\tau)}}\right]}{2V^{(\tau)} + 2F_{b,t}^{(\tau)} + \bar{N}^{(\tau)} - N_{t-1}^{(\tau+1)}}. \tag{28}$$

Appendix A proves that the following expression has indeterminate sign: $\bar{N}^{(\tau)} - N_{t-1}^{(\tau+1)} \geq -F_{b,t}^{(\tau)}$, so the denominator satisfies $2V^{(\tau)} + 2F_{b,t}^{(\tau)} + \bar{N}^{(\tau)} - N_{t-1}^{(\tau+1)} \geq 2V^{(\tau)} + F_{b,t}^{(\tau)} > 0$. Taking the partial derivative of equation (28) with respect to central bank purchases ($F_{b,t}^{(\tau)}$) and applying the quotient rule gives:

$$\frac{\partial S_t^{*(\tau)}}{\partial F_{b,t}^{(\tau)}} = \frac{\left(b\big[\bar{N}^{(\tau)}\big]^2 \frac{\partial P_t^{(\tau)}}{\partial F_{b,t}^{(\tau)}} - N_{t-1}^{(\tau+1)} \frac{\partial}{\partial F_{b,t}^{(\tau)}}\left[1 - \frac{P_{t-1}^{(\tau+1)}}{P_t^{(\tau)}}\right]\right) \cdot \Psi_t}{\Psi_t^2}$$

$$- \frac{2\left(bP_t^{(\tau)}\big[\bar{N}^{(\tau)}\big]^2 - N_{t-1}^{(\tau+1)}\,\mathrm{E}_t\left[1 - \frac{P_{t-1}^{(\tau+1)}}{P_t^{(\tau)}}\right]\right)}{\Psi_t^2}, \tag{29}$$

where $\Psi_t \equiv \left(2V^{(\tau)} + 2F_{b,t}^{(\tau)} + \bar{N}^{(\tau)} - N_{t-1}^{(\tau+1)}\right) > 0$. The spread's indeterminacy stems from the bond-price channel unique to this state of the world. Because $\Delta Z_t^{(\tau)} \neq 0$ in this low-inventory environment, the central bank purchase lowers $\lambda_{z,t}(Z_t^{(\tau)})$ and raises bond prices, which creates competing forces in carrying costs and price appreciation in the spread condition (28). This channel is absent under **MM**'s high-inventory environment, as we show next.

4.4.2 High Inventory — MM Inventory Drawdown: $Q_{s,t}^{(\tau)*} = 0$

When **MM**'s beginning-of-period inventory is elevated due to one-way trading and order flow imbalances, they can cover the central bank purchase over and above the inventory lower bound,

$$N_{t-1}^{(\tau+1)} > \bar{N}^{(\tau)} + F_{b,t}^{(\tau)}. \tag{30}$$

The unconstrained optimum requires $Q_{s,t}^{(\tau)*} < 0$, which violates feasibility. The non-negativity constraint therefore binds at:

$$Q_{s,t}^{(\tau)*} = 0. \tag{31}$$

MM absorbs the central bank purchase entirely from its own accumulated inventory, with no offsetting purchase from the arbitrageur. The buffer argument confirms this is always feasible: condition (30) in this environment implies $F_{b,t}^{(\tau)} < N_{t-1}^{(\tau+1)} - \bar{N}^{(\tau)} < N_{t-1}^{(\tau+1)}$, so the central bank purchase can never exhaust **MM**'s inventory in this state of the world. Substituting (31) into the arbitrageur stock-flow equation (21) shows that:

$$\Delta Z_t^{(\tau)} \;=\; Z_t^{(\tau)} - Z_{t-1}^{(\tau+1)} \;=\; 0, \tag{32}$$

demonstrating that arbitrageur positions are unaffected and duration risk is never transferred. Substituting (32) into the equilibrium yield condition (12) shows that $\Delta y_t^{(\tau)} = 0$, and therefore:

$$\frac{\partial y_t^{*(\tau)}}{\partial F_{b,t}^{(\tau)}} = 0. \tag{33}$$

Solving for $S_t^{*(\tau)}$ in this state of the world yields:

$$S_t^{*(\tau)} = \frac{bP_t^{(\tau)}\left[N_{t-1}^{(\tau+1)} - F_{b,t}^{(\tau)}\right]^2 - N_{t-1}^{(\tau+1)}\left[1 - \dfrac{P_{t-1}^{(\tau+1)}}{P_t^{(\tau)}}\right]}{2V^{(\tau)} + F_{b,t}^{(\tau)}}. \tag{34}$$

Taking the partial derivative of equation (34) with respect to $F_{b,t}^{(\tau)}$ shows an unambiguous negative effect. Bond prices are independent of $F_{b,t}^{(\tau)}$ since $\Delta Z_t^{(\tau)} = 0$ and $\Delta\lambda_{z,t}(Z_t^{(\tau)}) = 0$.[35] The only term in the numerator of (34) that depends on $F_{b,t}^{(\tau)}$ is the carrying cost, since $N_{t-1}^{(\tau+1)} - F_{b,t}^{(\tau)} > \bar{N}^{(\tau)} > 0$ holds throughout **MM**'s high-inventory environment. Applying the quotient rule yields:

$$\frac{\partial S_t^{*(\tau)}}{\partial F_{b,t}^{(\tau)}} = \frac{-2bP_t^{(\tau)}\left(N_{t-1}^{(\tau+1)} - F_{b,t}^{(\tau)}\right) \;-\; S_t^{*(\tau)}}{\left(2V^{(\tau)} + F_{b,t}^{(\tau)}\right)}. \tag{35}$$

Both terms in the numerator of (35) are non-positive. The first term is strictly negative since $N_{t-1}^{(\tau+1)} - F_{b,t}^{(\tau)} > 0$ in this state of the world, and the second term is non-positive since $S_t^{*(\tau)} \geq 0$, which implies the numerator of (34) is non-negative. The denominator is also positive. Therefore:

$$\frac{\partial S_t^{*(\tau)}}{\partial F_{b,t}^{(\tau)}} < 0. \tag{36}$$

The central bank purchase compresses the bid-ask spread when **MM**'s inventories are high. Because $\Delta Z_t^{(\tau)} = 0$, bond prices do not respond, lowering **MM**'s quadratic carrying cost through the direct reduction in its inventory burden $N_{t-1}^{(\tau+1)} - F_{b,t}^{(\tau)}$.

The two states are distinguished by whose balance sheet is affected by the central bank purchase. In the low-inventory state, the central bank purchases affect the arbitrageur's balance sheet by $\Delta Z_t^{(\tau)}$, which is strictly negative driving bond yields down with an indeterminate effect on spreads following a central bank purchase. Conversely, in the high-inventory state, a central bank purchase leaves the arbitrageur's portfolio unaffected ($\Delta Z_t^{(\tau)} = 0$) with no pass-through to bond yields and, instead, removes securities from the market maker's balance sheet, leading to a compression in bid-ask spreads.

[35] Given that $\partial P_t^{(\tau)}/\partial F_{b,t}^{(\tau)} = 0$ and $\partial[P_{t-1}^{(\tau+1)}]/\partial F_{b,t}^{(\tau)} = 0$, making the price appreciation term in (34) independent of $F_{b,t}^{(\tau)}$ in the high-inventory environment.

5 Implications for Balance Sheet Policy

Our stylized model provides a framework premised on balance sheet capacity for distinguishing between market functioning and accommodation LSAPs. One benefit of developing a model to interpret the data is that it sheds light on precise mechanisms, which both helps to distinguish our paper from other work and also provides some implications for balance sheet policy. We highlight four policy implications that flow from our theoretical framework.

First, our model emphasizes the importance of balance sheet constraints in shaping LSAP transmission. One closely related line of research has emphasized flow and stock effects of LSAPs transmission (D'Amico and King, 2013). For example, D'Amico et al. (2026) study the Federal Reserve's pandemic LSAPs and emphasize that flow effects were much more important in the spring and summer of 2020 and that stock effects became more important thereafter. They conclude, as we do, that the purchases in the spring of 2020, though large in duration-adjusted terms, had negligible effects on longer-term Treasury yields but materially reduced price dislocations. However, D'Amico et al. (2026) emphasize the role of Fed communications and expectations of duration risk removal in driving the differential effects between market functioning and accommodation purchases. While communication plays an important role in investor expectations for duration risk, a channel we also emphasize for accommodation shocks, our model suggests that it is not entirely at the Fed's discretion whether LSAPs affect yields or liquidity conditions. Rather, our model implies that impairments in dealer balance sheets due to order flow imbalances and one-way trading can also shape the transmission of LSAPs. In this respect, our mechanism aligns closely with the evidence in Duffie et al. (2023) who highlight the role of occasionally binding constraints on the intermediation capacity of the bond market.

Second, our model implies that in order to sustain progress in improving liquidity, market functioning purchases need to be maintained as long as the market order flow remains imbalanced. The market maker's inventory accumulation equation (13) shows that the flow of central bank purchases has to match the net flows of other investors to prevent an accumulation of securities on intermediaries' balance sheets. Thus, if selling pressures continue, spreads can widen out absent ongoing central bank purchases. The Federal Reserve's response in 2020 is a useful practical example. Even after accumulating over a $1 trillion in securities by April of 2020 *"to support smooth market functioning"*, the Fed maintained market functioning purchases through the summer *"to sustain smooth market functioning."*[36]

[36] See the March 23, 2020 FOMC Statement and the June 10, 2020 FOMC Statement, both available at:

Third, with respect to balance sheet management, the model predicts that the cumulative shift in expected-duration removal, including the holding period by the central bank, is a critical determinant of the yield effects of accommodation LSAPs. Therefore, securities purchased for accommodation likely need to be maintained on central bank balance sheets for some time to achieve the intended effects on longer-term interest rates (Potter, 2016). In contrast, securities purchased for market functioning can be sold once liquidity improves and order-flow is re-balanced without an outsized impact on liquidity conditions. This result stems from the market maker's convex cost of inventory accumulation. The Bank of England's intervention in the Gilt market in fall of 2022 provides supportive evidence that market functioning purchases can have their intended effects of restoring liquidity and can be unwound quickly once selling pressures abate (Hauser, 2022; Pinter, 2023).

Finally, another implication of our model is that, aside from maturity specific dislocations, the fiscal authority is limited in its ability to restore market functioning without retiring debt. While we do not explicitly model the liability side of the central bank's balance sheet, the ability to issue reserves to fund asset purchases endows the central bank with the capacity to offset large order flow imbalances through large-scale market functioning purchases. Therefore, unlike duration adjustments, which could in principle be carried out by central bank LSAPs or changes in government debt management, the central bank is uniquely positioned to address market functioning at scale. However, there could be other benefits to having the fiscal authority conduct market functioning purchases for reasons that are outside the scope of our model, including the separation of market functioning from other monetary policy objectives (see e.g., Duffie and Keane, 2023).

https://www.federalreserve.gov/monetarypolicy/fomchistorical2020.htm

6 Final Remarks

Characterizing the effects of LSAPs and the mechanisms through which they work are not simple endeavors. This was perhaps most poignantly underlined by former Federal Reserve Chairman Ben Bernanke when he famously quipped: "*Well, the problem with QE is it works in practice, but it doesn't work in theory.*"[37] However, the theoretical underpinnings of LSAP policy have matured since Bernanke's famous quote. Research has since emphasized that the effects of LSAPs depend largely on whose balance sheets are affected by central bank purchases (Carpenter et al., 2015).

Our work suggests that the state of dealer balance sheets is particularly important in shaping the effects of LSAPs. Under regular trading flows, large central bank purchases are effectively intermediated to remove duration risk from the market, lowering bond yields consistent with the preferred habitat theory in Vayanos and Vila (2021). However, we show that when bond market dealers are forced to absorb one-way trading flows, LSAPs act to relieve pressures on dealer balance sheets and improve liquidity with little to no yield effects. This latter mechanism provides a distinct role for market functioning purchases. By incorporating bond market intermediation into a preferred habitat model, our framework gives rise to endogenous, state-dependent effects of central bank asset purchases and is able to rationalize our empirical evidence of accommodation and market functioning shocks.

Finally, while the COVID-19 experience is highly informative for understanding market functioning dynamics, inventory constraints and trading imbalances in bond markets are not concerns isolated solely to the pandemic. As U.S. government debt issuance grows and primary dealers absorb a larger share of this issuance, there has been increased attention on potential fragilities in the U.S. Treasury market in recent years (Duffie, 2023; Kashyap et al., 2025). Given these conditions, large shocks have the potential to create renewed bond market stresses.[38] Therefore, our findings are likely to be informative for understanding and responding to any future episodes of market stress.

[37] This lack of theoretical grounding refers to the famous Wallace neutrality assertion that the size and composition of the central bank balance sheet has no effect on macroeconomic outcomes. See Brookings quote at `https://www.brookings.edu/wp-content/uploads/2014/01/20140116_bernanke_remarks_transcript.pdf`

[38] For example, Dobrev et al. (2025) show that Treasury market order flow imbalances spiked again in April 2025 amid market turbulence.

References

Adrian, T., A. Capponi, M. Fleming, E. Vogt, and H. Zhang (2020): "Intraday market making with overnight inventory costs," *Journal of Financial Markets*, 50, 100564.

Antolín-Díaz, J. and J. F. Rubio-Ramírez (2018): "Narrative sign restrictions for SVARs," *American Economic Review*, 108, 2802–2829.

Arias, J. E., D. Caldara, and J. F. Rubio-Ramírez (2019): "The systematic component of monetary policy in SVARs: An agnostic identification procedure," *Journal of Monetary Economics*, 101, 1–13.

Arias, J. E., J. F. Rubio-Ramírez, and D. F. Waggoner (2018): "Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications," *Econometrica*, 86, 685–720.

Bauer, M. and G. D. Rudebusch (2014): "The signaling channel for Federal Reserve bond purchases," *International Journal of Central Banking*.

Bhattarai, S. and C. J. Neely (2022): "An Analysis of the Literature on International Unconventional Monetary Policy," *Journal of Economic Literature*, 60, pp. 527–597.

Bräuning, F. and H. Stein (2026): "The effect of primary dealer constraints on intermediation in the Treasury market," *The Review of Financial Studies*.

Canova, F. and G. De Nicolo (2002): "Monetary disturbances matter for business fluctuations in the G-7," *Journal of Monetary Economics*, 49, 1131–1159.

Carpenter, S., S. Demiralp, J. Ihrig, and E. Klee (2015): "Analyzing Federal Reserve asset purchases: From whom does the Fed buy?" *Journal of Banking & Finance*, 52, 230–244.

D'Amico, S., W. English, D. López-Salido, and E. Nelson (2012): "The Federal Reserve's large-scale asset purchase programmes: rationale and effects," *The Economic Journal*, 122, F415–F446.

D'Amico, S., M. Gillet, S. Schulhofer-Wohl, and T. Seida (2026): "Open-Ended Treasury Purchases: From Market Functioning to Financial Easing," Staff Reports 1183, Federal Reserve Bank of New York.

D'Amico, S. and T. B. King (2013): "Flow and stock effects of large-scale treasury purchases: Evidence on the importance of local supply," *Journal of Financial Economics*, 108, 425–448.

D'Amico, S. and T. Seida (2024): "Unexpected Supply Effects of Quantitative Easing and Tightening," *The Economic Journal*.

Dobrev, D., E. Liu, J. Kim, and M. Rodriguez (2025): "Order Flow Imbalances and Amplification of Price Movements: Evidence from U.S. Treasury Markets," Feds notes, Board of Governors of the Federal Reserve System (U.S.).

Duffie, D. (2023): "Resilience redux in the US Treasury market," in *Jackson Hole Economic Policy Symposium*, Federal Reserve Bank of Kansas City.

Duffie, D., M. J. Fleming, F. M. Keane, C. Nelson, O. Shachar, and P. V. Tassel (2023): "Dealer Capacity and U.S. Treasury Market Functionality," Staff Reports 1070, Federal Reserve Bank of New York.

Duffie, D. and F. M. Keane (2023): "Market-Function Asset Purchases," Staff Reports 1054, Federal Reserve Bank of New York.

Faust, J. (1998): "The robustness of identified VAR conclusions about money," in *Carnegie-Rochester conference series on public policy*, Elsevier, vol. 49, 207–244.

Fleming, M. J., H. Liu, R. Podjasek, and J. Schurmeier (2021): "The Federal Reserve's Market Functioning Purchases," Staff Reports 998, Federal Reserve Bank of New York.

Gagnon, J., M. Raskin, J. Remache, B. Sack, et al. (2011): "The Financial Market Effects of the Federal Reserve's Large-Scale Asset Purchases," *International Journal of Central Banking*, 7, 3–43.

Greenwood, R., S. Hanson, J. S. Rudolph, and L. H. Summers (2014): "Government debt management at the zero lower bound," .

Greenwood, R., S. Hanson, and D. Vayanos (2024): "Supply and demand and the term structure of interest rates," *Annual Review of Financial Economics*, 16, 115–151.

Hanson, S. G. (2014): "Mortgage convexity," *Journal of Financial Economics*, 113, 270–299.

Hauser, A. (2022): "Thirteen days in October: how central bank balance sheets can support monetary and financial stability," Speech delivered at the ECB's 2022 Conference on Money Markets, Bank of England.

Kashyap, A. K., J. C. Stein, J. L. Wallen, and J. Younger (2025): "Treasury market dysfunction and the role of the central bank," *Brookings Papers on Economic Activity*, 2025, 221–265.

Kilian, L. and H. Lütkepohl (2017): *Structural vector autoregressive analysis*, Cambridge University Press.

Krishnamurthy, A. and A. Vissing-Jorgensen (2011): "The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy," *Brookings Papers on Economic Activity*, 43, 215–287.

Li, D., L. Petrasek, and M. Tian (2026): "Risk-averse dealers in a risk-free market—The role of trading desk risk limits," *Journal of Financial Economics*, 181, 104290.

Pinter, G. (2023): "An anatomy of the 2022 gilt market crisis," Staff Working Paper 1019, Bank of England.

Potter, S. M. (2016): "Discussion of 'Evaluating Monetary Policy Operational Frameworks' by Ulrich Bindseil," in *Jackson Hole Economic Policy Symposium*, Federal Reserve Bank of Kansas City.

Ray, W., M. Droste, and Y. Gorodnichenko (2024): "Unbundling Quantitative Easing: Taking a Cue from Treasury Auctions," *Journal of Political Economy*, 132, 3115–3172.

Shen, P. and R. M. Starr (2002): "Market-makers' supply and pricing of financial market liquidity," *Economics letters*, 76, 53–58.

Swanson, E. T. (2020): "Measuring the effects of Federal Reserve forward guidance and asset purchases on financial markets," *Journal of Monetary Economics*.

Uhlig, H. (2005): "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 52, 381–419.

Vayanos, D. and J.-L. Vila (2009): "A preferred-habitat model of the term structure of interest rates," *Mimeo*.

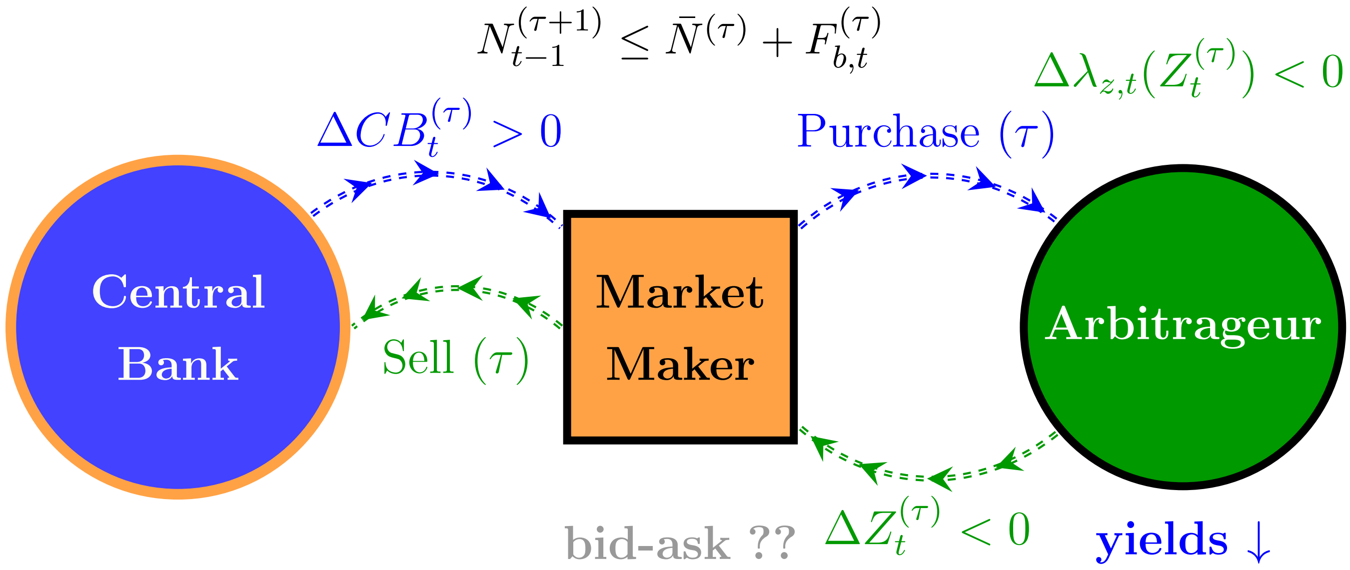
——— (2021): "A preferred-habitat model of the term structure of interest rates," *Econometrica*, 89, 77–112.

Vissing-Jorgensen, A. (2021): "The Treasury Market in Spring 2020 and the Response of the Federal Reserve," *Journal of Monetary Economics*, 124, 19–47.

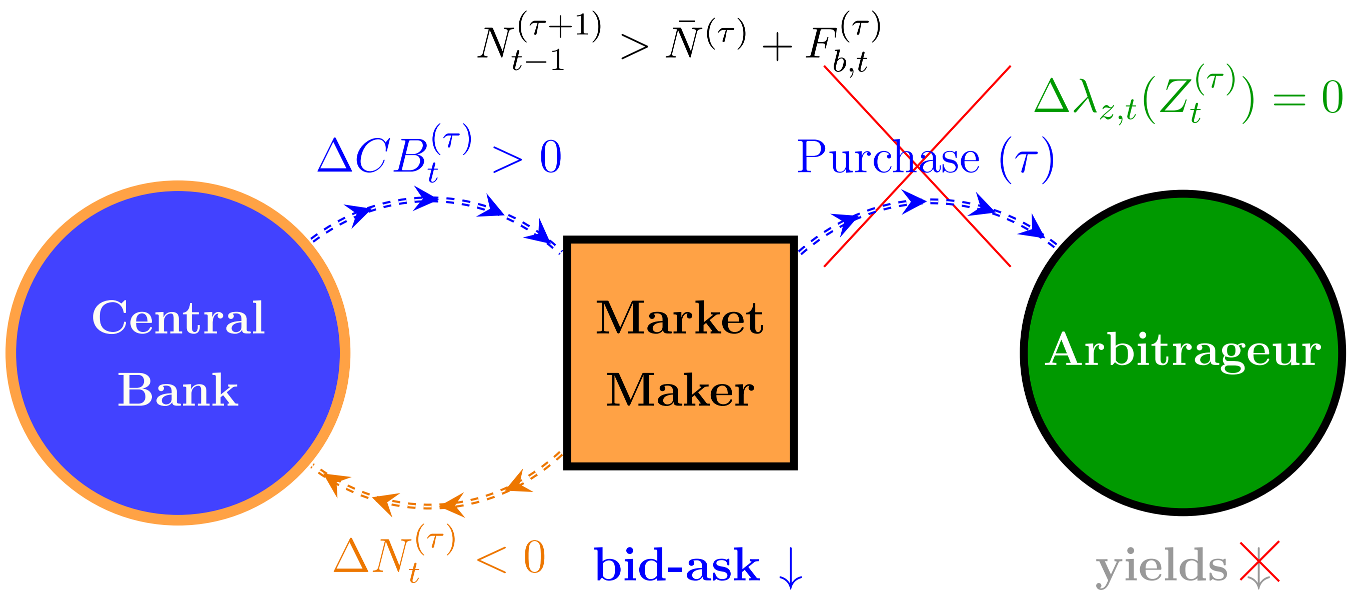
Williams, J. C. (2014): "Monetary policy at the zero lower bound: Putting theory into practice," *Hutchins Center Working Papers*.

Wolf, C. K. (2020): "Svar (mis) identification and the real effects of monetary policy shocks," *American Economic Journal: Macroeconomics*, 12, 1–32.

Woodford, M. (2012): "Methods of policy accommodation at the interest-rate lower bound," in *Proceedings-Economic Policy Symposium-Jackson Hole*, Federal Reserve Bank of Kansas City, 185–288.



(a) LSAPs ($\Delta CB_t^{(\tau)} > 0$) under **Low-Inventory** ($N_{t-1}^{(\tau+1)} \leq \bar{N}^{(\tau)} + F_{b,t}^{(\tau)}$)



(b) LSAPs ($\Delta CB_t^{(\tau)} > 0$) under **High-Inventory** ($N_{t-1}^{(\tau+1)} > \bar{N}^{(\tau)} + F_{b,t}^{(\tau)}$)

Figure 1: LSAPs under Two Different Inventory Environments

Panel (a) shows that in an orderly well-functioning trading flow environment, LSAPs ($\Delta CB_t^{(\tau)} > 0$) remove duration risk from the market by reducing arbitrageurs' Treasury positions ($\Delta Z_t^{(\tau)} < 0$), thereby transferring the duration risk from arbitrageurs' balance sheet to the central bank's—resulting in the VV-predicted expansionary effect on Treasury yields. Panel (b) shows that disruptions in trading flows in a high primary-dealer-inventory environment 'short-circuit' the expansionary effect on yields but relieve pressure and improve liquidity by lowering bid-ask spreads.

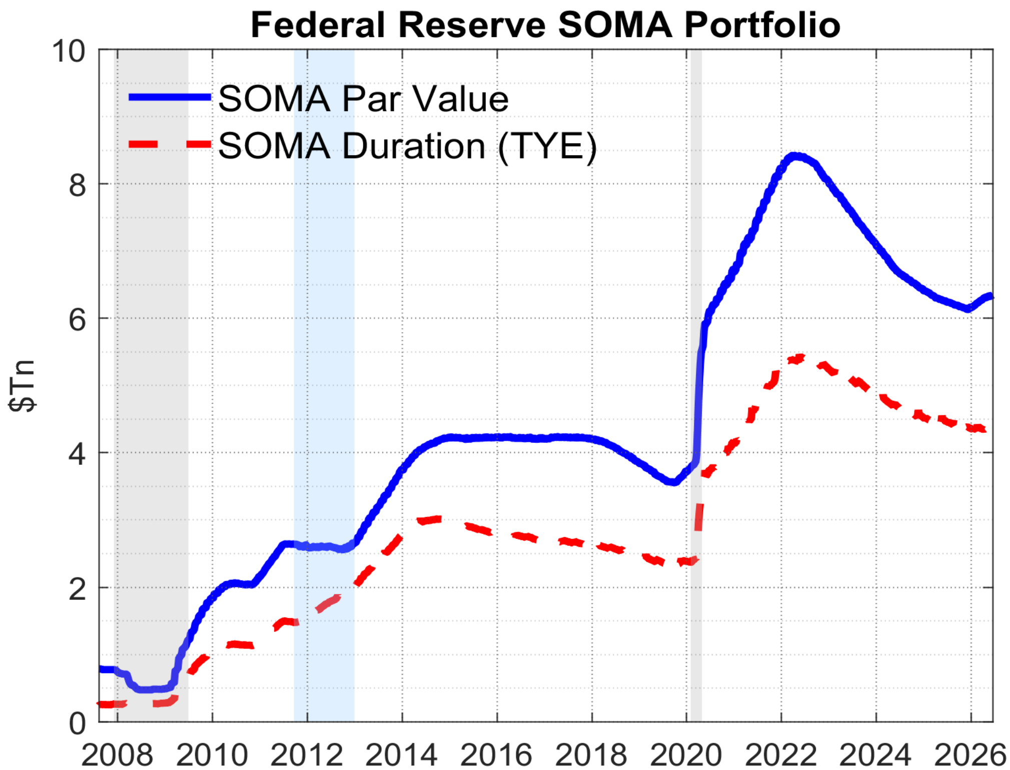


Figure 2: Par Value and 10-year-equivalent (TYE) Duration

Source: Par value from the FRBNY. Authors' calculations for TYE. Grey shading represents NBER recession dates. Lighter shading denotes the Maturity Extension Program period.

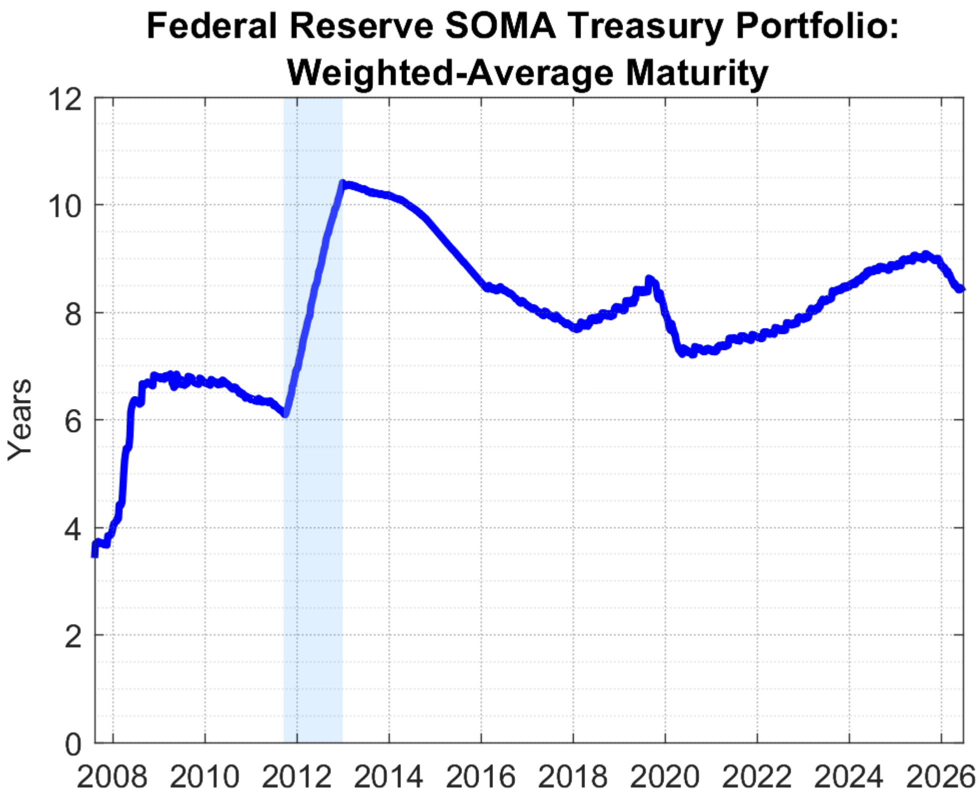
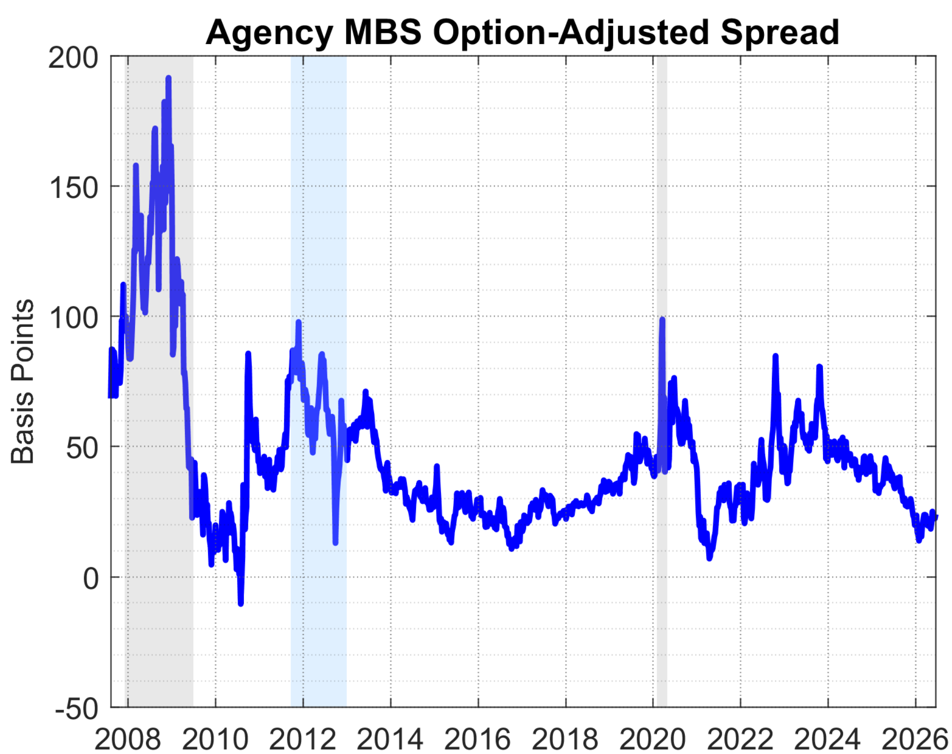


Figure 3: Weighted-Average Maturity of the Fed's SOMA Treasury Holdings

Source: FRBNY and authors' calculations (based on par value). Shading represents the Federal Reserve's Maturity Extension Program.

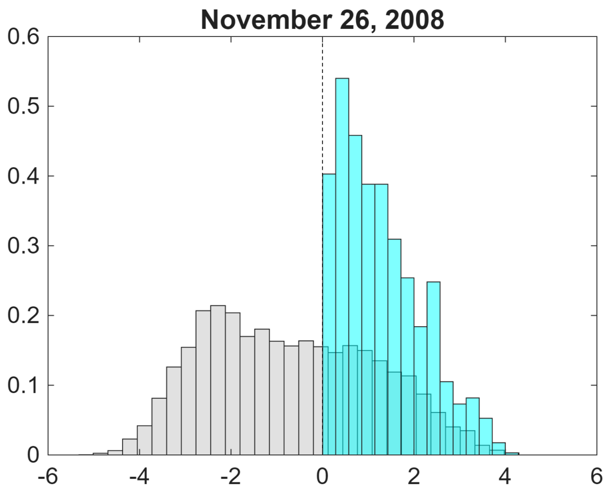
Figure 4: Aggregate Market Liquidity Measures

(a)

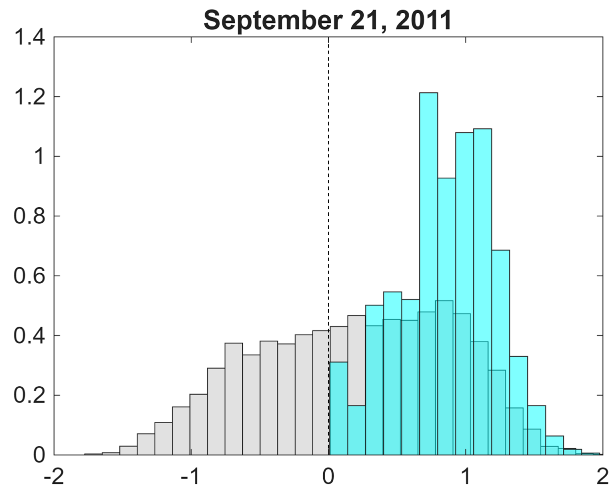


(b)

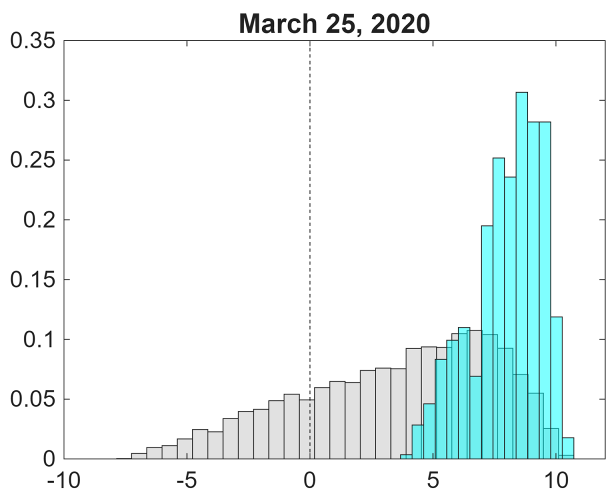
Source: Bloomberg and NBER



(a) Accommodation Shock: QE 1



(b) Accommodation Shock: MEP



(c) Market Functioning Shock: COVID-19

Figure 5: Posterior distributions of the accommodation shock [panels (a) & (b)] and the market functioning shock on panel (c) for our three key events. The lighter histograms correspond to the **baseline specification with Assumptions 1–3**. The darker histogram plots the same distribution after incorporating the **narrative restrictions with Assumptions 4 and 5**.

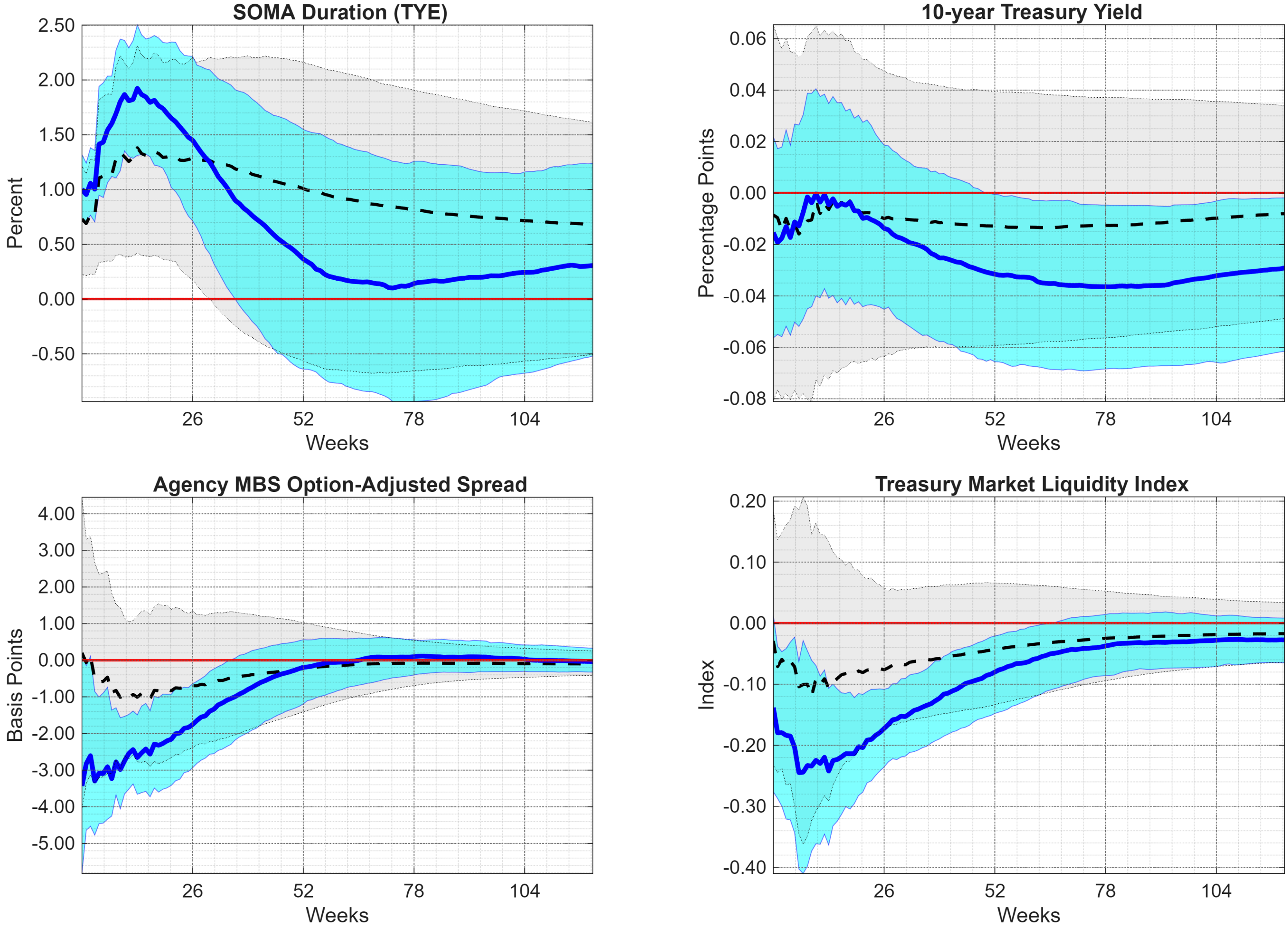


Figure 6: Responses to a Market Functioning Shock

Duration is measured in 10-year-equivalent (TYE) Tn USD balances in the Federal Reserve's SOMA Account — the dotted red line in Figure (2) — enters the VAR in log levels. Shaded areas denote the 16% - 84% probability interval of the posterior distribution for the two restriction schemes. The dash line and gray shaded ares denote responses corresponding to the **Baseline Specification: Assumptions 1–3**. The solid blue line and blue shaded area denote responses corresponding to the **Narrative Restrictions: Assumptions 4 and 5**.

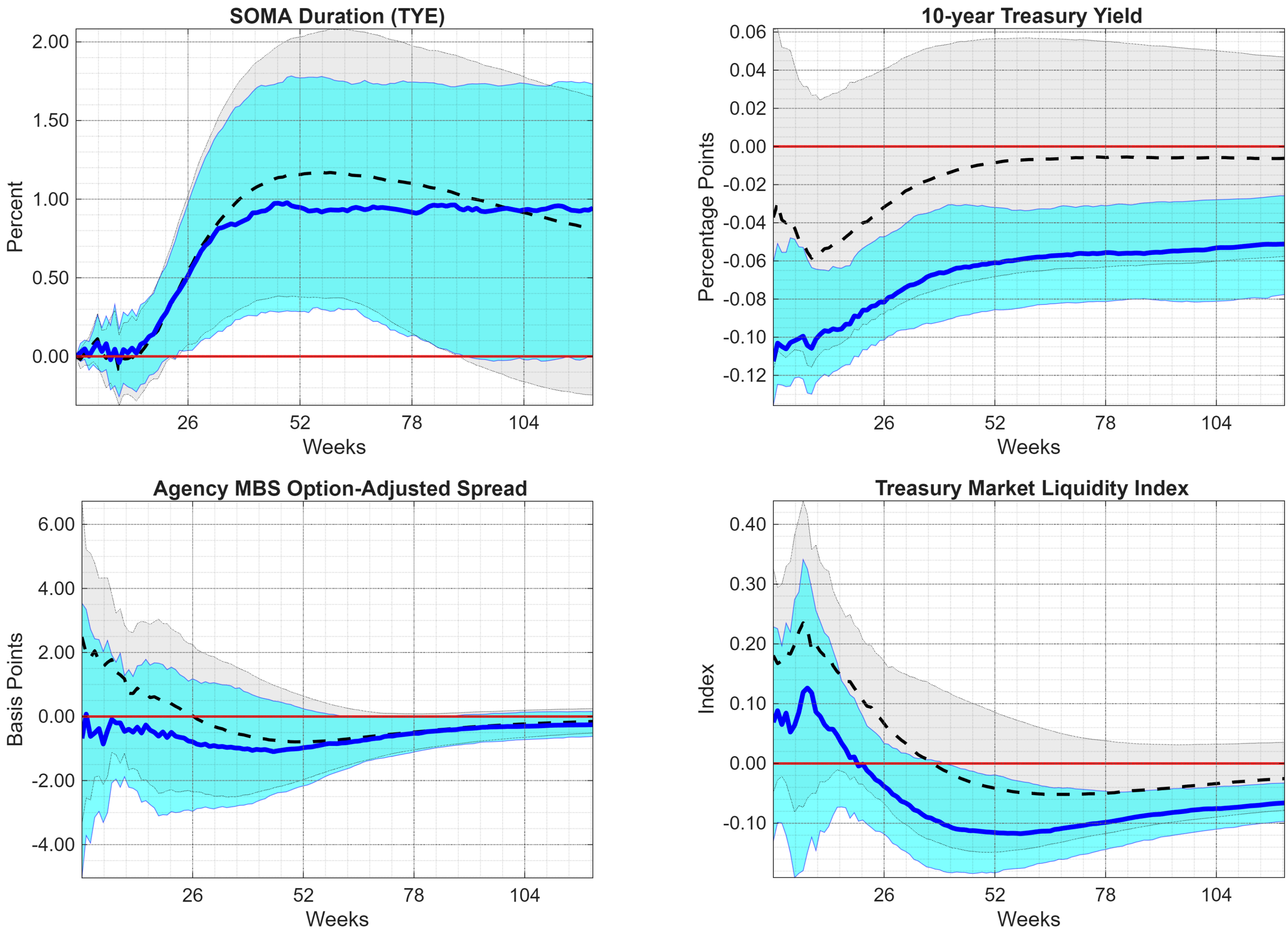


Figure 7: Responses to an Accommodation Shock

Duration is measured in 10-year-equivalent (TYE) Tn USD balances in the Federal Reserve's SOMA Account — the dotted red line in Figure (2) — enters the VAR in log levels. Shaded areas denote the 16% - 84% probability interval of the posterior distribution for the two restriction schemes. The dash line and gray shaded ares denote responses corresponding to the **Baseline Specification: Assumptions 1–3**. The solid blue line and blue shaded area denote responses corresponding to the **Narrative Restrictions: Assumptions 4 and 5**.