Unwinding Quantitative Easing: State Dependency and Household Heterogeneity*

Cristiano Cantore†
Pascal Meichtry‡

January 2023

Abstract

This paper studies the macroeconomic effect of state dependency of central bank asset market operations and their interactions with household heterogeneity. We build a New Keynesian model with borrowers and savers in which quantitative easing and tightening operate through portfolio rebalancing between short-term and long-term government bonds. We quantify the aggregate impact of an occasionally binding zero lower bound in determining an asymmetry between the effects of asset purchases and sales. When being close to the lower bound, raising the nominal interest rate before unwinding quantitative easing minimizes the economic costs of monetary policy normalization. Furthermore, our results imply that household heterogeneity in combination with state dependency amplifies the revealed asymmetry, while aggregate effects remain unaffected with only heterogeneous agents.

JEL Classification: E21, E32, E52, E58

Keywords: Unconventional Monetary Policy, Quantitative Tightening, Quantitative Easing, Heterogeneous Agents, Zero Lower Bound, State Dependency

*We thank Saleem Bahaj, Kenza Benhima, Pau Belda, Gianluca Benigno, Andrea Ferrero, Richard Harrison, Mike Joyce, Iryna Kaminska, Ricardo Reis, Martin Seneca, and Vincent Sterk for helpful comments. We are also grateful to seminar participants at the Bank of England and the UNIL Macro Lunch as well as participants at the 53rd Annual Conference of the Money, Macro and Finance Society, and the NIESR-CFM Quantitative Easing & Quantitative Tightening workshop. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Bank of England or any of its committees.

†Bank of England and University of Surrey. Email: cristiano.cantore@gmail.com

‡University of Lausanne. Email: pascal.meichtry@unil.ch

1 Introduction

In recent years, large-scale asset purchases have considerably increased the size of central banks' balance sheets. At the same time, as interest rates can fall back to low levels, future crises might once more call for similar unconventional policy tools to stabilize the economy. Central banks are therefore inclined to reduce the quantity of long-term bonds in their books to have sufficient leeway for monetary stimulus when the next negative shock happens.

Various studies have investigated the macroeconomic impact of quantitative easing (QE), generally finding evidence for meaningful effects on output and inflation (see, e.g., Baumeister & Benati, 2013; Haldane, Roberts-Sklar, Wieladek, & Young, 2016; Joyce, Miles, Scott, & Vayanos, 2012; Kapetanios, Mumtaz, Stevens, & Theodoridis, 2012; Weale & Wieladek, 2016). In contrast, evidence on unwinding asset purchases is scarce, mainly because there have to date only been a few attempts to actively do it.

Nevertheless, it seems reasonable to assume that balance sheet reductions do not necessarily have macroeconomic effects that are equal but opposite to expansions. For example, the Federal Reserve's unwind experience in 2017-2019 revealed strong asymmetries in the form of larger liquidity effects compared to periods of balance sheet expansion (Smith & Valcarcel, 2021). Furthermore, the effectiveness of unwinding might be closely linked to the state of the economy and financial markets — similar to QE itself having worked particularly well when frictions in financial markets were high (Bailey, Bridges, Harrison, Jones, & Mankodi, 2020; Haldane et al., 2016). Finally, unwinding past asset purchases is most likely executed at a slower pace and more gradually and its impact would probably be different from entering QE because of the interaction with policy rates (Vlieghe, 2018, 2021).

Understanding the implications of reducing the central bank's balance sheet is key to dampening the negative side effects on the economy and deciding when and how fast to take that step. Given the lack of empirical evidence on the subject, this issue has to be studied theoretically.

In this paper, we therefore present a two-agent New Keynesian model with borrowers and savers (TANK-BS) that we use to study: i) the asymmetric macroeconomic effects of QE and quantitative tightening (QT) driven by state dependency in the form of a zero lower bound (ZLB) on the nominal short-term interest rate; and ii) the interactions between QE/QT, the ZLB, and household heterogeneity. We thereby define QT as an *active* reduction of a central bank's balance sheet in the form of a sale of assets back to the secondary market, aimed to decrease the amount of liquidity within the economy. Our focus will be on long-term bonds from the government only.

Similar to QE, tightening works through different transmission mechanisms. This paper focuses on the portfolio balance channel.[1] Asset purchases or sales by a central bank change the relative supply of assets the private sector holds, implying movements in relative asset prices and

[1] There is a debate over the relative importance of different transmission channels of asset market operations. Several papers have demonstrated the significance of the portfolio balance channel for the effectiveness of QE (see, e.g., D'Amico & King, 2013; Joyce, Lasaosa, Stevens, & Tong, 2011). We deem it as equally important for large-scale asset sales as those will also change the relative supply of assets in the economy and the portfolio composition of households, hence implying potentially considerable real effects.

yields. Various studies show that QE programs have indeed raised financial asset prices and reduced longer-term interest rates, often substantially (Christensen & Rudebusch, 2012; Joyce et al., 2011; Krishnamurthy & Vissing-Jorgensen, 2011).

In our model, the two types of agents can borrow and save in short-term and long-term government bonds. The key assumption for the portfolio balance channel of QE/QT to work is the imperfect substitutability between assets, according to which investors value bonds along the yield curve differently (Andrés, López-Salido, & Nelson, 2004). Following Harrison (2017), we capture this idea using portfolio adjustment costs that investors have to pay whenever their preferred relative portfolio composition changes. Since asset market operations alter the relative supply and prices between short-term and long-term bonds, they incentivize asset holders to rebalance their portfolios. This, in turn, directly affects their average returns, because any adjustment is costly, and implies changes in their demand.

A large-scale asset sale in the model has an effect on bond returns which translates into an increase in the long-term interest rate and a decrease in the short-term real rate. These effects propagate to the real economy through changes in the portfolio allocation of all households and general-equilibrium effects on real wages, driving down individual consumption. The *direct* effects of QT through the bond market contribute thereby more persistently to the drop in consumption for both agents compared to the *indirect* effect through net labor income changes, among others due to a favorable tax cut. A major difference across the two household types is (countercyclical) profit income. It has a strong positive impact on savers' income such that their consumption drops by much less in relative terms compared to the case of borrowers.

Assuming the presence of state dependency in the form of a (non-)binding ZLB, we are then interested in how doing QE and unwinding it affects aggregate variables such as consumption and real output. The role of the lower bound and whether the nominal interest rate is available as an additional policy tool of the central bank will thereby be the main driver of the asymmetry we focus on.[2] As previous research has found, asset purchases are most effective if the ZLB is binding (Gertler & Karadi, 2013), but there also seems to be a role for asset market operations if policy rates are unconstrained (Sims & Wu, 2021). By analyzing the impact of state dependency on unwinding QE, we thus also address the question of when central banks should actually unwind.

In line with standard intuition, we find that a binding ZLB magnifies the macroeconomic effects of asset market operations by central banks. The response of the short-term real interest rate when the economy is in (or close to) a liquidity trap flips sign and is larger in magnitude. After a QT shock, the short-term real rate decreases when away from the ZLB, while it increases when at the lower bound, generating a further decrease in aggregate demand. As a result, when dealing with the risk of hitting the ZLB, our model implies that a central bank can minimize the economic costs of monetary policy normalization by prioritizing a policy rate hike before starting to sell assets. The likelihood of ending in a liquidity trap is thereby higher when the policy rate is close to the lower bound and QT starts too early or if the tightening is done too fast relative to the normalization of the short-term rate.

[2] Away from the ZLB, the TANK-BS model is symmetric.

The second aim of the paper is to study the interaction between state dependency of QE/QT and household heterogeneity. The empirical literature provides evidence of heterogeneous effects of QE on households across the income distribution (Montecino & Epstein, 2015; Mumtaz & Theophilopoulou, 2017; Saiki & Frost, 2014). On the other hand, quantitative models have recently found strong distributional effects of QE (Cui & Sterk, 2021). Moreover, there is a large literature showing how heterogeneity can amplify the real effects of conditional monetary policy (see, among others, Auclert, 2019; Bilbiie, 2018, 2020; Bilbiie, Känzig, & Surico, 2022; Debortoli & Galí, 2017). Against this backdrop, we want to study how the presence of heterogeneous households affects the asymmetry between QE and QT.

We find that household heterogeneity alone does not amplify the aggregate effects of asset market operations when the economy is off the ZLB. This result is in line with the one in the complementary work of Sims, Wu, and Zhang (2022b). Differently from us, they use a heterogeneous-agent New Keynesian (HANK) model with uninsurable income risk and QE introduced from the firm's side, as in Gertler and Karadi (2013). The lack of amplification for QE in their framework arises because most agents of the economy react in the same way as in the representative-agent New Keynesian (RANK) counterpart. Only very few households at the bottom of the wealth distribution behave differently and increase their consumption in response to a QE shock. Given that those agents represent a very small share of the population in the economy, it only has a marginal effect on aggregate consumption.

Our story here is different. We show that the lack of amplification via heterogeneity is due to a composition effect of changes in the balance sheet of the two household types and those changes almost entirely cancel out when moving from RANK to TANK-BS. Without borrowers in the model, all the impact of a QT shock on aggregate demand comes from a combination of direct effects (drop in bond demand and interest income) and indirect general-equilibrium effects (drop in real wage due to lower aggregate demand) on the income of the representative agent buying bonds from the central bank. When moving to TANK-BS, borrowers replace part of the savers in the population. While the latter behave like the representative agent in RANK, their share and thus their relative contribution to total spending are lower. The attenuated drop in aggregate demand through savers is compensated by a decrease in labor income of borrowers who have a larger marginal propensity to consume (MPC). The net effect of the lower cut in spending coming from savers and the additional decrease through borrowers is almost neutral. In the background, profit income is as before an essential element because the higher the proportion of savers the less each agent benefits from the increased (countercyclical) earnings of firms.

Finally, we show that household heterogeneity, when combined with state dependency, amplifies the aggregate effects of asset market operations. When asset sales are performed at the ZLB, the direct and indirect effects on borrowers discussed above together generate a stronger decline in labor income of high-MPC borrowers than the decline in spending contributed by savers.

Related literature. Our paper is related to several strands of the literature on asset market operations which we summarize hereafter.[3] On the empirical side, the literature has identified

[3] A thorough review of the literature is beyond the scope of this paper.

various channels through which QE affects the macroeconomy. See Bernanke (2020) and Bhattarai and Neely (2022) for comprehensive reviews. As discussed in the motivation, we focus here on the portfolio balance channel which is one of the key transmission mechanisms through which QE worked in the past.[4]

From a theoretical perspective, QE has mainly been studied in RANK setups (see, among others, Chen, Cúrdia, & Ferrero, 2012; Falagiarda, 2014; Gertler & Karadi, 2013; Harrison, 2012, 2017; Harrison, Seneca, & Waldron, 2021; Sims & Wu, 2021; Sims, Wu, & Zhang, 2022a). On the other hand, the bulk of the literature on household heterogeneity and monetary policy (e.g. Auclert, 2019; Bilbiie, 2008, 2020; Kaplan, Moll, & Violante, 2018) has mostly focused on conventional monetary policy. The only two papers we are aware of that merge these two pieces of literature are Cui and Sterk (2021) and Sims et al. (2022b). As discussed in the motivation, while we find a similar result as in the latter, our setup is different because we focus mainly on the effect of asset market operations coming through the balance sheet of households. In Cui and Sterk (2021) instead, the impact of QE on the macroeconomy emerges from the household side as well. They use a model with liquid and illiquid wealth, in the HANK tradition, and focus on the different MPCs out of the two types of wealth. Hence, in their model, household heterogeneity plays a direct role in the transmission mechanism of QE, which they show to be significant on output and inflation. Here we use a much simpler setup, allowing only for two types of agents as in Eggertsson and Krugman (2012) or Bilbiie, Monacelli, and Perotti (2013) and abstracting from liquid and illiquid wealth, while focusing on the impact of QE on households' bonds positions at different maturities. Furthermore, differently from Cui and Sterk (2021) and Sims et al. (2022b), we are not just interested in the interaction of heterogeneity and QE but also on the effects of the ZLB, which both papers abstract from.[5]

The works cited so far are primarily focused on QE. Empirically, this is obviously due to the lack of enough episodes of large-scale asset sales or, more generally, central bank balance sheet reductions. On the theoretical side, a few exceptions are Benigno and Benigno (2022), Cui and Sterk (2021), Karadi and Nakov (2021), Sims et al. (2022a), Wei (2022), and Wen (2014). To the best of our knowledge, Wen (2014) is the first theoretical attempt on QE exit strategies and its impact on firms. We focus instead on households and the impact of unwinding QE on their portfolios. Cui and Sterk (2021) analyze the impact of the speed of QE exit, captured by the persistence of the policy in the model. They show that the quicker the exit, the lower the real impact of the policy, which is driven by agents anticipating the dampening effects of exiting QE. By keeping the nominal interest rate pegged, however, they do not look at the interaction between conventional and unconventional monetary policy as we do in this paper. Karadi and Nakov (2021) and Sims et al. (2022a) look at the optimal conditions to exit QE. The former present

[4] See Christensen and Rudebusch (2012), D'Amico and King (2013), Froemel, Joyce, and Kaminska (2022), and Joyce et al. (2011) for empirical evidence on the portfolio balance channel. Related to this, see Andrés et al. (2004) and Vayanos and Vila (2009, 2021) for the theoretical foundation of imperfect substitutability between assets along the yield curve and preferred-habitat theory, respectively.

[5] Cui and Sterk (2021) assume in their model simulations for QE that the interest rate is pegged at zero. However, they do not compare simulations with and without the peg.

a model in which banks' balance sheet constraints bind only occasionally, so that asset purchases are not always effective. Unlike them, we are not conducting any normative analysis and focus on the implications of asset market operations via portfolio rebalancing of households' assets. Wei (2022) uses the preferred-habitat model of Vayanos and Vila (2021) to quantify how many interest rate hikes QT is equivalent to. Our focus is instead on the macroeconomic implications and we study the interaction of asset market operations with conventional monetary policy rather than treating the two as substitutes. A similar idea is advocated by Benigno and Benigno (2022) who study optimal monetary policy normalization when exiting a liquidity trap. Besides the policy rate, they view reserves as an additional tool of monetary authorities to influence macroeconomic aggregates, while we disregard liquidity in order to keep the central bank balance sheet simple and to stress the transmission through portfolio rebalancing. Somewhat contrary to our finding, their analysis implies that efforts to reduce the size of the central bank balance sheet ideally start before the policy rate is raised.

The last strand of the literature this paper addresses is related to state-dependent QE/QT and possible asymmetries between the two. Policymakers have discussed at length the possible causes and effects of state dependency, focusing mostly on different states of financial markets (Bailey et al., 2020; Haldane et al., 2016; Vlieghe, 2021). To maintain tractability and because our focus is on household portfolio compositions, we abstract in this paper from financial markets and focus on state dependency driven by the ZLB. With respect to asymmetries, we directly address the idea of policymakers that QT is likely to impact the economy by less than asset purchases. Potential explanations for this view include a milder reaction of bond markets as visible during the Federal Reserve's 2017-2019 unwind (Neely, 2019), the vanishing of signaling effects of asset market operations once policy rates are well above zero (Bullard, 2019), or differences in the nature and scope of QE/QT episodes and the prevailing economic and financial conditions (Smith & Valcarcel, 2021; Vlieghe, 2018, 2021)

Outline. The rest of the paper is organized as follows. Section 2 presents the TANK-BS model economy and describes the calibration and the solution method. Section 3 discusses the simulation results and section 4 concludes.

2 Asset market operations in a borrower-saver model

This section presents the main elements of the model used for our analysis. Further details on the derivation, a thorough description of the steady state, and an overview of all model equations are in Appendix A.

The model economy consists of four sectors: households, firms, a government and a central bank. The household sector is populated by two different types, savers and borrowers, who differ in their degree of patience, modeled as in Bilbiie et al. (2013) and Eggertsson and Krugman (2012). Firms are modeled as in standard New Keynesian models, with nominal frictions that generate sticky prices. The government finances public spending by issuing bonds and levying lump-sum taxes. It also implements redistributive policies by taxing firms' profits. Finally, the monetary authority follows a Taylor rule to set the nominal interest rate and participates in the market for long-term bonds. The design of asset market operations follows Harrison (2017).

2.1 Households

There is a continuum of households with a share λ being borrowers (B) who are constrained in terms of how much they can borrow. The remaining $1-\lambda$ are savers (S) with unconstrained access to asset markets. Borrowers are assumed to be less patient than savers, such that $\beta^S > \beta^B$. As will become clear later, this difference in the discount factors will induce lending from S to B in equilibrium.

The period utility function of household type $j = \{B, S\}$ is given by

$$U\left(c_t^j, N_t^j\right) = \theta_t \left(\frac{(c_t^j)^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}} - \zeta^j \frac{(N_t^j)^{1+\varphi}}{1+\varphi} \right),$$

where c_t is real consumption, N_t are hours worked, θ_t is a preference shock that follows an AR(1) process, σ is the elasticity of intertemporal substitution, $\frac{1}{\varphi}$ is the Frisch elasticity of labor supply, and ζ indicates how leisure is valued relative to consumption.

Both household types have access to bonds issued by the government. Following Harrison (2017), we differentiate between real short-term (b^j) and long-term ($b^{j,L}$) bonds. The former are one-period assets: a bond purchased in period $t-1$ pays a real return $r_{t-1} = \frac{R_{t-1}}{\Pi_t}$ at time t, where R is the gross nominal interest rate and $\Pi_t = \frac{P_t}{P_{t-1}}$ is the gross inflation rate. On the other hand, we assume that longer-term government debt is captured by perpetuities with coupon payments that decay exponentially over time as in Woodford (2001). Denoting by $\widetilde{B}_t^{j,L}$ the nominal long-term bond holdings of a saver and by V_t the nominal price of each of these bonds, we can write the value of long-term bond holdings as $B_t^{j,L} = V_t \widetilde{B}_t^{j,L}$. By defining χ as the long-term bond coupon decay rate, Harrison (2017) then shows that the (ex-post) nominal return on long-term bonds is $R_t^L = \frac{1+\chi V_t}{V_{t-1}}$. This formulation allows us to express long-term bonds in the budget constraint in terms of a single stock variable and a single (one-period) bond return instead of having to keep track of issued bonds and their prices over time. In real terms, a long-term bond $b_{t-1}^{j,L}$ therefore pays $r_t^L = \frac{R_t^L}{\Pi_t}$ in interest one period later.

Households face portfolio adjustment costs whenever they change the allocation of their assets between short-term and long-term bonds. In the style of Chen et al. (2012) and Harrison (2017), this adjustment cost is specified as

$$\Psi_t^j = \frac{\nu}{2} \left(\delta^j \frac{b_t^j}{b_t^{j,L}} - 1 \right)^2,$$

where $\delta^j = \frac{b^{j,L}}{b^j}$ is the steady-state ratio of long-term bonds to short-term bonds and $\nu > 0$ captures how costly deviations from a household's preferred steady-state portfolio mix are.[6]

Introducing adjustment costs implies a direct role for asset market operations to stimulate the economy, namely through the portfolio balance channel. If the central bank purchases bonds of a specific maturity, it thereby lowers the relative supply of those assets and so increases their price. Investors will rebalance their portfolios, which is costly due to the presence of Ψ and affects their average portfolio returns, thus implying a real impact through changes in individual and aggregate demand.[7] The adjustment cost captures in a parsimonious way the preferred-habitat theory which assumes that investors have preferences for specific maturities (Vayanos & Vila, 2009, 2021). In other words, these agents view different assets along the yield curve as imperfect substitutes (Andrés et al., 2004).

2.1.1 Savers

Unconstrained agents can save and borrow in both short-term and long-term bonds and receive dividends from their share holdings in monopolistically competitive firms. Besides these asset returns, savers also earn labor income and pay taxes. They each maximize their lifetime utility from consumption and leisure subject to their budget constraint in real terms, taking prices and wages as given:

$$\max_{c_t^S, N_t^S, b_t^S, b_t^{S,L}} \mathbb{E}_t \sum_{t=0}^{\infty} \left(\beta^S\right)^t \theta_t \left(\frac{(c_t^S)^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}} - \zeta^S \frac{(N_t^S)^{1+\varphi}}{1+\varphi} \right) \quad \text{subject to}$$

$$c_t^S + b_t^S + b_t^{S,L} = r_{t-1} b_{t-1}^S + r_t^L b_{t-1}^{S,L} + w_t N_t^S + \frac{1-\tau^D}{1-\lambda} d_t - t_t - \Psi_t^S - \frac{tr}{1-\lambda} ,$$

where b_t^S and $b_t^{S,L}$ are real short-term and long-term government bonds held by a saver, respectively, with corresponding interest rates r and r^L as described above. Furthermore, w_t is the real wage, d_t are real dividends from firms' profits equally distributed to savers, t_t are real lump-sum taxes levied by the government, Ψ_t^S are portfolio adjustment costs described above, and tr are steady-state transfers from savers to hand-to-mouth agents that ensure consumption equality between the two household types in steady state.[8] Profits of intermediate firms that are owned by savers are taxed at a rate of τ^D. The government redistributes the tax revenues as a direct transfer to constrained households.

Solving the decision problem (see Appendix A.1) results in the following consumption-leisure

[6]The proposed adjustment cost function only captures the impact of changes in the relative supply of an asset and thus deviations from a household's desired portfolio composition (so-called stock effects). Harrison (2017) or Harrison et al. (2021) consider in addition the impact of fundamental changes in that portfolio mix (flow effects).

[7]Asset market operations prove to be ineffective in baseline New Keynesian models. Changes in the portfolio allocation of households have no impact on real economic variables as shown, among others, by Eggertsson and Woodford (2003).

[8]We use a symmetric steady state with $c^B = c^S = c$ as a benchmark, modeled similar to Bilbiie et al. (2022).

choice condition and Euler equations for short-term and long-term bonds:

$$
\begin{aligned}
w_t &= \zeta^S \, (N_t^S)^{\varphi} \, (c_t^S)^{\frac{1}{\sigma}} \, , \\
1 &= \beta^S \, R_t \, \mathbb{E}_t \left[\frac{\theta_{t+1}}{\theta_t} \left(\frac{c_{t+1}^S}{c_t^S} \right)^{-\frac{1}{\sigma}} \frac{1}{\Pi_{t+1}} \right] - \frac{\nu \, \delta^S}{b_t^{S,L}} \left(\delta^S \, \frac{b_t^S}{b_t^{S,L}} - 1 \right) \, , \\
1 &= \beta^S \, \mathbb{E}_t \left[\frac{\theta_{t+1}}{\theta_t} \left(\frac{c_{t+1}^S}{c_t^S} \right)^{-\frac{1}{\sigma}} \frac{R_{t+1}^L}{\Pi_{t+1}} \right] + \frac{\nu \, \delta^S \, b_t^S}{\left(b_t^{S,L} \right)^2} \left(\delta^S \, \frac{b_t^S}{b_t^{S,L}} - 1 \right) \, .
\end{aligned}
$$

2.1.2 Borrowers

Constrained households have access to both types of government bonds as well and consume their disposable income together with transfers (net of taxes) from the government. Different from savers, they face a borrowing constraint such that the total amount borrowed in each period cannot exceed a given limit.[9] Each borrower therefore solves the following problem:

Borrowers are assumed to be less patient than savers, such that $\beta^S > \beta^B$. As will become clear later, this difference in the discount factors will induce lending from S to B in equilibrium.

$$
\max_{c_t^B, N_t^B, b_t^B, b_t^{B,L}} \mathbb{E}_t \sum_{t=0}^{\infty} \left(\beta^B \right)^t \, \theta_t \left(\frac{(c_t^B)^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}} - \zeta^B \, \frac{(N_t^B)^{1+\varphi}}{1+\varphi} \right) \quad \text{subject to}
$$

$$
\begin{aligned}
& c_t^B + b_t^B + b_t^{B,L} \leq r_{t-1} \, b_{t-1}^B + r_t^L \, b_{t-1}^{B,L} + w_t \, N_t^B + \frac{\tau^D}{\lambda} d_t - t_t - \Psi_t^B + \frac{tr}{\lambda} \, , \\
& -b_t^B - b_t^{B,L} \leq \overline{D} \, ,
\end{aligned}
$$

where $\overline{D} \geq 0$ is the exogenous borrowing limit. We assume that this constraint binds for all periods and borrowers thus have a high MPC.

Besides the borrowing constraint, the optimality conditions are very similar to the one of the savers, yielding:

$$
\begin{aligned}
w_t &= \zeta^B \, (N_t^B)^{\varphi} \, (c_t^B)^{\frac{1}{\sigma}} \, , \\
1 &= \beta^B \, R_t \, \mathbb{E}_t \left[\frac{\theta_{t+1}}{\theta_t} \left(\frac{c_{t+1}^B}{c_t^B} \right)^{-\frac{1}{\sigma}} \frac{1}{\Pi_{t+1}} \right] - \frac{\nu \, \delta^B}{b_t^{B,L}} \left(\delta^B \, \frac{b_t^B}{b_t^{B,L}} - 1 \right) + \psi_t^B \, , \\
1 &= \beta^B \, \mathbb{E}_t \left[\frac{\theta_{t+1}}{\theta_t} \left(\frac{c_{t+1}^B}{c_t^B} \right)^{-\frac{1}{\sigma}} \frac{R_{t+1}^L}{\Pi_{t+1}} \right] + \frac{\nu \, \delta^B \, b_t^B}{\left(b_t^{B,L} \right)^2} \left(\delta^B \, \frac{b_t^B}{b_t^{B,L}} - 1 \right) + \psi_t^B \, ,
\end{aligned}
$$

where $\psi_t^B \geq 0$ is the Lagrangian multiplier on the borrowing constraint, with complementary

[9] In equilibrium, constrained agents will borrow in both short-term and long-term bonds. Although they are termed government bonds, borrowers actually borrow from savers so that b_t^B and $b_t^{B,L}$ can alternatively be interpreted as bonds issued by B to S. Hence, the implicit assumption here is that public and private bonds are perfect substitutes.

slackness condition $\psi_t^B\left(b_t^B+b_t^{B,L}+\bar{D}\right)=0$. If the constraint is binding, $\psi_t^B>0$ so that the marginal utility of consuming today is larger than the expected marginal utility of saving in any of the two bonds.

2.2 Firms

The firm sector is standard and features two different types of agents: monopolistically competitive intermediate goods producers and perfectly competitive final goods firms.

Final goods producers. The final goods sector aggregates differentiated intermediate goods according to a CES production function:

$$y_t=\left(\int_0^1 y_t(i)^{\frac{\varepsilon-1}{\varepsilon}}\,di\right)^{\frac{\varepsilon}{\varepsilon-1}},$$

where ε is the elasticity of substitution. Final goods producers maximize their profits, resulting in a demand for each intermediate input of

$$y_t(i)=\left(\frac{P_t(i)}{P_t}\right)^{-\varepsilon}y_t\,,$$

where $P_t(i)$ is the price of intermediate good i and $P_t^{1-\varepsilon}=\int_0^1 P_t(i)^{1-\varepsilon}\,di$ the aggregate price index.

Intermediate goods producers. Varieties of intermediate goods i are produced by a continuum of monopolistically competitive firms with production function $y_t(i)=z_t\,N_t(i)$, where technology z_t follows an AR(1) process. Cost minimization implies real marginal costs $mc_t=\frac{w_t}{z_t}$.

Intermediate goods firms set prices subject to a quadratic adjustment cost à la Rotemberg (1982) with the degree of nominal price rigidity governed by ϕ_p:

$$\Psi_t^p=\frac{\phi_p}{2}\left(\frac{P_t(i)}{P_{t-1}(i)}-1\right)^2 y_t\,.$$

Following Bilbiie (2020), we also assume that the government imposes an optimal subsidy on sales, τ^S, to induce marginal cost pricing in steady state. This subsidy is financed by a lump-sum tax on firms such that $t_t^F=\tau^S y_t$. Thus, real profits of each intermediate goods producer i are given by

$$d_t(i)=\left(1+\tau^S\right)\frac{P_t(i)}{P_t}\,y_t(i)-w_t\,N_t(i)-\Psi_t^p-t_t^F\,.$$

Appendix A.2 shows the solution to the price-setting problem which leads to the standard Phillips curve:

$$\left(1+\tau^S\right)(1-\varepsilon)+\varepsilon\,mc_t-\phi_p\,(\Pi_t-1)\,\Pi_t+\beta^S\,\mathbb{E}_t\left[\frac{\theta_{t+1}}{\theta_t}\left(\frac{c_{t+1}^S}{c_t^S}\right)^{-\frac{1}{\sigma}}\phi_p\,(\Pi_{t+1}-1)\,\Pi_{t+1}\,\frac{y_{t+1}}{y_t}\right]=0\,.$$

Abstracting from price adjustment costs, the optimal subsidy that induces marginal cost pricing turns out to be $\tau^S=(\varepsilon-1)^{-1}$. Finally, using the expression for the lump-sum tax and aggregating over firms yields total real profits:

$$d_t = \left[1 - mc_t - \frac{\phi_p}{2}\left(\Pi_t - 1\right)^2\right] y_t \ .$$

2.3 Government and Monetary Policy

Monetary and fiscal policy are combined in one entity. The government budget constraint is given by

$$b_t + b_t^L = r_{t-1}\, b_{t-1} + r_t^L\, b_{t-1}^L + \Omega_t + g_t - t_t \ ,$$

where b_t and b_t^L are total real short-term and long-term bonds issued by the government, respectively, Ω_t are net purchases of long-term bonds by the central bank, and g_t is real government spending which follows an AR(1) process. Note that subsidy expenses and tax revenues from firms' profits are balanced in every period and thus do not appear in the budget constraint above.

We assume that lump-sum taxes are set by the following rule:

$$\frac{t_t}{t} = \left(\frac{t_{t-1}}{t}\right)^{\rho^{\tau,t}} \left(\frac{b_t + b_t^L}{b + b^L}\right)^{\rho^{\tau,b}} \left(\frac{g_t}{g}\right)^{\rho^{\tau,g}} \ .$$

Moreover, total supply of long-term bonds follows an AR(1) process:

$$\log\left(\frac{b_t^L}{b^L}\right) = \rho_{BL} \log\left(\frac{b_{t-1}^L}{b^L}\right) + \varepsilon_t^{b^L} \ .$$

Turning to the central bank, net asset purchases of long-term bonds are defined as

$$\Omega_t = b_t^{CB,L} - r_t^L\, b_{t-1}^{CB,L} \ ,$$

where $b_t^{CB,L}$ denotes the value of long-term bonds purchased by the central bank. The inclusion of central bank asset purchases in the consolidated budget constraint implies that asset market operations are financed by the central government, which itself will pay for it with either tax revenues from households or through the issuance of new short-term debt.

The central bank has two policy tools. First, it conducts QE/QT by deciding on which fraction q_t of the total market value of long-term bonds to buy/sell:

$$b_t^{CB,L} = q_t\, b_t^L \ ,$$

where we model q_t as a AR(1) process:

$$\log\left(\frac{q_t}{q}\right) = \rho_q \log\left(\frac{q_{t-1}}{q}\right) + \varepsilon_t^q \ .$$

Besides asset market operations, the monetary authority can implement conventional monetary policy by setting the nominal short-term interest rate, R, according to a standard Taylor rule:

$$\log\left(\frac{R_t}{R}\right) = \rho_r \log\left(\frac{R_{t-1}}{R}\right) + (1-\rho_r)\left[\phi_\pi \log\left(\frac{\Pi_t}{\Pi}\right)\right] + \varepsilon_t^m \,,$$

where ε_t^m is an i.i.d. policy shock.

2.4 Aggregation and market clearing

Aggregate consumption and aggregate hours are given by

$$\begin{aligned} c_t &= \lambda c_t^B + (1-\lambda)c_t^S \,, \\ N_t &= \lambda N_t^B + (1-\lambda)N_t^S \,. \end{aligned}$$

Market clearing for short-term and long-term bonds, respectively, requires

$$\begin{aligned} b_t &= b_t^H \,, \\ b_t^L &= b_t^{H,L} + b_t^{CB,L} \,, \end{aligned}$$

with households' total demand for short-term bonds $b_t^H = \lambda b_t^B + (1-\lambda)b_t^S$ and for long-term bonds $b_t^{H,L} = \lambda b_t^{B,L} + (1-\lambda)\,b_t^{S,L}$. By using the equation for asset market operations, we can write $b_t^{H,L} = (1-q_t)\,b_t^L$. This condition shows the direct impact of asset purchases and sales on long-term bond holdings and hence the portfolio mix of households.[10]

Finally, the aggregate resource constraint is given by

$$y_t = c_t + g_t + \frac{\phi_p}{2}\left(\Pi_t - 1\right)^2 y_t \,.$$

2.5 Steady state

We approximate our model around a deterministic steady state with zero net inflation and output normalized to one. Our assumption $\beta^S > \beta^B$ implies that the borrowing constraint will always bind in steady state:

$$\psi^B = \left(c^B\right)^{-\frac{1}{\sigma}}\left[1 - \frac{\beta^B}{\beta^S}\right] > 0 \,.$$

As a result, patient (impatient) agents will be net lenders (borrowers) in steady state.

The Euler equations of the saver yield for the nominal rates that $R = R^L = \left(\beta^S\right)^{-1}$ and we have $r = R$ and $r^L = R^L$. The presence of the optimal subsidy to firms results in zero profits ($d = 0$). Furthermore, we assume that labor supply is equalized across households ($N^B = N^S = N$), which implies that they will consume the same amount in steady state ($c^B = c^S = c$).

Regarding the steady-state ratio of bond holdings, δ^j, we impose the simplifying assumption that they are equal across household types such that individual demand variables can be replaced

[10] In Appendix A.1, we derive a no-arbitrage condition between short-term and long-term bonds. It shows that changes in households' portfolio composition caused by central bank asset market operations directly affect the long-term bond return, namely due to the presence of the portfolio adjustment cost.

by their household-level counterparts:

$$\delta^S = \delta^B = \delta = \frac{b^{H,L}}{b^H} \, .$$

We further define $\tilde{\delta} = \frac{b^L}{b}$ as the steady-state ratio between total long-term and short-term bonds. Finally, note that portfolio and price adjustment costs will be zero at steady state.

2.6 Calibration and simulation setup

Our calibration is summarized in Table 1. We target the case of the U.S. economy.

The parameters from the household sector are mostly taken from Bilbiie et al. (2013) who build a borrower-saver model similar to ours. In particular, we target a steady-state real interest rate of 4% annually. The baseline value for the savers' discount factor is therefore set to 0.99, while we will increase it to $\beta^S = 0.99955$ for some simulations later on. Regarding the production side, it is worth mentioning that we set taxes on profits to zero in order to rule out any impact from redistribution on the income of borrowers.

For the bond-related parameters, we choose $\chi = 0.975$ to match the average duration of ten-year US Treasury bonds in the non-stochastic steady state, following Harrison (2017) and Harrison et al. (2021) who draw on D'Amico and King (2013). The same value is also used by Sims et al. (2022b). The adjustment cost parameter ν is chosen such that the model matches the empirical evidence by Weale and Wieladek (2016) on the impact of a QE shock on real output, as discussed hereafter. Finally, the value of central bank's long-term bond holdings in steady state implies that households hold a share of 0.75, namely three-quarters of the stock of long-term debt, which is equivalent to the calibration in Gertler and Karadi (2013) and Karadi and Nakov (2021).

Output is normalized to one in steady state, while the target for net inflation is 0%, in line with Cui and Sterk (2021). Moreover, the persistence of the preference shock is set to 0.8, a high value as is common in the literature (see, e.g., Bianchi, Melosi, & Rottner, 2021). It allows to achieve a lasting ZLB spell of several quarters in our simulations. Finally, the chosen QE smoothing reflects the high persistence of asset market operations and is similar to the value of 0.8 in Sims and Wu (2021) or Sims et al. (2022a).

In each simulation we run below, the shock size is such that the central bank buys or sells long-term bonds worth 1% of annualized nominal GDP. We then match the output response to empirical evidence from the United States. The simulation results used for the matching are the impulse responses of the net effect of a QE shock that happens when the economy is in a liquidity trap, a situation brought about by a negative preference shock. See section 3.2 for more details. All the other simulations build on the parameterization from this exercise.

Weale and Wieladek (2016) show that the peak impact on U.S. real GDP of an asset purchase in the size of 1% of annualized nominal GDP has been around 0.58%.[11] We take this number as our target for the average output response during the first four quarters subsequent to a QE shock

[11] This number reflects the average of median peak effects of four different identification schemes in Weale and Wieladek (2016) that all leave the reaction of real GDP unrestricted.

at the ZLB, following the approach used in Cui and Sterk (2021). More specifically, we set the adjustment cost parameter ν accordingly to approximate this target.

To solve our model with the occasionally binding lower bound constraint, we use the dynare-OBC toolbox developed by Tom Holden.[12] Given that we approximate the model at first order, our simulation results will be perfect foresight transition paths in response to a QE or QT shock.

3 Results

In this section, we discuss the model simulations. We proceed in three steps. First, we study the impact of asset market operations when the economy is either close to or well above the ZLB and analyze the shock transmission to the real economy. Second, we examine the asymmetric macroeconomic effects of QE and QT due to state dependency. Finally, we compare our TANK-BS model to its representative-agent counterpart to isolate the implications of household heterogeneity.

3.1 Asset market operations and unwinding QE close to the ZLB

We start by illustrating what the TANK-BS model implies about the potential impact on macroeconomic aggregates of doing QE/QT and unwinding QE, conditional on an existing state dependency in the form of a lower bound on the nominal short-term interest rate. Figure 1 shows selected impulse responses to a QE and QT shock occurring when the economy is sufficiently far away from the ZLB and a QT shock which hits an economy that is already close to the ZLB. See Appendix B.1 for the entire set of impulse responses.

To explain how the model works, we begin by analyzing a standard QT shock, captured by the solid red line in the figure. When the central bank sells long-term bonds, the amount of assets available to other agents in the economy increases. The return of those bonds goes up and their price decreases. Together with the lower short-term interest rate, both household types therefore demand more long-term and less short-term bonds. Constrained agents borrow now more in the short-term asset because it has become cheaper, while savers purchase the long-term asset sold by the central bank. Overall, the lower demand for long-term bonds from the central bank is exactly offset by the higher demand from households so that the supply of long-term bonds remains fixed.

To understand the transmission of the shock to the real economy, it is useful to study the responses of the components of each agent's budget constraint to an asset market operation. Figure 2 shows that individual consumption of both household types decreases in response to a QT shock far enough off the ZLB, but that the underlying driving forces differ. We distinguish between direct effects of the asset sales (changes in bond demand and returns) and indirect general-equilibrium effects (changes in the real wage and profits).[13]

---

[12] See Holden (2016, 2022) for theory and computational details.

[13] The partition in Figure 2 can be captured by the budget constraints of the two household types: $c_t^j = \left[-b_t^j - b_t^{j,L} + r_{t-1} b_{t-1}^j + r_t^L b_{t-1}^{j,L} - \Psi_t^j\right] + \left[w_t N_t^j - t_t\right] + \left[d_t^j\right] + tr^j$, for $j = \{B,S\}$ and with $d_t^B = \frac{\tau^D}{\lambda} d_t$ and $d_t^S = \frac{1-\tau^D}{1-\lambda} d_t$. The square brackets represent the bond demand/interest, the net labor income, and the profit income component, respectively.

Table 1: Parameter values

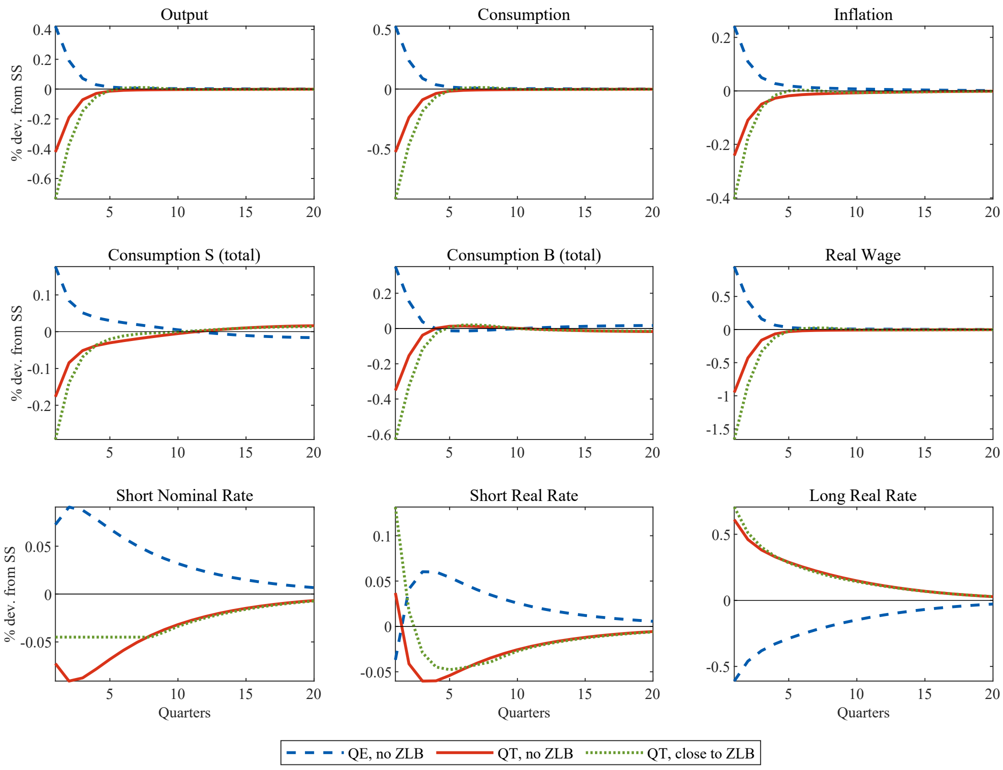
Parameter	Description	Value	Source / Target
λ	Proportion of borrowers	0.35	Bilbiie et al. (2013)
σ	Intertemporal elasticity of substitution	1	Conventional
$1/\varphi$	Frisch elasticity of labor supply	1	Conventional
β^S	Discount factor, saver	0.99	Annual steady-state interest rate of 4%; Bilbiie et al. (2013)
β^B	Discount factor, borrower	0.95	Bilbiie et al. (2013)
$\bar{D}$	Borrowing limit	0.5	Bilbiie et al. (2013)
ε	Elasticity of substitution between goods	6	Price markup of 20%
τ^D	Tax on profits	0	No redistribution
ϕ_p	Rotemberg price adjustment cost	42.68	3.5-quarters price duration
ϕ_π	Taylor rule coefficient on inflation	1.5	Conventional
χ	Long-term bond coupon decay rate	0.975	Average bond duration of 7-8 years
ν	Portfolio share adjustment cost	0.05	Empirical evidence on output response by Weale and Wieladek (2016)
$\tilde{\delta} = b^L/b$	Steady-state ratio of long-term to short-term bonds	0.3	Harrison (2017), Harrison et al. (2021)
$q = b^{CB,L}/b^L$	Steady-state CB long-term bond holdings	0.25	Households' long-term bond holdings
g/y	Steady-state government-spending-to-GDP ratio	0.2	Galí et al. (2007)
$(b+b^L)/y$	Steady-state total-debt-to-GDP ratio	0.8	U.S. average since 2009
Π	Steady-state gross inflation rate	1	Inflation target
Y	Steady-state output	1	Normalized
τ^S	Production subsidy	$(\varepsilon-1)^{-1}$	Marginal cost pricing
ρ_θ	Persistence of preference shock	0.8	Own choice
$\rho^{\tau,t}$	Tax smoothing in fiscal rule	0.7	Own choice
$\rho^{\tau,b}$	Tax response to total debt	0.33	Galí et al. (2007)
$\rho^{\tau,g}$	Tax response to government spending	0.1	Galí et al. (2007)
ρ_r	Interest rate smoothing	0.8	Sims and Wu (2019)
ρ_q	QE smoothing	0.9	Cui and Sterk (2021)

The first panel reveals that the change in savers' labor income through general equilibrium has a negative effect on consumption, but only on impact of the shock. After that, the cut in lump-sum taxes and, in particular, the strong increase in countercyclical profits push savers' income up and leads to a quick recovery. Instead, the medium-term negative consumption response is mainly driven by developments in their portfolio allocation. By buying long-term bonds from the central bank, savers give up some of their income because changes in the bond portfolio are costly. This drop in income is larger than their gains from selling short-term bonds together with the increase in interest income coming from more long-term bonds in their portfolio and the higher real rate on these assets.[14] This effect depresses consumption of savers and thus aggregate demand.

The bottom panel of Figure 2 shows some commonalities for borrowers. Their bond demand

[14]Strictly speaking, the rise in savers' long-term bond holdings is larger than the decrease in short-term bonds. Similarly, their interest income from long-term bonds increases by more than the income from short-term bonds falls.

Figure 1: Impulse responses to a QE/QT shock and a QT shock near the ZLB



Notes: This figure depicts the impulse responses of selected variables to a QE (dashed blue line) and a QT (solid red line) shock occurring far enough above the ZLB, and a QT shock happening close to the ZLB (dotted green line, simulated with $\beta^S = 0.99955$). The shock for each simulation is an asset purchase/sale of size 1% of annualized GDP. Responses for individual consumption levels and hours are weighted by population shares of savers (S) and borrowers (B), respectively, and thus represent total responses.

and interest payments react similarly to the one for savers. The other negative income effect comes as before through net labor income. While borrowers do not change labor supply by a lot because they cannot afford to work much less, the lower spending from savers hurts them through the drop in the real wage.[15] This effect on labor income is again short-lived due to the cut in taxes that causes a fast rebound.

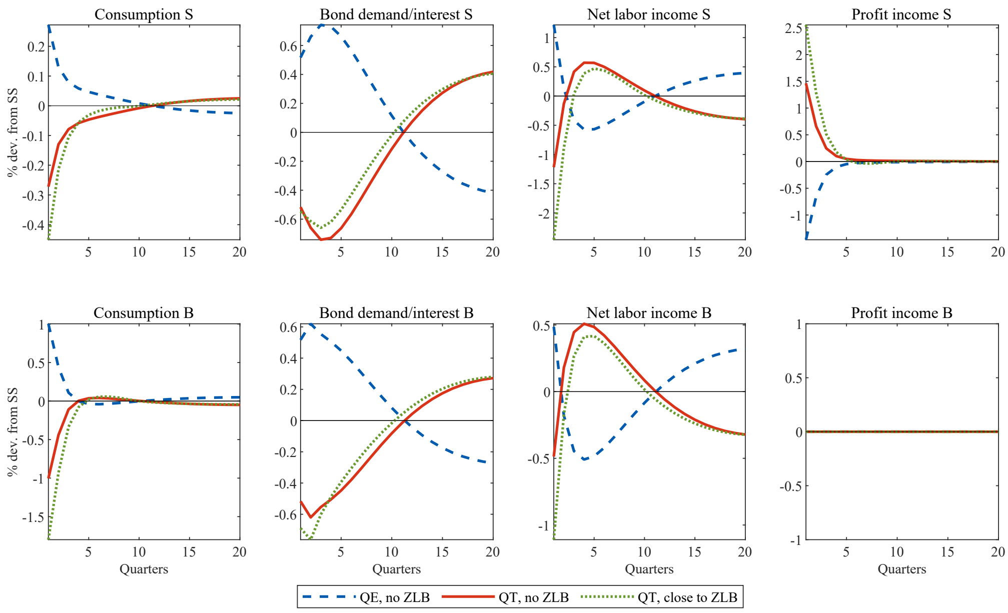
Overall, the direct effects of QT and the ensuing changes in returns are considerable for all households and the indirect effect through the labor market is counterbalanced by a cut in taxes. The major difference that leads to a weaker drop in individual consumption of savers, however, is the response of profits. They constitute a strong boost for them such that their individual consumption drops by much less in relative terms.

A key point to mention here is that QT is modeled as the exact opposite of QE. Given the linearity of the model, both policies have therefore the same impact in absolute terms – as long as the economy is far enough away from the ZLB such that the QT shock cannot push it into a liquidity trap. This is also visible from Figure 1. QE decreases the long-term rate and increases the short-term rate. These effects then propagate to the real economy via households demanding

Combined, the former effect is larger and leads to a negative net effect out of the bond-related variables in the saver's budget constraint, as depicted in Figure 2.

[15]The weak reaction of borrowers' labor supply is also visible in the full set of impulse responses in Appendix B.1.

Figure 2: Households' budget components to a QE/QT shock and a QT shock near the ZLB



Notes: This figure shows grouped components of the budget constraints of savers (top) and borrowers (bottom) in response to a QE (dashed blue line) and a QT (solid red line) shock occurring far enough above the ZLB, and a QT shock happening close to the ZLB (dotted green line, simulated with $\beta^S = 0.99955$). The shock for each simulation is an asset purchase/sale of size 1% of annualized GDP. Each panel consists of four columns, containing the responses of individual consumption, bond-related variables (bond demand, interest payments/income, net of adjustment cost), labor income net of taxes, and income from profits. All responses are shown in per-capita terms.

more short-term and less long-term bonds, which translates into a higher aggregate demand and leads to a rise in all main aggregate variables.

Starting from a state of the world with symmetric effects of QE and QT makes it possible to isolate the asymmetry emerging from the presence of a ZLB. By allowing for a binding lower bound on the nominal short-term interest rate, we introduce state dependency that can generate asymmetric effects of asset market operations, similar to the literature about fiscal policy and the government-spending multiplier (see, e.g., Christiano, Eichenbaum, & Rebelo, 2011). This idea is also motivated by previous research that confirmed a stronger effectiveness of asset purchases if the ZLB was binding (see Gertler & Karadi, 2013).

Assuming that the economy is currently in a situation where the log interest rate is close to (but not at) zero, even a mild QT shock can push it into a liquidity trap.[16] We illustrate this case by a simulation using our baseline calibration except that we set $\beta^S = 0.99955$. The implied lower steady-state real rate (annual: 0.18%) ensures that the ZLB will bind right on impact of the QT shock and for a total of eight quarters, given the same shock size as before.

This case is captured by the dotted green impulse responses in Figure 1. If the policy rate were unconstrained, it would drop on impact of the shock and show a hump-shaped course, mitigating the contractionary implications of the asset sales. However, with a binding ZLB, it cannot anymore decrease by that much, while long-term rates are still at a higher level. As a consequence, the short-

[16]We do not discuss here the case of QE done near the ZLB. Due to its expansionary effects, such an asset market operation would move the economy in any case away from the lower bound. Asset purchases can therefore even be an effective policy tool if the policy rate is unconstrained.

term real rate increases and both household types decrease their consumption by more compared to the unconstrained case, leading to larger drops in all aggregate variables and a deeper recession.

We can deduce from Figure 2 that the stronger decrease in savers' consumption right after the shock is substantially triggered by a magnified fall in labor income, which is again partly absorbed by positive profits. Borrowers are particularly hurt through the higher borrowing costs and the larger drop in the real wage.[17]

The above unveils a distinct asymmetry in the macroeconomic effects of QE and QT, precisely arising from the different states of the world and the (non-)availability of the nominal short-term interest rate to help to stabilize the economy. It also addresses the question of when central banks should actually unwind. It is obvious to see that the central bank needs to be sure that any tightening will not bring the policy rate back to zero. Otherwise, it risks strong adverse effects on the aggregate economy. As a result, when dealing with the risk of hitting the ZLB, our model implies that minimizing the economic costs of normalizing monetary policy requires the monetary authority to first raise the policy rate before starting with active asset sales. Such an approach is less harmful to the overall economy.

The likelihood of staying away from the ZLB depends on the optimal co-ordination between interest rate increases and QT with respect to the order, timing, and pace of actions. Selling assets before normalizing the policy rate increases the probability of ending in a liquidity trap and staying there for an extended period of time. A similar outcome awaits if QT starts when the short-term rate has not been raised enough or if the tightening is done too fast relative to the increases in the policy rate.

3.2 State-dependent asset market operations and their asymmetric impact

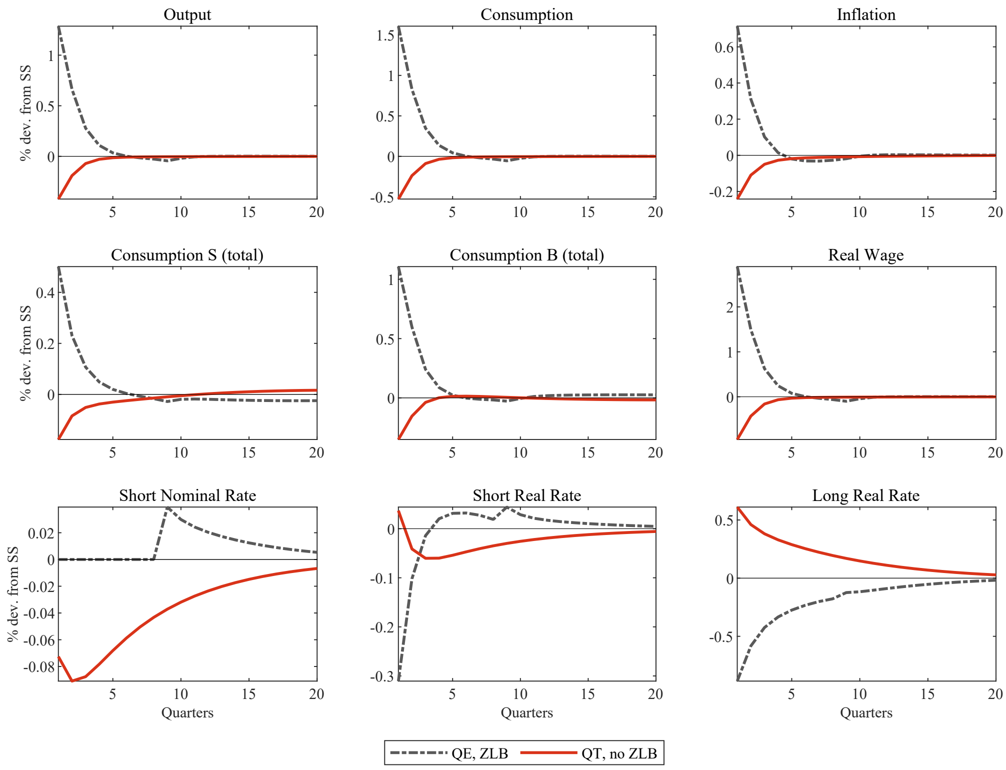
We now run a counterfactual exercise to compare QE and QT programs of similar size across different states of the economy. Based on the idea of state-dependent asset market operations, we compare two types of shocks: a QE shock that happens when the economy is in a liquidity trap, and a QT shock off the ZLB. Intuitively, central banks have heavily used large-scale asset purchase programs to fight the detrimental consequences of historically low interest rates in the past, often during times where the economy has been constrained at the ZLB. In contrast, we showed in the previous section that unwinding QE before the policy rate has reached a certain level is not advisable from our model's point of view.[18]

Figure 3 shows selected results of these simulations. Additional impulse responses are in Appendix B.2. We model the net effect of the QE shock by first simulating an asset purchase together with a negative preference shock and then deduct the impact of a mere preference shock. The size of the latter shock is chosen such that the economy is brought to the ZLB on impact and remains constrained for eight quarters. Generating a liquidity trap by a preference shock is a simple

[17]Overall, bond demand and supply variables respond similarly to QT, whether the economy is close to or away from the lower bound. See Appendix B.1 for the respective impulse responses.

[18]There is a huge debate on whether a CB should raise the policy rate first or should start with some tapering or active asset sales. See Forbes (2021) for a recent consideration.

Figure 3: Impulse responses to a QE shock at the ZLB and a QT shock off the ZLB



Notes: This figure depicts the impulse responses of selected variables to a QE shock when the ZLB on the policy rate is binding (dash-dotted gray line, showing the impact of QE net of a negative preference shock), and a QT shock occurring far enough above the ZLB (solid red line). The shock for each simulation is an asset purchase/sale of size 1% of annualized GDP. For QE, the size of the preference shock is chosen such that the ZLB binds for eight quarters. Responses for individual consumption levels and hours are weighted by population shares of savers (S) and borrowers (B), respectively, and thus represent total responses.

and effective way for our purpose to isolate the effects of state dependency (see, e.g., Christiano et al., 2011). Otherwise, the QT shock is equivalent to the shock in the previous section where we discussed its effects on macroeconomic aggregates and the associated transmission mechanism.

The figure reveals clear differences in the macroeconomic implications of the two shocks. As before, QE has a positive effect on aggregate demand while QT affects the economy negatively. When QE is done at the ZLB, however, its positive effect is magnified compared to the findings from the previous section without the lower bound. The resulting uneven responses of aggregate variables emerge from the prevalent state dependency, best visible from the asymmetric behavior of interest rates. The long-term rate response shows only minor (absolute) differences across the two shocks. On the other hand, while the short-term real interest rate increases after a QE shock when the economy is away from the ZLB, it flips sign when at the lower bound and falls considerably due to the inability of the policy rate to react.[19]

Our findings highlight the significance of the occasionally binding lower bound for the asymmetric implications between QE and QT. If conventional monetary policy is constrained and the economy is stuck in a liquidity trap, QE helps to stimulate aggregate demand and will have a larger

[19]This result resembles Gertler and Karadi (2013) who showed that central bank asset purchases lead to a larger drop in long-term rates the longer short-term rates are constrained.

effect than in normal times. With the nominal interest rate being at the ZLB, the rise in output and prices following a QE shock decreases the real rate considerably and thus fosters spending by households and boosts real wages.[20] This, in turn, results in an even higher output and constitutes an expansionary spiral.

3.3 Household heterogeneity and state dependency in interaction

As a final exercise, we study how household heterogeneity affects the asymmetry between QE and QT. For this purpose, we compare the impulse responses resulting from our borrower-saver model (named TANK-BS) with the ones from a standard representative-agent framework (named RANK) without heterogeneity on the household side. See Appendices B.3 and B.4 for the entire set of impulse responses. The shocks we focus on are the same as in the previous section, namely an asset purchase at the ZLB and an asset sale away from it.

The motivation for such an exercise comes from the implications of heterogeneity in households' income, wealth, or consumption and saving decisions found in the literature. Studies focusing on conventional monetary policy find substantial amplification (e.g. Auclert, 2019; Bilbiie, 2018, 2020; Bilbiie et al., 2022; Debortoli & Galí, 2017), driven by heterogeneity in MPCs out of a transitory income shock. Sims et al. (2022b) instead focus on QE and find no amplification coming from household heterogeneity. In our setup, borrowers have a higher MPC than savers. Any policy measure that relaxes their borrowing constraint frees up some individual income which is spent immediately and boosts aggregate demand and consumption. It appears therefore natural to study if amplification also arises after asset market operations.

Figure 4 shows the results for a QT shock when the economy is far enough off the ZLB such that the nominal short-term rate remains unconstrained. Adding household heterogeneity to a RANK model seems to have only a minor impact on the aggregate effects of QT (and due to the model linearity also of QE), which is in line with the finding in Sims et al. (2022b).

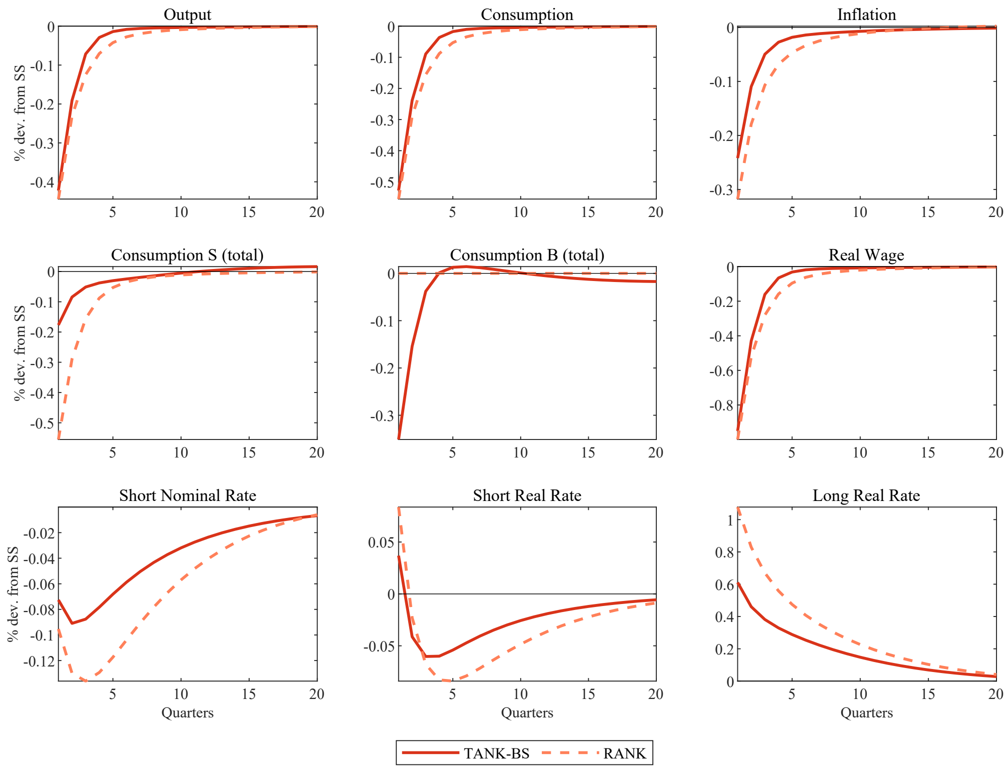
The reason for this lack of amplification via heterogeneity lies in a composition effect of changes in households' balance sheets that roughly cancel out when moving from RANK to TANK-BS. Without borrowers in the model, the propagation of the shock works entirely through the income of the saver. The representative agent purchases the bonds sold by the central bank, which drives down their income and thus aggregate demand. Compared to TANK-BS, we observe a higher effect on the demands of short-term and long-term bonds of the total responses across savers (as they are the only household type) and a larger effect on the long-term real rate.[21] Together with the lower increase in gains out of firms' profits, this magnifies the income drop of savers from buying long-term bonds from the central bank, therefore decreasing their consumption more than in the TANK-BS case.

When moving to TANK-BS, savers behave like the representative agent in RANK. They affect,

[20] The boost originating from the drop in the short-term real rate is so large that it generates an increase in real wages that induces borrowers to work less when QE is done at the ZLB. On the contrary, asset purchases away from the ZLB would induce them to increase their labor supply. See Appendix B.2 with the full set of impulse responses.

[21] Bond demands of savers in the RANK model are only more sensitive in *total* terms. Once we look at *per-capita* bond demands, the effect of a shock will be lower in RANK compared to TANK-BS due to the higher share of savers.

Figure 4: Impulse responses to a QT shock off the ZLB: RANK vs. TANK-BS



Notes: This figure depicts the impulse responses of selected variables to a QT shock occurring far enough above the ZLB, for the borrower-saver model (TANK-BS, solid red line) and its representative-agent counterpart for $\lambda = 0$ (RANK, dashed light red line). The shock for each simulation is an asset sale of size 1% of annualized GDP. Responses for individual consumption levels and hours are weighted by population shares of savers (S) and borrowers (B), respectively, and thus represent total responses.

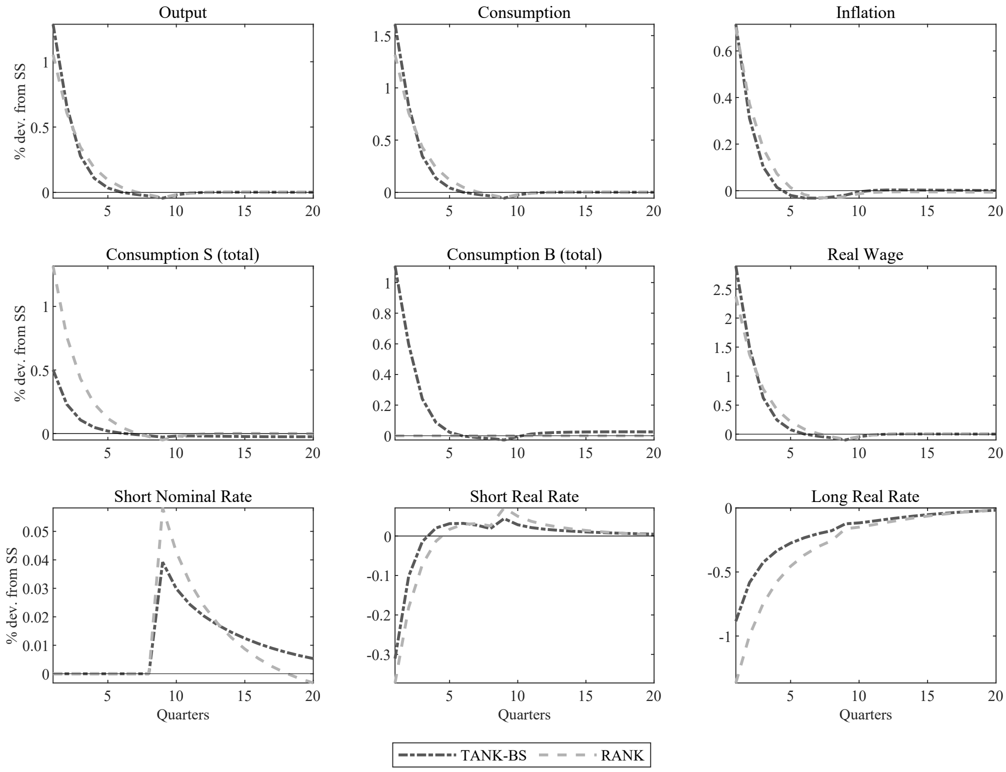
however, aggregate demand by relatively less given their lower share in the population and hence the higher profit income per agent. The reduced contribution to the fall in spending is compensated by a decrease in labor income of borrowers who have a larger MPC.[22] The net effect of the lower drop in aggregate consumption coming from savers and the additional decrease through borrowers is almost neutral. Even though this finding is consistent with the complementary work of Sims et al. (2022b), the story is different. The lack of amplification in their model arises because only very few households at the bottom of the wealth distribution respond other than the representative agent to a QE shock and their impact on aggregate consumption is therefore marginal.

Unlike a state of the world without a binding ZLB, household heterogeneity starts to matter more when combined with state dependency. Figure 5 shows that this case leads to amplified aggregate effects of asset purchases in TANK-BS.

The reasoning combines what has been described so far. First, the presence of the ZLB generates an asymmetric behavior of the short-term real rate, pushing consumption of both household types and hence aggregate demand upwards. Second, there is an extra boost from the presence of constrained households with a high MPC, such that an increase in their labor income through higher wages has a strong multiplier impact on aggregate demand. Together, these two elements

[22] See Figure 10 in Appendix B.3 for more details on each agent's budget constraint components.

Figure 5: Impulse responses to a QE shock at the ZLB: RANK vs. TANK-BS



Notes: This figure depicts the impulse responses of selected variables to a QE shock net of a negative preference shock when the ZLB on the policy rate is binding, for the borrower-saver model (TANK-BS, dash-dotted gray line) and its representative-agent counterpart for $\lambda = 0$ (RANK, dashed light gray line). The shock for each simulation is an asset purchase of size 1% of annualized GDP. The size of the preference shock is chosen such that the ZLB binds for eight quarters. Responses for individual consumption levels and hours are weighted by population shares of savers (S) and borrowers (B), respectively, and thus represent total responses.

lead to a larger increase in aggregate variables in TANK-BS.

Compared to the case of an asset market operation done off the ZLB discussed before, the direct and indirect effects of QE on borrowers together more than offset the reaction of savers in TANK-BS and the changes in their balance sheets no longer cancel out.[23] When an asset purchase is done when the lower bound binds, the impact of the increased labor income of borrowers with their high MPC exceeds the reduced contribution by savers in terms of spending, with a strong reaction of profits per agent being crucial again.

In order to quantify the asymmetry arising from state dependency in this model, Table 2 lists the responses of the main aggregate variables to the two shocks we have analyzed in this section, on impact and cumulated over four periods, and for both the RANK and the TANK-BS model.

The impact multipliers reveal two results. First and as in the previous section, the impact of QE on macroeconomic aggregates is larger than the absolute impact of QT. This holds for both models and constitutes a *within-model asymmetry*. Doing QE at the ZLB instead of unwinding it off the ZLB has a macroeconomic effect on impact that is more than two times stronger in RANK and about three times stronger in TANK-BS.

Second, as discussed before, household heterogeneity amplifies the aggregate effects of asset

[23] See Figure 12 in Appendix B.4 for more details on each agent's budget constraint components.

Table 2: Multipliers on impact and cumulated (in %)

	Output		**Inflation**		**Consumption**	
	QE	QT	QE	QT	QE	QT
RANK (impact)	1.05	-0.44	0.70	-0.32	1.32	-0.56
TANK-BS (impact)	1.29	-0.42	0.71	-0.24	1.61	-0.53
RANK (cumulative)	2.18	-0.86	1.32	-0.67	2.72	-1.08
TANK-BS (cumulative)	2.32	-0.71	1.14	-0.43	2.90	-0.89

Notes: This table summarizes the aggregate effects of a QE shock when the ZLB on the policy rate is binding and a QT shock occurring far enough above the ZLB, for the borrower-saver model (TANK-BS) and its representative-agent counterpart (RANK). The shock for each simulation is an asset purchase/sale of size 1% of annualized GDP. The table contains the multipliers both on impact of the shock and cumulated over the first four periods after the shock.

market operations only when it appears in combination with state dependency. This *across-model asymmetry* is therefore very weak in the case of our simulated QT shock, but sizable for QE simulated at the ZLB.[24] Moving from RANK to TANK-BS, the macroeconomic impact multiplier of QE is around 20% higher for output and consumption, but about the same for inflation. This result might arise because heterogeneity affects the slope of the aggregate demand curve but not the one of the Phillips curve. As a direct consequence, introducing household heterogeneity amplifies the within-model asymmetry.

4 Conclusion

In this paper, we present a two-agent New Keynesian model with borrowers and savers that is used to study state dependency of asset market operations and their interactions with household heterogeneity. Central bank asset purchases and sales operate via portfolio rebalancing between short-term and long-term government bonds held by the two types of households in the economy. These assets are imperfect substitutes due to portfolio adjustment costs in place. State dependency arises through the presence of an occasionally binding ZLB on the nominal short-term interest rate. Therefore, asymmetry between QE and QT in this context is driven by whether the nominal rate is available as a policy tool or is constrained by the lower bound.

We find that a binding ZLB magnifies the macroeconomic effects of asset market operations by central banks. This is due to the behavior of the short-term real rate when the economy is at (or close to) the lower bound. Consequently, when dealing with the risk of hitting the ZLB, our simulations imply that a central bank can mitigate the adverse effects of monetary policy normalization by prioritizing a policy rate hike over asset sales and thus by avoiding to tighten too early or too fast.

Moreover, we find that the role of household heterogeneity in amplifying the effects of asset

[24] Whether the aggregate effects of QT are slightly stronger or weaker depends on the calibrated parameter values. However, for a realistic calibration, QT has always around the same aggregate impact on output and total consumption in both models.

market operations also depends on the state of the economy. Away from the ZLB, household heterogeneity does not imply amplification. On the contrary, when asset market operations occur in a liquidity trap, we find substantial amplification for aggregate output and consumption.

Despite the lack of evidence, our model intends to contribute to a better understanding of the potential effects of balance sheet reductions. Given the widespread belief that the effects of QE and QT are not exactly of equal but opposite size, further work on the implications of monetary policy normalization are indispensable. In particular, it would be essential to analyze transmission channels other than portfolio rebalancing, to extend the heterogeneity dimension to a continuum of households, or to additionally consider frictions on the firms' side.

References

Andrés, J., López-Salido, J. D., & Nelson, E. (2004). Tobin's Imperfect Asset Substitution in Optimizing General Equilibrium. *Journal of Money, Credit and Banking*, *36*(4), 665–690.

Auclert, A. (2019). Monetary policy and the redistribution channel. *American Economic Review*, *109*(6), 2333–2367.

Bailey, A., Bridges, J., Harrison, R., Jones, J., & Mankodi, A. (2020, December). *The central bank balance sheet as a policy tool: past, present and future* (Staff Working Paper No. 899). Bank of England.

Baumeister, C., & Benati, L. (2013). Unconventional Monetary Policy and the Great Recession: Estimating the Macroeconomic Effects of a Spread Compression at the Zero Lower Bound. *International Journal of Central Banking*, *9*(2), 165–212.

Benigno, G., & Benigno, P. (2022, May). *Managing Monetary Policy Normalization* (Staff Report No. 1015). Federal Reserve Bank of New York.

Bernanke, B. S. (2020). The New Tools of Monetary Policy. *American Economic Review*, *110*(4), 943–983.

Bhattarai, S., & Neely, C. J. (2022). An Analysis of the Literature on International Unconventional Monetary Policy. *Journal of Economic Literature*, *60*(2), 527–97.

Bianchi, F., Melosi, L., & Rottner, M. (2021). Hitting the elusive inflation target. *Journal of Monetary Economics*, *124*, 107–122.

Bilbiie, F. O. (2008, May). Limited asset markets participation, monetary policy and (inverted) aggregate demand logic. *Journal of Economic Theory*, *140*(1), 162–196.

Bilbiie, F. O. (2018, January). *Monetary policy and heterogeneity: An analytical framework* (CEPR Discussion Paper No. 12601). Centre for Economic Policy Research.

Bilbiie, F. O. (2020). The New Keynesian cross. *Journal of Monetary Economics*, *114*, 90–108.

Bilbiie, F. O., Känzig, D. R., & Surico, P. (2022). Capital and income inequality: An aggregate-demand complementarity. *Journal of Monetary Economics*, *126*, 154–169.

Bilbiie, F. O., Monacelli, T., & Perotti, R. (2013). Public Debt and Redistribution with Borrowing Constraints. *The Economic Journal*, *123*(566), F64–F98.

Bullard, J. (2019, March). *When Quantitative Tightening Is Not Quantitative Tightening* (On the Economy Blog). Federal Reserve Bank of St. Louis. Retrieved from https://www.stlouisfed.org/on-the-economy/2019/march/bullard-when-quantitative-tightening-not-quantitative-tightening

Chen, H., Cúrdia, V., & Ferrero, A. (2012). The macroeconomic effects of large-scale asset purchase programmes. *The Economic Journal*, *122*(564), F289–F315.

Christensen, J. H. E., & Rudebusch, G. D. (2012). The Response of Interest Rates to US and UK

Quantitative Easing. *Economic Journal*, *122*(564), F385–F414.

Christiano, L., Eichenbaum, M., & Rebelo, S. (2011). When Is the Government Spending Multiplier Large? *Journal of Political Economy*, *119*(1), 78–121.

Cui, W., & Sterk, V. (2021). Quantitative easing with heterogeneous agents. *Journal of Monetary Economics*, *123*, 68–90.

D'Amico, S., & King, T. B. (2013). Flow and stock effects of large-scale treasury purchases: Evidence on the importance of local supply. *Journal of Financial Economics*, *108*(2), 425–448.

Debortoli, D., & Galí, J. (2017). *Monetary policy with heterogeneous agents: Insights from TANK models* (Economics Working Paper No. 1686). Universitat Pompeu Fabra.

Eggertsson, G. B., & Krugman, P. (2012). Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach. *The Quarterly Journal of Economics*, *127*(3), 1469–1513.

Eggertsson, G. B., & Woodford, M. (2003). The Zero Bound on Interest Rates and Optimal Monetary Policy. *Brookings Papers on Economic Activity*, *34*(1), 139–235.

Falagiarda, M. (2014). Evaluating quantitative easing: a DSGE approach. *International Journal of Monetary Economics and Finance*, *7*(4), 302–327.

Forbes, K. (2021, August). *Unwinding Monetary Stimulus in an Uneven Economy: Time for a New Playbook* (Speech, Jackson Hole Economic Policy Symposium). Massachusetts Institute of Technology. Retrieved from https://www.kansascityfed.org/research/jackson-hole-economic-symposium/macroeconomic-policy-in-an-uneven-economy/

Froemel, M., Joyce, M., & Kaminska, I. (2022, May). *The local supply channel of QE: Evidence from the Bank of England's gilt purchases* (Staff Working Paper No. 980). Bank of England.

Galí, J., López-Salido, J. D., & Vallés, J. (2007, March). Understanding the Effects of Government Spending on Consumption. *Journal of the European Economic Association*, *5*(1), 227–270.

Gertler, M., & Karadi, P. (2013). QE 1 vs. 2 vs. 3. . . : A Framework for Analyzing Large-Scale Asset Purchases as a Monetary Policy Tool. *International Journal of Central Banking*, *9*(1), 5–53.

Haldane, A., Roberts-Sklar, M., Wieladek, T., & Young, C. (2016, October). *QE: The Story so far* (Staff Working Paper No. 624). Bank of England.

Harrison, R. (2012, January). *Asset purchase policy at the effective lower bound for interest rates* (Staff Working Paper No. 444). Bank of England.

Harrison, R. (2017, September). *Optimal Quantitative Easing* (Staff Working Paper No. 678). Bank of England.

Harrison, R., Seneca, M., & Waldron, M. (2021). *Monetary policy options in a 'low for long' era.* (Mimeo)

Holden, T. D. (2016, July). *Computation of solutions to dynamic models with occasionally* *binding constraints* (EconStor Preprints No. 144569). ZBW - Leibniz Information Centre for Economics.

Holden, T. D. (2022, September). *Existence and uniqueness of solutions to dynamic models with occasionally binding constraints* (Discussion Paper No. 09/2022). Deutsche Bundesbank.

Joyce, M., Lasaosa, A., Stevens, I., & Tong, M. (2011). The Financial Market Impact of Quantitative Easing in the United Kingdom. *International Journal of Central Banking*, *7*(3), 113–161.

Joyce, M., Miles, D., Scott, A., & Vayanos, D. (2012). Quantitative Easing and Unconventional Monetary Policy – An Introduction. *Economic Journal*, *122*(564), F271–F288.

Kapetanios, G., Mumtaz, H., Stevens, I., & Theodoridis, K. (2012). Assessing the Economy-wide Effects of Quantitative Easing. *Economic Journal*, *122*(564), 316–347.

Kaplan, G., Moll, B., & Violante, G. L. (2018). Monetary Policy According to HANK. *American Economic Review*, *108*(3), 697–743.

Karadi, P., & Nakov, A. (2021). Effectiveness and addictiveness of quantitative easing. *Journal of Monetary Economics*, *117*, 1096–1117.

Krishnamurthy, A., & Vissing-Jorgensen, A. (2011). The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy. *Brookings Papers on Economic Activity*, *42*(2), 215–287.

Montecino, J. A., & Epstein, G. (2015, October). *Did Quantitative Easing Increase Income Inequality?* (Working Paper No. 28). Institute for New Economic Thinking.

Mumtaz, H., & Theophilopoulou, A. (2017). The impact of monetary policy on inequality in the UK. An empirical analysis. *European Economic Review*, *98*, 410–423.

Neely, C. J. (2019, April). *What to Expect from Quantitative Tightening* (Economic Synopses No. 8). Federal Reserve Bank of St. Louis. Retrieved from https://research.stlouisfed.org/publications/economic-synopses/2019/04/05/what-to-expect-from-quantitative-tightening

Rotemberg, J. J. (1982). Monopolistic price adjustment and aggregate output. *The Review of Economic Studies*, *49*(4), 517–531.

Saiki, A., & Frost, J. (2014). Does unconventional monetary policy affect inequality? Evidence from Japan. *Applied Economics*, *46*(36), 4445–4454.

Sims, E., & Wu, J. C. (2019, July). *The Four Equation New Keynesian Model* (NBER Working Paper No. 26067). National Bureau of Economic Research.

Sims, E., & Wu, J. C. (2021). Evaluating Central Banks' tool kit: Past, present, and future. *Journal of Monetary Economics*, *118*, 135–160.

Sims, E., Wu, J. C., & Zhang, J. (2022a). The Four Equation New Keynesian Model. *Review of Economics and Statistics*, *Forthcoming*.

Sims, E., Wu, J. C., & Zhang, J. (2022b, August). *Unconventional Monetary Policy According to HANK* (NBER Working Paper No. 30329). National Bureau of Economic Research.

Smith, A. L., & Valcarcel, V. J. (2021, January). *The Financial Market Effects of Unwinding the Federal Reserve's Balance Sheet* (Research Working Paper No. RWP 20-23). Federal Reserve Bank of Kansas City.

Vayanos, D., & Vila, J.-L. (2009, November). *A Preferred-Habitat Model of the Term Structure of Interest Rates* (NBER Working Paper No. 15487). National Bureau of Economic Research.

Vayanos, D., & Vila, J.-L. (2021). A Preferred-Habitat Model of the Term Structure of Interest Rates. *Econometrica*, *89*(1), 77–112.

Vlieghe, G. (2018, September). *The yield curve and QE* (Speech, Imperial College Business School, London). Bank of England. Retrieved from https://www.bankofengland.co.uk/speech/2018/gertjan-vlieghe-imperial-college-business-school-london

Vlieghe, G. (2021, July). *Running out of room: revisiting the 3D perspective on low interest rates* (Speech, London School of Economics). Bank of England. Retrieved from https://www.bankofengland.co.uk/speech/2021/july/gertjan-vlieghe-speech-at-the-london-school-of-economics

Weale, M., & Wieladek, T. (2016). What are the macroeconomic effects of asset purchases? *Journal of Monetary Economics*, *79*, 81–93.

Wei, B. (2022, July). *Quantifying "Quantitative Tightening" (QT): How Many Rate Hikes Is QT Equivalent To?* (Working Paper No. 2022-8). Federal Reserve Bank of Atlanta.

Wen, Y. (2014). When and how to exit quantitative easing? *Federal Reserve Bank of St. Louis Review*, *96*(3), 243–265.

Woodford, M. (2001). Fiscal Requirements for Price Stability. *Journal of Money, Credit and Banking*, *33*(3), 669–728.

A Borrower-saver model derivations

This part provides more details on the derivations of the model presented in section 2.

A.1 Household problem

Each household of type $j = \{B, S\}$ faces the following optimization problem:

$$\max_{c_t^j, N_t^j, b_t^j, b_t^{j,L}} \mathbb{E}_t \sum_{t=0}^{\infty} \left(\beta^j\right)^t \theta_t \left(\frac{(c_t^j)^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}} - \zeta^j \frac{(N_t^j)^{1+\varphi}}{1+\varphi} \right) \quad \text{subject to}$$

$$c_t^j + b_t^j + b_t^{j,L} \leq r_{t-1} b_{t-1}^j + r_t^L b_{t-1}^{j,L} + w_t N_t^j + d_t^j - t_t - \frac{\nu}{2} \left(\delta^j \frac{b_t^j}{b_t^{j,L}} - 1 \right)^2 + tr^j ,$$

$$0 \leq \mathbb{I}^j \left(b_t^B + b_t^{B,L} + \bar{D} \right) ,$$

where $d_t^B = \frac{\tau^D}{\lambda} d_t$, $d_t^S = \frac{1-\tau^D}{1-\lambda} d_t$, $tr^B = \frac{tr}{\lambda}$, and $tr^S = -\frac{tr}{1-\lambda}$. Moreover, $\mathbb{I}^j$ is an indicator function with values $\mathbb{I}^S = 0$ and $\mathbb{I}^B = 1$.

The resulting optimality conditions for each agent are:

$$U_{c,t}^j = \theta_t (c_t^j)^{-\frac{1}{\sigma}} ,$$

$$U_{N,t}^j = -\theta_t \zeta^j (N_t^j)^{\varphi} ,$$

$$w_t = -\frac{U_{N,t}^j}{U_{c,t}^j} ,$$

$$U_{c,t}^j + U_{c,t}^j \frac{\nu \delta^j}{b_t^{j,L}} \left(\delta^j \frac{b_t^j}{b_t^{j,L}} - 1 \right) = \beta^j R_t \mathbb{E}_t \left[U_{c,t+1}^j \frac{1}{\Pi_{t+1}} \right] + \mathbb{I}^j \psi_t^B ,$$

$$U_{c,t}^j - U_{c,t}^j \frac{\nu \delta^j b_t^j}{\left(b_t^{j,L}\right)^2} \left(\delta^j \frac{b_t^j}{b_t^{j,L}} - 1 \right) = \beta^j \mathbb{E}_t \left[U_{c,t+1}^j \frac{R_{t+1}^L}{\Pi_{t+1}} \right] + \mathbb{I}^j \psi_t^B ,$$

$$0 = \mathbb{I}^j \psi_t^B \left(b_t^B + b_t^{B,L} + \bar{D} \right) ,$$

where $\psi_t^B \geq 0$ is the Lagrangian multiplier on the borrowing constraint. It holds that $\psi_t^B > 0$ whenever the constraint is binding.

From the expressions above, we can derive the following Euler equations for short-term and long-term bonds, where we already imposed $\delta^S = \delta^B = \delta$ as specified in the description of the steady state (see section 2.5):

$$1 = \beta^j R_t \mathbb{E}_t \left[\frac{\theta_{t+1}}{\theta_t} \left(\frac{c_{t+1}^j}{c_t^j} \right)^{-\frac{1}{\sigma}} \frac{1}{\Pi_{t+1}} \right] - \frac{\nu \delta}{b_t^{j,L}} \left(\delta \frac{b_t^j}{b_t^{j,L}} - 1 \right) + \mathbb{I}^j \psi_t^B ,$$

$$1 = \beta^j \mathbb{E}_t \left[\frac{\theta_{t+1}}{\theta_t} \left(\frac{c_{t+1}^j}{c_t^j} \right)^{-\frac{1}{\sigma}} \frac{R_{t+1}^L}{\Pi_{t+1}} \right] + \frac{\nu \delta b_t^j}{\left(b_t^{j,L}\right)^2} \left(\delta \frac{b_t^j}{b_t^{j,L}} - 1 \right) + \mathbb{I}^j \psi_t^B .$$

Combining the two equations leads to an expression for the nominal return on long-term bonds as a function of the nominal rate on short-term bonds and the bond holdings of households:

$$\mathbb{E}_t R^L_{t+1} = \frac{1 - \frac{\delta\, b^j_t}{\left(b^{j,L}_t\right)^2}\, \widetilde{\Psi}^j_t - \mathbb{I}^j\, \psi^B_t}{1 + \frac{\delta}{b^{j,L}_t}\, \widetilde{\Psi}^j_t - \mathbb{I}^j\, \psi^B_t}\, R_t \ ,$$

where $\widetilde{\Psi}^j_t = \nu \left(\delta\, \frac{b^j_t}{b^{j,L}_t} - 1\right)$. This equation is a no-arbitrage condition between the two types of bonds and captures the key impact channel of asset market operations on bond returns. When the central bank buys or sells long-term bonds, it changes the quantity of assets available to the rest of the economy. Holding bond supply fixed, this implies that households' portfolio mix is not at its desired level and induces costly portfolio rebalancing. The impact of the adjustment cost and of changes in bond demands is directly visible from the equation above. It can be shown that the fraction is larger than one whenever $\delta < \frac{b^{j,L}_t}{b^j_t}$ and smaller than one otherwise.

A.2 Intermediate goods producer problem

The price-setting problem of an intermediate goods firm is

$$\max_{\{P_{t+k}(i)\}_{k=0}^{\infty}} \mathbb{E}_t \sum_{k=0}^{\infty} \Lambda_{t,t+k} \left[\left(1+\tau^S\right) \frac{P_{t+k}(i)}{P_{t+k}}\, y_{t+k}(i) - mc_{t+k}\, y_{t+k}(i) - \frac{\phi_p}{2} \left(\frac{P_{t+k}(i)}{P_{t-1+k}(i)} - 1 \right)^2 y_{t+k} - t^F_{t+k} \right]$$

$$\text{s.t.} \quad y_{t+k}(i) = \left(\frac{P_{t+k}(i)}{P_{t+k}} \right)^{-\varepsilon} y_{t+k} \ ,$$

where $\Lambda_{t,t+k} = (\beta^S)^k \left(\frac{U^S_{c,t+k}}{U^S_{c,t}} \right)$ is the stochastic discount factor for payoffs in period $t+k$. The optimality condition of this optimization problem is

$$\mathbb{E}_t \left\{ \Lambda_{t,t} \left[\left(1+\tau^S\right) (1-\varepsilon)\, P_t(i)^{-\varepsilon}\, P_t^{\varepsilon-1}\, y_t + mc_t\, \varepsilon\, P_t(i)^{-\varepsilon-1}\, P_t^{\varepsilon}\, y_t - \phi_p \left(\frac{P_t(i)}{P_{t-1}(i)} - 1 \right) \frac{y_t}{P_{t-1}(i)} \right] \right.$$
$$\left. + \Lambda_{t,t+1}\, \phi_p \left(\frac{P_{t+1}(i)}{P_t(i)} - 1 \right) \frac{P_{t+1}(i)}{P_t(i)^2}\, y_{t+1} \right\} = 0 \ .$$

Since all firms are identical and face the same demand from final goods producers, they will all set the same price. This yields the following optimal price-setting condition:

$$\phi_p\, (\Pi_t - 1)\, \Pi_t - \mathbb{E}_t \left[\Lambda_{t,t+1}\, \phi_p\, (\Pi_{t+1} - 1)\, \Pi_{t+1} \frac{y_{t+1}}{y_t} \right] = \left(1+\tau^S\right)(1-\varepsilon) + \varepsilon\, mc_t \ .$$

A.3 Steady state

For the approximation of the model around a deterministic steady state, we assume a long-run inflation rate of unity ($\Pi = 1$), normalize output to one (by setting $z = N = 1$) and set $\theta = 1$.

The Euler equations of the saver gives $R = R^L = \left(\beta^S\right)^{-1}$. Using this in the Euler equations of the borrower implies that the borrowing constraint binds in steady state ($\psi^B > 0$) because we assumed $\beta^S > \beta^B$. We further impose for labor supply that $N^B = N^S = N$. Together with the steady-state transfer on the part of households, this results in $c^B = c^S = c$. Finally, the optimal subsidy to firms induces $mc = 1$ and thus zero profits ($d = 0$).

For the real returns, we get $r = R$ and $r^L = R^L$, which pins down the nominal bond price $V = 1/\left(R^L - \chi\right)$. The weights on hours are found through the labor supply equations, $\zeta^j = w\left(N^j\right)^{-\varphi}\left(c^j\right)^{\sigma}$ with $j = \{B, S\}$ and where $w = y$ from the expression for labor demand. Due to equalized levels of labor supply and consumption across household types, $\zeta^S = \zeta^B$. Finally, as portfolio adjustment costs are zero in steady state ($\Psi^j = 0$), the aggregate resource constraint determines consumption through $c = \left(1 - \frac{g}{y}\right) y$.

With respect to the bond-related variables, we impose $\delta^S = \delta^B = \delta = \frac{b^{H,L}}{b^H}$. This expression can be rewritten by using bond market clearing as $b^L = \frac{\delta b}{1-q}$, where we define $\tilde{\delta} = \frac{b^L}{b}$. Moreover, we write the annual steady-state total government debt-to-GDP ratio (in quarterly terms) as $b_y^{tot} = \frac{b+b^L}{4y}$, where the denominator captures annualized output. In order to find an expression for short-term government debt, we rewrite the last equation as $b = 4 b_y^{tot}\left(\frac{1-q}{1-q+\delta}\right) y$, or $b = 4 b_y^{tot}\left(\frac{1}{1+\tilde{\delta}}\right) y$. Market clearing then gives $b^H = b$.

Regarding the central bank, bond holdings are $b^{CB,L} = q b^L$. This pins down net asset purchases $\Omega = \left(1 - r^L\right) b^{CB,L}$ and households' total demand for long-term bonds $b^{H,L} = b^L - b^{CB,L}$. A borrower's bond holdings are then determined through the (binding) borrowing constraint, with $b^B = -\frac{\overline{D}}{(1+\delta)}$ and $b^{B,L} = \overline{D} - b^B$. A saver's holdings are pinned down by market clearing, with $b^S = \frac{b^H - \lambda b^B}{1-\lambda}$ and $b^{S,L} = \frac{b^{H,L} - \lambda b^{B,L}}{1-\lambda}$. Finally, lump-sum taxes are given by $t = g + \Omega - b(1-r) - b^L(1-r^L)$ and the steady-state transfer by $tr = \lambda[c^B + (1-r)b^B + (1-r^L)b^{B,L} - wN^B - \frac{\tau^D}{\lambda} d + t]$.

A.4 Model summary

Table 3 lists all equations of the TANK-BS model.

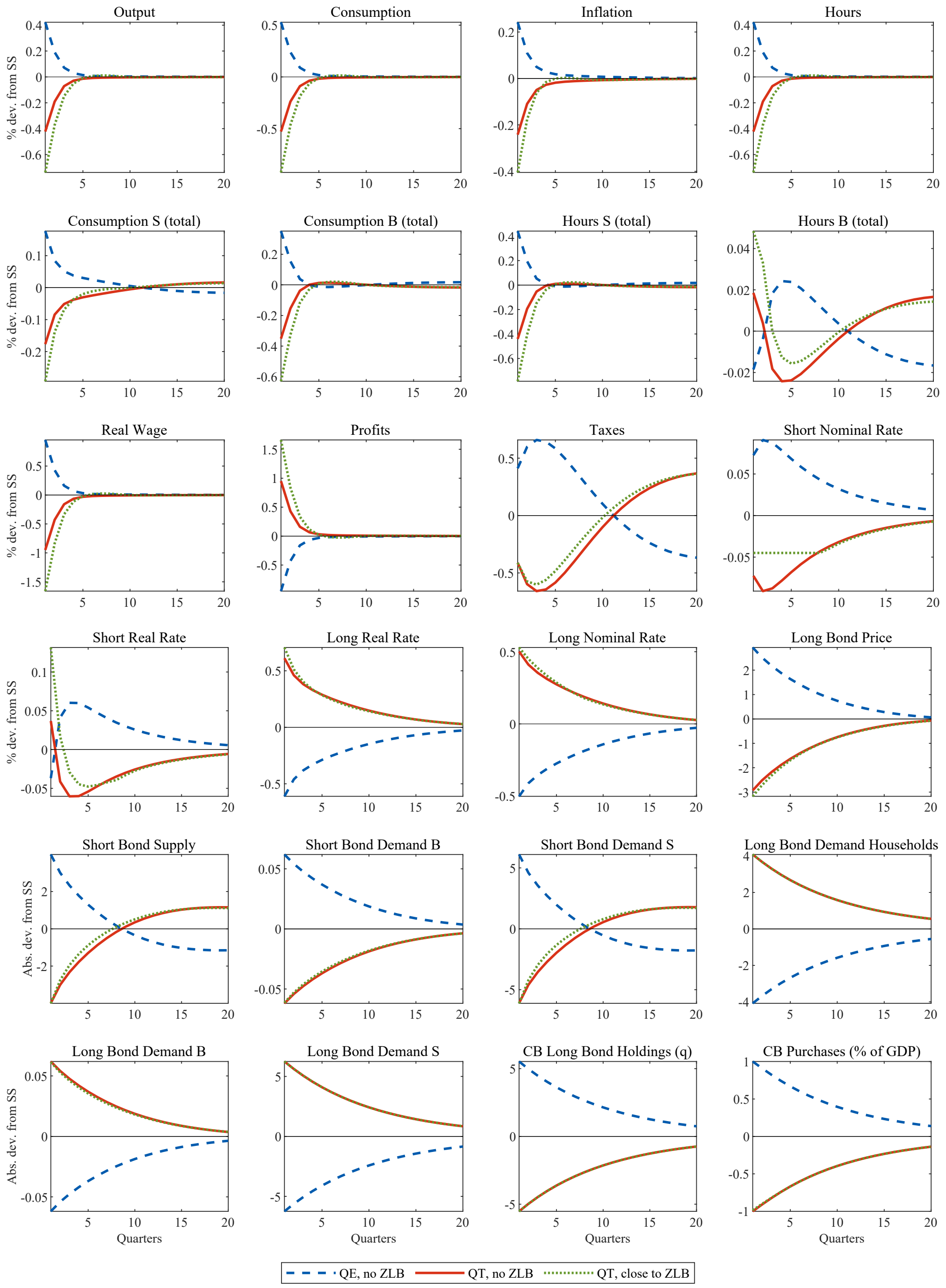
Table 3: Model overview of the TANK-BS model with asset market operations

Labor supply	$w_t = \zeta^j (N_t^j)^{\varphi} (c_t^j)^{1/\sigma}, \ \ j=\{B,S\}$
Euler short-term bonds, S	$1 = \beta^S \mathbb{E}_t \left[\frac{\theta_{t+1}}{\theta_t} \left(\frac{c_{t+1}^S}{c_t^S} \right)^{-1/\sigma} \frac{R_t}{\Pi_{t+1}} \right] - \frac{\nu\,\delta^S}{b_t^{S,L}} \left(\delta^S \frac{b_t^S}{b_t^{S,L}} - 1 \right)$
Euler long-term bonds, S	$1 = \beta^S \mathbb{E}_t \left[\frac{\theta_{t+1}}{\theta_t} \left(\frac{c_{t+1}^S}{c_t^S} \right)^{-1/\sigma} \frac{R_{t+1}^L}{\Pi_{t+1}} \right] + \frac{\nu\,\delta^S b_t^S}{\left(b_t^{S,L}\right)^2} \left(\delta^S \frac{b_t^S}{b_t^{S,L}} - 1 \right)$
Budget constraint, S	$c_t^S + b_t^S + b_t^{S,L} = r_{t-1}\, b_{t-1}^S + r_t^L\, b_{t-1}^{S,L} + w_t\, N_t^S + \frac{1-\tau^D}{1-\lambda} d_t - t_t - \Psi_t^S - \frac{tr}{1-\lambda}$
Euler short-term bonds, B	$1 = \beta^B \mathbb{E}_t \left[\frac{\theta_{t+1}}{\theta_t} \left(\frac{c_{t+1}^B}{c_t^B} \right)^{-1/\sigma} \frac{R_t}{\Pi_{t+1}} \right] - \frac{\nu\,\delta^B}{b_t^{B,L}} \left(\delta^B \frac{b_t^B}{b_t^{B,L}} - 1 \right) + \psi_t^B$
Euler long-term bonds, B	$1 = \beta^B \mathbb{E}_t \left[\frac{\theta_{t+1}}{\theta_t} \left(\frac{c_{t+1}^B}{c_t^B} \right)^{-1/\sigma} \frac{R_{t+1}^L}{\Pi_{t+1}} \right] + \frac{\nu\,\delta^B b_t^B}{\left(b_t^{B,L}\right)^2} \left(\delta^B \frac{b_t^B}{b_t^{B,L}} - 1 \right) + \psi_t^B$
Budget constraint, B	$c_t^B + b_t^B + b_t^{B,L} = r_{t-1}\, b_{t-1}^B + r_t^L\, b_{t-1}^{B,L} + w_t\, N_t^B + \frac{\tau^D}{\lambda} d_t - t_t - \Psi_t^B + \frac{tr}{\lambda}$
Borrowing constraint	$-b_t^B - b_t^{B,L} \leq \bar{D}$
Portfolio adjustment cost	$\Psi_t^j = \frac{\nu}{2} \left(\delta^j \frac{b_t^j}{b_t^{j,L}} - 1 \right)^2, \ \ j=\{B,S\}$
Labor demand	$w_t = mc_t \frac{y_t}{N_t}$
Production function	$y_t = z_t\, N_t$
Profits, aggregate	$d_t = \left[1 - mc_t - \frac{\phi_p}{2} (\Pi_t - 1)^2 \right] y_t$
Phillips curve	$\phi_p (\Pi_t - 1)\,\Pi_t = \varepsilon\, mc_t + \left(1+\tau^S\right)(1-\varepsilon) + \beta^S \mathbb{E}_t \left[\frac{\theta_{t+1}}{\theta_t} \left(\frac{c_{t+1}^S}{c_t^S} \right)^{-\frac{1}{\sigma}} \phi_p (\Pi_{t+1} - 1)\,\Pi_{t+1} \frac{y_{t+1}}{y_t} \right]$
Government budget constraint	$b_t + b_t^L = r_{t-1}\, b_{t-1} + r_t^L\, b_{t-1}^L + \Omega_t + g_t - t_t$
Real short-term interest rate	$r_t = \frac{R_t}{\mathbb{E}_t \Pi_{t+1}}$
Nominal long-term bond return	$R_t^L = \frac{1+\chi V_t}{V_{t-1}}$
Real long-term bond return	$r_t^L = \frac{R_t^L}{\Pi_t}$
Net bond purchases, CB	$\Omega_t = b_t^{CB,L} - r_t^L\, b_{t-1}^{CB,L}$
Value bond purchases, CB	$b_t^{CB,L} = q_t\, b_t^L$
Taylor rule	$\log\left(\frac{R_t}{R}\right) = \rho_r \log\left(\frac{R_{t-1}}{R}\right) + (1-\rho_r) \left[\phi_\pi \log\left(\frac{\Pi_t}{\Pi}\right) \right] + \varepsilon_t^m$
QE shock rule	$\log\left(\frac{q_t}{q}\right) = \rho_q \log\left(\frac{q_{t-1}}{q}\right) + \varepsilon_t^q$
Fiscal rule	$\frac{t_t}{t} = \left(\frac{t_{t-1}}{t}\right)^{\rho^{\tau,t}} \left(\frac{b_t + b_t^L}{b + b^L}\right)^{\rho^{\tau,b}} \left(\frac{g_t}{g}\right)^{\rho^{\tau,g}}$
Aggregate consumption	$c_t = \lambda c_t^B + (1-\lambda) c_t^S$
Aggregate labor	$N_t = \lambda N_t^B + (1-\lambda) N_t^S$
Short-term bonds market clearing	$b_t = \lambda b_t^B + (1-\lambda) b_t^S$
Long-term bonds market clearing	$b_t^L = \left(\lambda\, b_t^{B,L} + (1-\lambda)\, b_t^{S,L} \right) + b_t^{CB,L}$
Resource constraint	$y_t = c_t + g_t + \frac{\phi_p}{2} (\Pi_t - 1)^2\, y_t$
Other shock rules	$\log\left(\frac{x_t}{x}\right) = \rho_x \log\left(\frac{x_{t-1}}{x}\right) + \varepsilon_t^x, \ \ x = \{g, b^L, z, \theta\}$

B Full sets of impulse responses

B.1 QE/QT and QT near the ZLB

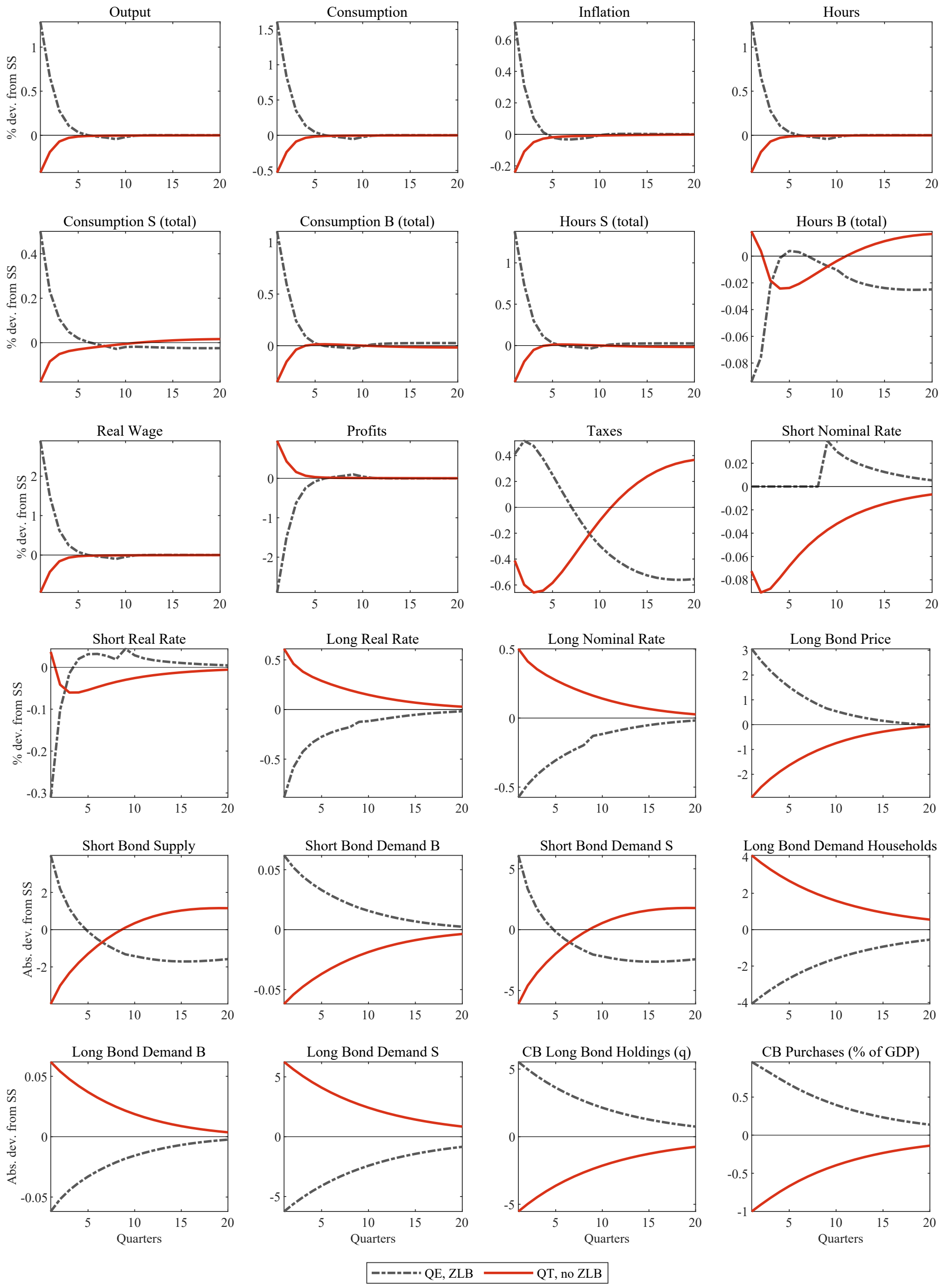
Figure 6: Impulse responses to a QE/QT shock and a QT shock near the ZLB



Notes: This figure depicts the impulse responses to a QE (dashed blue line) and a QT (solid red line) shock occurring far enough above the ZLB, and a QT shock happening close to the ZLB (dotted green line, simulated with $\beta^S = 0.99955$). The shock for each simulation is an asset purchase/sale of size 1% of annualized GDP. Responses for individual consumption levels and hours are weighted by population shares of savers (S) and borrowers (B), respectively, and thus represent total responses.

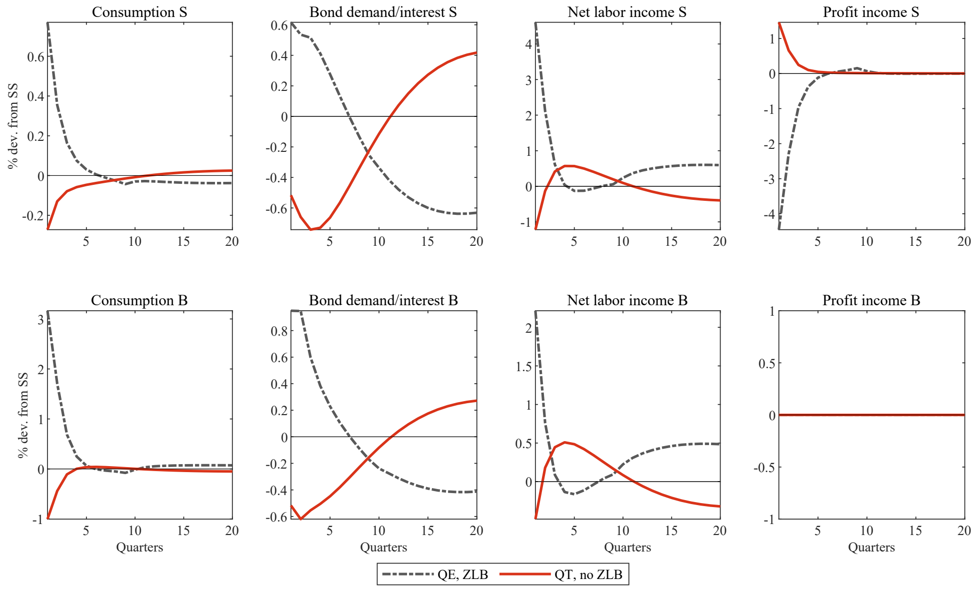
B.2 QE at the ZLB and QT off the ZLB

Figure 7: Impulse responses to a QE shock at the ZLB and a QT shock off the ZLB



Notes: This figure depicts the impulse responses to a QE shock when the ZLB on the policy rate is binding (dash-dotted gray line, showing the impact of QE net of a preference shock), and a QT shock occurring far enough above the ZLB (solid red line). The shock for each simulation is an asset purchase/sale of size 1% of annualized GDP. For QE, the size of the preference shock is chosen such that the ZLB binds for eight quarters. Responses for individual consumption levels and hours are weighted by population shares of savers (S) and borrowers (B), respectively, and thus represent total responses.

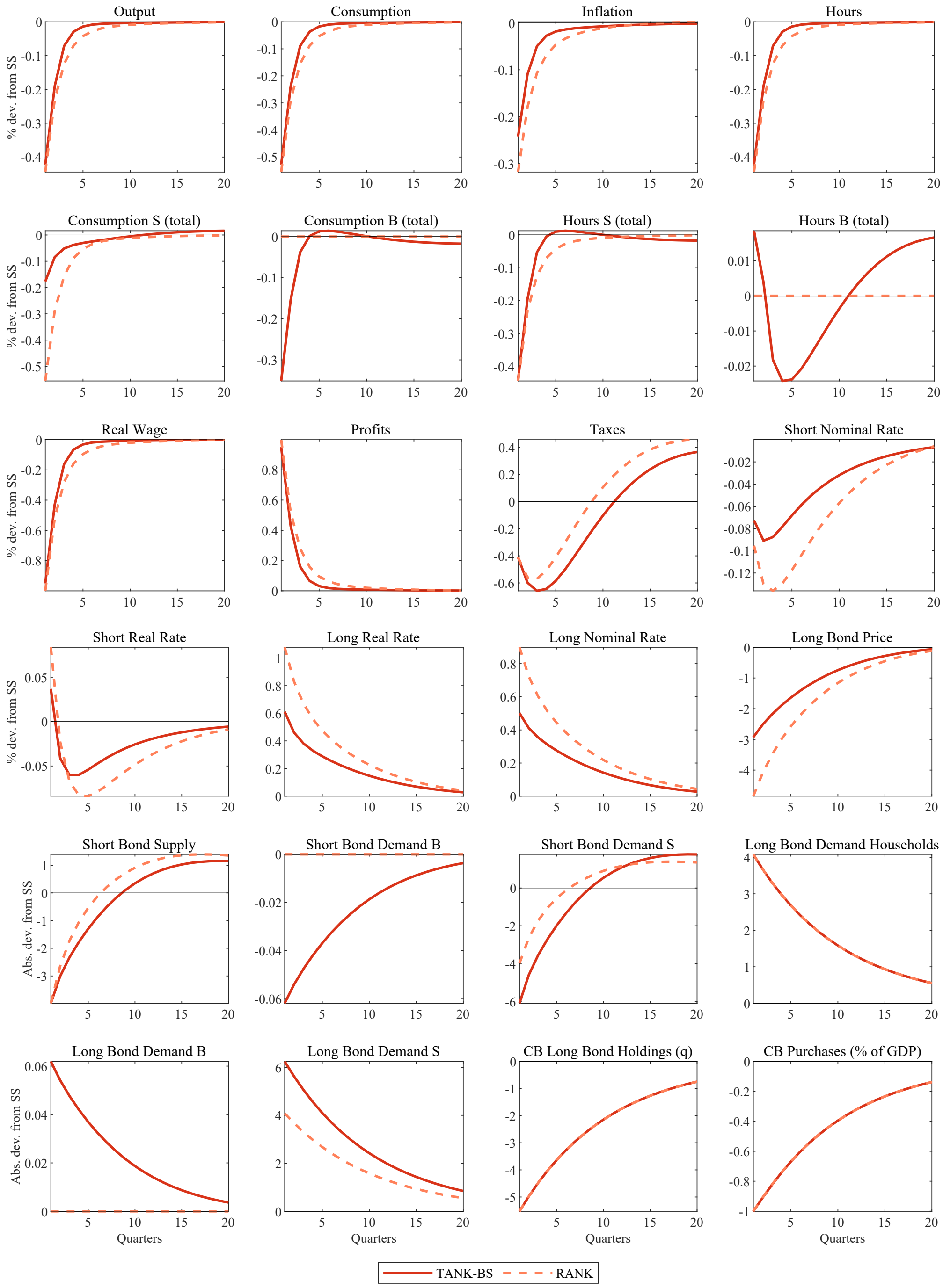
Figure 8: Households' budget components to a QE shock at the ZLB and a QT shock off the ZLB



Notes: This figure shows grouped components of the budget constraints of savers (top) and borrowers (bottom) in response to a QE shock when the ZLB on the policy rate is binding (dash-dotted gray line, showing the impact of QE net of a negative preference shock), and a QT shock occurring far enough above the ZLB (solid red line). The shock for each simulation is an asset purchase/sale of size 1% of annualized GDP. For QE, the size of the preference shock is chosen such that the ZLB binds for eight quarters. Each panel consists of four columns, containing the responses of individual consumption, bond-related variables (bond demand, interest payments/income, net of adjustment cost), labor income net of taxes, and income from profits. All responses are shown in per-capita terms.

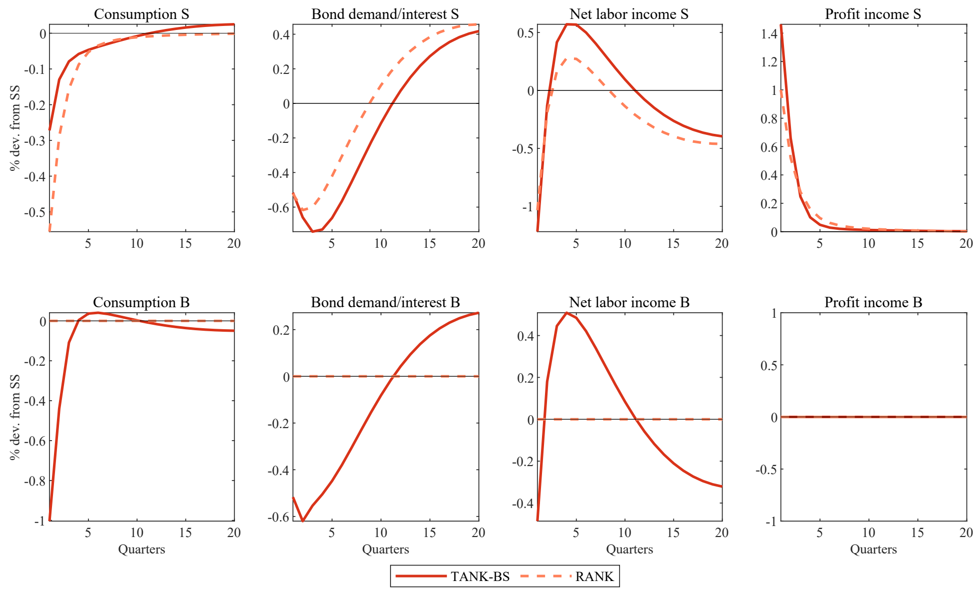
B.3 QT off the ZLB: RANK vs. TANK-BS

Figure 9: Impulse responses to a QT shock off the ZLB: RANK vs. TANK-BS



Notes: This figure depicts the impulse responses to a QT shock occurring far enough above the ZLB, for the borrower-saver model (TANK-BS, solid red line) and its representative-agent counterpart for $\lambda = 0$ (RANK, dashed light red line). The shock for each simulation is an asset sale of size 1% of annualized GDP. Responses for individual consumption levels and hours are weighted by population shares of savers (S) and borrowers (B), respectively, and thus represent total responses.

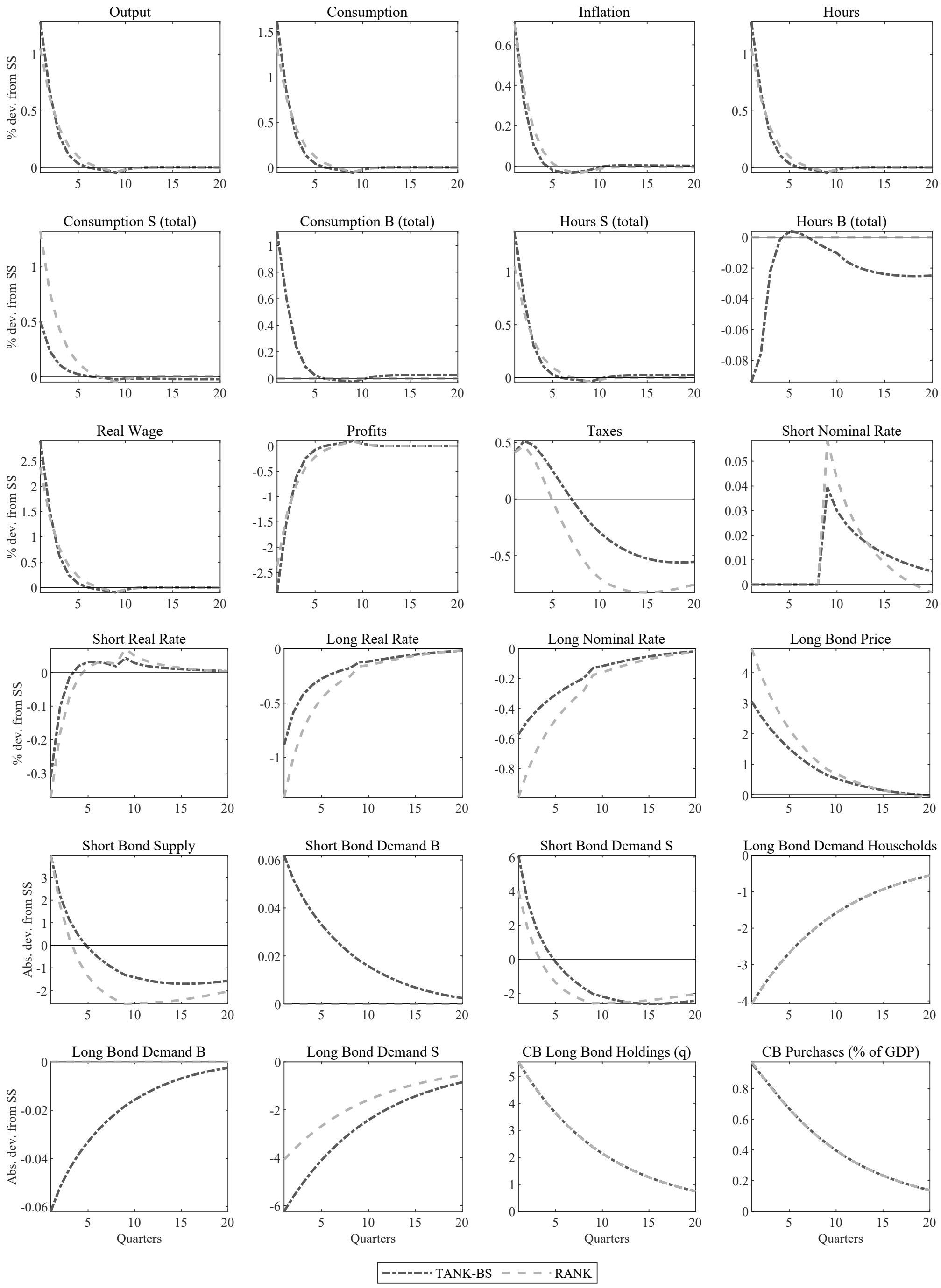
Figure 10: Households' budget components to a QT shock off the ZLB: RANK vs. TANK-BS



Notes: This figure shows grouped components of the budget constraints of savers (top) and borrowers (bottom) in response to a QT shock occurring far enough above the ZLB, for the borrower-saver model (TANK-BS, solid red line) and its representative-agent counterpart for $\lambda = 0$ (RANK, dashed light red line). The shock for each simulation is an asset sale of size 1% of annualized GDP. Each panel consists of four columns, containing the responses of individual consumption, bond-related variables (bond demand, interest payments/income, net of adjustment cost), labor income net of taxes, and income from profits. All responses are shown in per-capita terms.

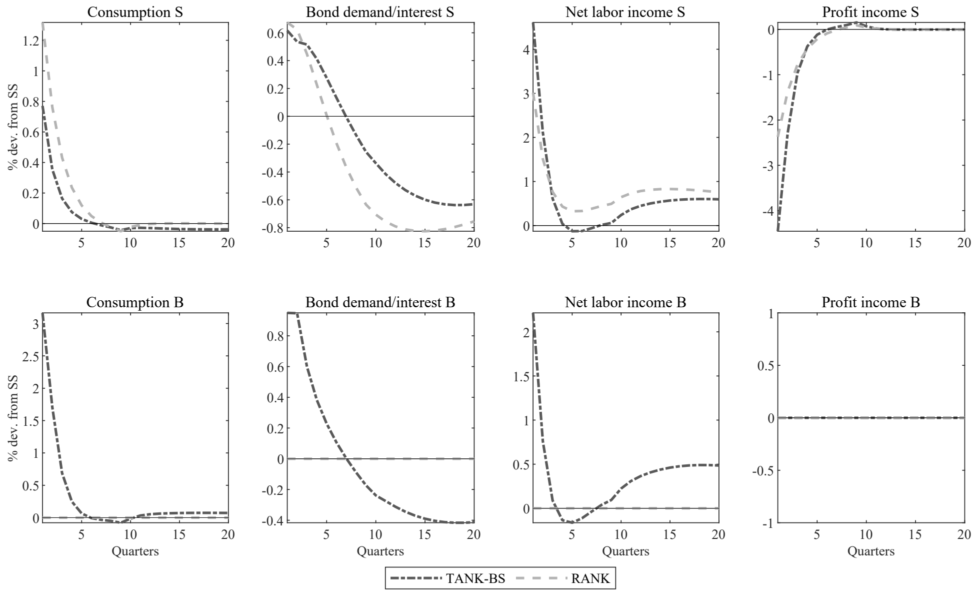
B.4 QE at the ZLB: RANK vs. TANK-BS

Figure 11: Impulse responses to a QE shock at the ZLB: RANK vs. TANK-BS



Notes: This figure depicts the impulse responses to a QE shock net of a negative preference shock when the ZLB on the policy rate is binding, for the borrower-saver model (TANK-BS, dash-dotted gray line) and its representative-agent counterpart for $\lambda = 0$ (RANK, dashed light gray line). The shock for each simulation is an asset purchase of size 1% of annualized GDP. The size of the preference shock is chosen such that the ZLB binds for eight quarters. Responses for individual consumption levels and hours are weighted by population shares of savers (S) and borrowers (B), respectively, and thus represent total responses.

Figure 12: Households' budget components to a QE shock at the ZLB: RANK vs. TANK-BS



Notes: This figure shows grouped components of the budget constraints of savers (top) and borrowers (bottom) in response to a QE shock net of a negative preference shock when the ZLB on the policy rate is binding, for the borrower-saver model (TANK-BS, dash-dotted gray line) and its representative-agent counterpart for $\lambda = 0$ (RANK, dashed light gray line). The shock for each simulation is an asset purchase of size 1% of annualized GDP. The size of the preference shock is chosen such that the ZLB binds for eight quarters. Each panel consists of four columns, containing the responses of individual consumption, bond-related variables (bond demand, interest payments/income, net of adjustment cost), labor income net of taxes, and income from profits. All responses are shown in per-capita terms.