Quantitative Easing with Heterogenous Agents*

Wei Cui[†] and Vincent Sterk[†‡]

[†]University College London and CfM
[‡]CEPR

February, 2021

Abstract

We study the effects of Quantitative Easing (QE) in a heterogeneous-agents model with liquid and partially liquid wealth, as well as nominal rigidities. The direct macroeconomic effect of QE is determined by the difference in marginal propensities to consume out of the two types of wealth, which is large according to empirical studies. Therefore, the effects of QE on aggregate output and inflation are large, according to the model. Indeed, the estimated model reveals that QE interventions greatly dampened the U.S. Great Recession. However, QE may have strong and adverse distributional effects, compared to interest rate policy.

JEL: E21, E30, E50, E58

Key words: Monetary Policy, Large-scale Asset Purchases, Liquidity

*Email: `w.cui@ucl.ac.uk` and `v.sterk@ucl.ac.uk`. We are grateful for editorial guidance by Yuriy Gorodnichenko and Eric Sims. We also thank seminar and conference participants at the Banco de Portugal, the Bank of England, Deutsche Bundesbank, Copenhagen Business School, EIEF, King's College London, University College London, University of Bristol, University of Durham, University of Essex, the 2018 HKUST-Jinan Conference on Macroeconomics, the 2018 Quantitative Macroeconomic Workshop at the Reserve Bank of Australia, AMCM conference in Lillehammer, the 2019 EABCN conference in Warsaw, the EEA/ESEM meetings 2019, ESSIM 2019, the 2019 NYFED - HKMA joint conference, and SED 2019 for helpful comments. We also thank Jihyun Bae and Carlo Galli for excellent research assistance. An earlier version of this paper is available under the title "Quantitative Easing" as CEPR discussion paper 13322, and CfM discussion paper, CFM-DP2018-30, November 2018.

1 Introduction

It has been over ten years since the U.S. Federal Reserve (Fed) initiated a colossal expansion of its balance sheet; the largest since the Great Depression. The 2008 financial crisis compelled the Fed to start providing loans to the banking sector, which was suffering from a freeze of interbank lending. However, as banks recovered from the crisis the Fed did not shrink its balance sheet but instead expanded it further, buying up assets such as long-term government debt and mortgage-backed securities in large quantities. This was done in a bid to stimulate aggregate demand, which slumped during the Great Recession. Known as *Quantitative Easing (QE)*, these interventions acted as a placeholder for conventional monetary policy, which had become powerless as the policy rate had hit the zero lower bound. Similar interventions took place in the UK and the Euro Area, as well as in Japan during the early 2000s.

Since the Covid-19 crisis, central banks have started new rounds of QE. However, in doing so they receive little guidance from economic theory, as this type of policy is completely ineffective in modern textbook models, such as the standard representative-agent New Keynesian (NK) model. This result is known as "Wallace Neutrality", see Wallace (1981), and more recently Woodford (2012). However, when households face uninsurable income risk, as well as borrowing constraints and liquidity frictions, which is realistically the case, Wallace Neutrality breaks down and QE becomes a potential tool for monetary policy.

This paper presents a Heterogeneous-Agents New Keynesian (HANK) model with liquidity frictions to provide a better understanding of how household heterogeneity and incomplete markets affect the pass-through of liquidity policy to the real economy. Households can save in fully liquid assets, and also hold less liquid shares in mutual funds. We further allow for sticky prices. In this setting, liquidity policy such as QE can have powerful effects on aggregate demand, but can also create strong side effects which may exacerbate inequality across households.

In the model, asset purchases by the central bank change the composition of liquid versus less liquid assets held by the public (as well as the prices of these assets). This variation in asset composition shifts the aggregate demand for goods, and hence the real economy, since households have different propensities to consume out of different types of assets, due to the presence of liquidity frictions. The aggregate demand effect in turn feeds back into asset markets. We focus on liquidity policy at times when the nominal interest rates are constrained at zero, i.e. QE, and compare the macroeconomic and distributional effects of such policy to those of changes in nominal interest rates (the conventional policy lever in the New Keynesian model).

More specifically, we model QE as a purchase of long-term debt by the central bank, financed by the issuance of reserves.[1] On the sell side of the transaction are mutual funds, who receive

[1] We will show that the long-term debt can be interpreted either as government debt or household debt.

newly created deposits in exchange for the long-term debt being sold.[2] They, however, derive little value from liquidity and immediately trade in the deposits for newly issued debt, which offers a higher yield than deposits. This response pushes up the price of long-term debt, i.e., the long-term real interest rate falls. The lower cost of long-term borrowing in turn induces households to hold more liquidity.[3] Hence, the newly created deposits end up in the hands of households, who value liquidity for self-insurance reasons. Because households have high Marginal Propensities to Consume (MPCs) out of liquid wealth, the additional liquidity boosts aggregate spending on goods, which increases aggregate output.

We first present a simple formula which captures the essence of the QE transmission mechanism in the model and which can be used for back-of-the-envelope calculations. The key insight conveyed by this formula is that the direct effect of QE depends on the *difference* between the MPCs out of deposits and less liquid sources of wealth. Empirical estimates in the literature suggest that the gap between these two MPCs is large. An increase in household liquidity may therefore boost aggregate demand substantially. In Section 2, we will discuss in more detail how QE leads to the creation of deposits, and how this liquidity can end up in the hands of households.

Sections 3 and 4 present the quantitative model. We evaluate the model's implications for consumption at the micro level and show that it generates a large gap in MPCs out of liquid and less liquid wealth, in line with empirical studies. A subset of the parameters is estimated by Maximum Likelihood, using the data on household deposits, as well as other macroeconomic time series. With the parametrized model at hand, we evaluate the macroeconomic and distributional effects of QE.

Our first main finding is that QE can have strong stimulative effects on the macro economy. In particular, we find that QE had a large and positive impact on output and inflation during the Great Recession, preventing a much deeper downturn. This result follows from a counterfactual simulation in which we shut down QE interventions. The exercise also reveals that the effects of QE during its first round were stronger than during the second and the third rounds.

These results underscore the importance of liquidity in an incomplete-markets world. Most work in New Keynesian economics (even with financial frictions) abstracts from money/liquid assets altogether. With a representative agent the presence of liquidity tends to have very limited implications for policy transmission (see, e.g., the textbook treatments in Woodford (2003) and Galí (2008)). A key point of our paper is that, once household heterogeneity and incomplete markets are integrated into the model, liquidity emerges as a quantitatively powerful lever of monetary policy.

[2] Deposit creation is necessarily triggered because mutual funds –unlike banks– cannot directly hold reserves at the central bank. In the U.S., only a small fraction of the QE assets were purchased from banks, see Section 2 for more discussion and for empirical evidence.

[3] In the baseline model, the long-term debt is issued by the government. An increase in the price of debt then allows the government to cut taxes, which allows households to purchase more liquid assets. However, we show that the same outcomes are obtained when the long-term debt is instead issued by households. Moreover, we conduct several robustness checks regarding assumptions on the fiscal policy response to QE.

A second main finding is that changes in aggregate liquidity, as engineered by QE operations, have much stronger distributional effects than changes in nominal interest rates. We arrive at this result by comparing the effects of QE expansions to those of a "conventional" cut in interest rates (without liquidity effects). For a similar macroeconomic expansion, the redistributive effects of QE are much greater. Our results suggest that expansionary QE policy reduces consumption/income inequality *initially* but increases inequality later on, primarily via a growing discrepancies in wealth accumulation across agents. This is in contrast to a cut in the policy rate, which has similar aggregate effects but persistently reduces inequality, as found empirically by Coibion et al. (2017).

In the comparison between QE and interest rate policy, we abstract from the liquidity effects of the latter, which in reality is often implemented via Open Market Operations (OMO) as well. We do so in order to maximally contrast the liquidity channel and the interest rate channel. But we do acknowledge that, realistically, conventional policy implemented via OMO may create liquidity effects. However, empirical evidence in Section 2 suggests that the liquidity effects of conventional policy are smaller than those of QE. This may be the case because conventional OMO typically involves a swap of reserves/money for *short-term* government bonds. To the extent that short-term government debt is closer to reserves in terms of liquidity than long-term debt, the liquidity effects of conventional policy are smaller than those of QE.

We conduct a number of robustness checks, regarding assumptions on fiscal policy and the behavior of mutual funds. In Section 5, we further consider robustness with respect to the introduction of capital, which we abstract from in the baseline model in order to highlight the key transmission channel. The key channel in the baseline model works through household consumption, and to allow for a better a comparison with smaller-scale New Keynesian models which often abstract from capital. However, investment is generally considered to be an crucial component of the transmission of QE and it is therefore important to investigate how its introduction affects the transmission of QE in the model.

We find that the effect of a QE expansion on output is similar to that in the baseline without capital. However, the increase in output is now driven by both consumption and investment. Intuitively, the boom in investment is driven both by direct channels, as investors replace government bonds by capital investment, and indirect equilibrium channels, as the increase in aggregate consumption demand triggers an increase in goods demand, and hence investment demand.

Finally, the technical aspects of this paper may be of independent interest. In particular, we show how to keep track of the distribution of liquid wealth in a parsimonious yet accurate way, exploiting the fact that households' holdings of deposits are low in the data. We then exploit the model's tractability to devise a fast solution method so that the model can be estimated via a standard Maximum Likelihood procedure and can be used to assess the effects of QE.

Related literature. We contribute to a fast-growing literature which explores various, mostly complementary channels via which QE could affect inflation and the real economy. The literature has studied the transmission of QE via bank lending to firms, mortgages, portfolio rebalancing, stock markets, exchange rates, signaling of future policy, which are not mutually exclusive and may all be important in reality. Our model is not designed to (fully) capture all of these channels. Instead, our aim is specifically to understand how households facing uninsurable idiosyncratic risks because of incomplete markets can affect the transmission of QE.

A number of authors have studied QE in a representative-agent Dynamic Stochastic General Equilibrium (DSGE) model. Chen, Curdia and Ferrero (2012) analyze QE in a medium-scale DSGE model with segmented asset markets. They find that QE only has small effects. Large effects are found by Del Negro, Egertsson, Ferrero and Kiyotaki (2017), who develop a quantitative model to evaluate the effects of liquidity provisions during the financial crisis. In their model, liquidity interventions ease financial constraints on the production side of the economy. A similar channel operates in Gertler and Karadi (2012). Wen (2014) studies the QE exiting strategy and the impact on firms. By contrast, we focus on the role of QE as a direct instrument to manage aggregate demand, which has been used well *beyond* the financial crisis. Campbell (2014) considers the implications of QE for the occurrence of liquidity traps, whereas Harrison (2017) studies optimal QE policy in a representative-agent model with portfolio adjustment costs. Finally, Sims and Wu (2020) study a DSGE model in which QE stimulates economic activity by relaxing leverage constraints faced by financial intermediaries, a channel we abstract from.

We view our contribution as complementary to these studies, as we study channels created by incomplete markets on the household side.[4] Our findings underscore the importance of liquidity for macroeconomic outcomes. Under incomplete markets, liquidity and aggregate demand are closely linked, which enables the central bank to exercise control over aggregate demand via liquidity management, even when the nominal policy rate is fixed. We further emphasize that unconventional monetary policy has distributional effects under incomplete markets. Empirically, the existence of conventional monetary policy's distributional effects is well established, see for instance Doepke and Schneider (2006).

The emphasis on liquidity connects our work to the search models following Lagos and Wright (2005), which typically imply a degenerate distribution of liquidity. Recent contributions by Rocheteau, Weill and Wong (2019) study the interaction between (non-degenerate) distributions of liquid asset holdings and labor income in this type of model. We instead use a heterogeneous-agents model in the Bewley-Huggett-Aiyagari tradition and analyze liquidity provision policy via the central bank's balance sheet. By adding nominal rigidities, we allow for a quantitative comparison to

[4]It would be interesting to explore how this incomplete-markets channel interacts with other channels proposed in the literature, although this is beyond the scope of the present paper.

conventional interest rate channels studied in the New Keynesian literature.[5]

The neutrality of central bank balance-sheet policies in complete-markets models was originally established by Wallace (1981), and reiterated more recently by Woodford (2012). The underlying theoretical argument is a variation on the Modigliani-Miller and is related to the Ricardian Equivalence, cf. Barro (1974). Perhaps in part because of this neutrality result, much of the recent NK literature on unconventional monetary policy has focused on Forward Guidance rather than on QE, see e.g. Del Negro, Giannoni and Patterson (2012) and McKay, Nakamura and Steinsson (2016).

The importance of household liquidity for monetary policy is emphasized by Bilbiie and Ragot (2016). They show that liquidity frictions change the output-inflation trade-off, as inflation affects the extent to which households can self-insure using nominal assets. Cui (2016) studies monetary-fiscal interactions in a model in which the liquidity of different asset classes differs endogenously, but without QE. Caballero and Farhi (2017) consider monetary policy –including QE– in a model with safe and risky assets and heterogeneity in risk tolerance. Finally, Ray (2018) introduces segmented markets and "preferred habitats" to representative-household NK model, creating a disconnect between long and short rates. While our model also features such a disconnect, household heterogeneity takes center stage in the transmission of QE in our model.

Our results further highlight the quantitative importance of interactions between nominal rigidities and market incompleteness in the transmission of QE. A number of recent papers have studied such interactions in the context of conventional monetary policy, see e.g. Gornemann, Kuester and Nakajima (2016), Auclert (2016), Luetticke (2015), Ravn and Sterk (2016), Kaplan, Moll and Violante (2017), Debortoli and Galí (2017), and Hagedorn, Luo, Mitman and Manovskii (2017).

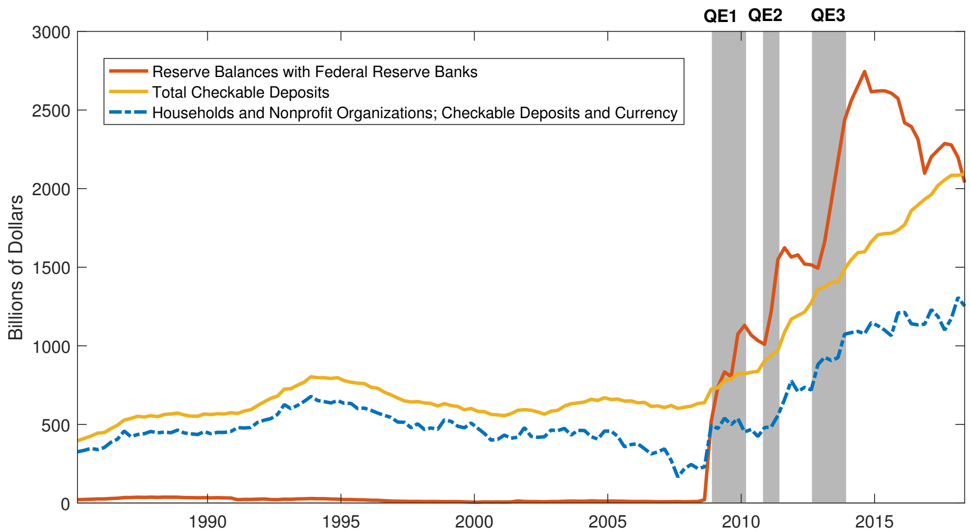
Finally, various authors have studied the empirical effects of large-scale asset purchases, generally finding evidence for expansionary macro effects. For example, Weale and Wieladek (2016) find that in the U.S., an asset purchase of 1% of annual GDP leads to an increase in real GDP of 0.58% and an increase in inflation of 0.62%. Di Maggio, Kermani and Palmer (2016) use data on mortgage origination to provide evidence that QE stimulated aggregate consumption substantially. A survey of the broader literature on this topic can be found in Bhattarai and Neely (2016).

2 A first glance at the macroeconomic effects of increased household liquidity

Before presenting the full model, we provide a more detailed empirical account of the transmission of Quantitative Easing to households in the United States, in light of the model to be described in the next section. Figure 1 shows the evolution of reserves at the Fed, the aggregate amount of checkable

[5]Heterogeneity also plays a role in Sterk and Tenreyro (2018), who study the distributional effects of open-market operations in a flexible-price model.

Figure 1: Reserves and deposits in the US



Source: Federal Reserve Board, Flow of Funds accounts. The grey areas denote rounds of QE purchases by the Federal Reserve.

deposits, and the amount of deposits/currency held by households and non-profits. As large-scale asset purchases began, all three series increased sharply. This strongly suggests that QE triggered the creation of additional deposits, which in large part ended up being held by households. In Appendix A, we further draw a comparison between the Flow of Funds data shown in Figure 1 and data from the Survey of Consumer Finances, both of which show a similar pattern since 2009. There, we also provide supporting empirical evidence based on a local projection exercise.

2.1 Transmission of liquidity

The role of banks. Banks played an important intermediary role in the transmission of QE, for at least two reasons. First, only banks can hold reserves at the Federal Reserve and create deposits. Second, when purchasing assets the Fed trades with primary dealers, which are typically banks.

However, these facts do not imply that banks were the owners of the assets that were purchased under QE. Indeed, banks played only a modest role as sellers of assets to the Fed. This is shown by for instance Carpenter, Demiralp, Ihrig and Klee (2015), who investigate in detail which investors the Fed purchased from. They show that during the two decades prior to QE, banks held only about 7% of the total amount of treasury debt and mortgage-backed securities. They also provide more formal econometric analysis implying only a minor role for banks. Indeed, the Fed could not even have purchased the QE assets exclusively from banks. Flow of Funds data show that at the start of QE, commercial banks held about \$1.2 trillion worth of treasury debt and MBS. While this number may seem large, it is less than half of the total value of assets that were purchased under QE: Figure 1 shows that the increase in reserves following QE was about \$2.5 trillion.

The role of mutual funds and other non-bank entities. Using Flow of Funds data, Carpenter et al. (2015) report that the major sellers of assets to the Fed during QE were household-held mutual funds, pension funds, broker-dealers, and insurance companies. Importantly, none of these entities has the ability to hold reserves at the Federal Reserve, which implies that QE purchases from these entities must have led to deposit creation through banks.

To understand this point, consider a mutual fund which sells \$100 million worth of assets to the Fed. In turn, the Fed finances the purchase by issuing \$100 million worth or reserves. Immediately after the sale, the fund will hold \$100 million of additional deposits at a bank, which in turn balances this liability by holding \$100 million of additional reserves at the Fed. The bank thus serves as a middleman who increases its deposit liabilities and reserve holdings by the same amount.

Of course, the mutual fund may subsequently choose to offload the additional cash/deposits from its balance sheet, as cash offers a low return and the fund may have little use for extra liquidity. Indeed, empirical evidence shows that this is what happened following QE in the U.S.. Goldstein, Witmer and Yang (2015) use micro-level data on the behavior of mutual funds following QE. They report that mutual funds did not increase their cash holdings following the asset sales to the Fed.

The role of households and the fiscal authority. What did the mutual funds do with the additional liquidity and how did it end up in the hands of households? One possibility is that, following QE, there was a direct net outflow of cash from the mutual funds, which happens when the funds increase dividends to households, when raises less new investment from households, or when households take wealth out of the fund. The empirical importance of this channel seems rather limited, however, as outflows from mutual funds who sold assets to the Fed were only moderate, see e.g., Goldstein et al. (2015).[6] They show that, instead, mutual funds mainly replaced the assets sold to the Fed with newly issued government debt and/or MBS.

Still, purchases of MBS can directly leave additional liquidity in the hands of households. For instance, Di Maggio, Kermani and Palmer (2016) show that QE led to an increase in new mortgage lending. In case of purchases of newly issued government debt, liquidity leaves the mutual fund sector via the government, which then uses it to lower taxes, to increase transfers, and/or to purchase goods and services (and we will discuss these aspects in our model). Again, the newly created liquidity ultimately flows to households (and to some extent firms), although less directly.

[6] They also find little evidence of a rebalancing towards other asset classes such as firm equity within mutual funds.

The role of asset prices. So far, our discussion on the empirical evidence has focused on the effect of QE on asset volumes. The empirical finance literature has extensively studied the effects of QE on asset prices often reporting substantial effects. For instance, Krishnamurthy and Vissing-Jorgenson (2011) find that QE1 and QE2 had substantial negative effects on, respectively, MBS and Treasury yields. Thus, QE seems to have successfully increased the total demand for the assets being purchased. In our model, such asset price effects (represented by lower long-term interest rates) are an important component of the QE transmission mechanism through aggregate demand, as we will discuss detail in the next section.

2.2 Macroeconomic effects

While the effects of QE of asset markets can be studied by exploiting high-frequency movements around QE events, the macroeconomic effects are much more difficult to tease out empirically, if only because macro data are available at a much lower frequency than financial data. Therefore, our approach is to build a full-blown structural model. The model not only incorporates the immediate effects of QE on asset prices, but also feedback effects from the macro economy on asset prices.

That said, it is possible to gauge some of the effects of QE on aggregate demand with a simple formula, which can be confronted with empirical evidence. To this end, let us postulate an aggregate consumption demand function $C(L, I, \Gamma)$, where L denotes the (nominal) value of fully liquid assets held by households (e.g. deposits), and I denotes the value of their illiquid, or partially liquid assets (e.g. assets owned via mutual funds). The third argument, Γ, contains other relevant aspects of individual states and the overall economy, such as asset prices and wages, and is denoted by a scalar for simplicity. The (average) marginal propensities to consume out of liquid and illiquid wealth are given by the respective derivatives of the aggregate demand function, and will be denoted by $MPC^L \equiv C_L(L, I, \Gamma)$ and $MPC^I \equiv C_I(L, I, \Gamma)$.

When the central bank conducts QE, it purchases I in exchange for L. The mechanics are the following. The central bank purchases long-term government debt from mutual funds (or financial institutions that hold government debt) by issuing reserves; the banks act as a pure middle man, sourcing the bonds from mutual funds in exchange for deposits, and then selling bonds to the Fed in exchange for reserves. Mutual funds have no use for the additional amount of liquidity, and are induced to offload the additional liquid assets from their balance sheets. They may do so either by paying dividends directly to households, by raising fewer new inflows from households, by purchasing new privately-issued assets such as mortgages and/or mortgages-back securities, or by purchasing newly issued government debt which would allow for lower taxes or higher government spending. Either way, the liquid wealths from QE ultimately end up in the hands of those agents who have a need for them, i.e. households, who use liquidity for self-insurance purposes.[7]

[7] In the quantitative model, we will discuss the transmission of liquidity to households in more detail, which could

Since these are all voluntary trades, QE does not directly change the total amount of wealth owned by households, i.e. any increase in L is matched by a decrease in I of the same magnitude. However, the intervention does affect the *composition* of wealth held by the public. Denoting the value of assets purchased under QE by Δ^{QE}, the consumption function becomes $C(L + \Delta^{QE}, I - \Delta^{QE}, \Gamma(\Delta^{QE}))$. By differentiating this function with respect to Δ^{QE}, we obtain the following formula for the marginal effect of QE on aggregate demand:

$$\frac{\partial C}{\partial \Delta^{QE}} = \underbrace{MPC^L - MPC^I}_{\text{direct effect}} + \underbrace{GE}_{\text{indirect effect}},$$

where $GE \equiv C_\Gamma(L, I, \Gamma)\frac{\partial \Gamma}{\partial \Delta^{QE}}$ denotes the general equilibrium effect. This formula splits the effects of QE into "direct" and "indirect" GE effects, in the spirit of a decomposition proposed by Kaplan et al. (2017) for conventional monetary policy.[8]

The first term captures the direct effect. It is the difference between the MPCs out of liquid and illiquid wealth. Intuitively, QE directly triggers a liquidity transformation: it lowers households' illiquid wealth holdings, while increasing their liquid wealth. The direct effect of this transformation on consumption depends on the difference in the marginal propensities to consume out of the two types of wealth. The second term captures the indirect general equilibrium effects triggered by QE.

Simple as it looks, the formula conveys a number of important insights. First, if the two types of wealth were equally liquid to households, as in many standard models, it would hold that $MPC^L = MPC^I$, other things equal. In this case, QE would have no direct effect on aggregate demand, echoing the neutrality result of Wallace (1981). Second, even in the extreme case in which $MPC^I = 0$, QE only has large effects to the extent that the marginal propensity to consume out of liquid wealth, MPC^L, is large. This point provides a way of understanding why for instance Chen et al. (2012) find that QE has small effects on the real economy, as it is well known that MPCs tend to be very small in representative-agent models. On the other hand, models with incomplete markets and borrowing constraints are well known to generate much higher MPCs out of liquid wealth. Moreover, when certain types of assets are subject to liquidity frictions, the MPCs out of these types of wealth tend to be small, even in incomplete-markets models.

Finally, the indirect GE effects depend crucially on the structure of the economy and in particular on price stickiness. With flexible prices, an increase in aggregate consumption demand is typically dampened by an increase in prices. With sticky prices, the increase in aggregate demand might be further amplified.

Are strong direct effects of QE in line with the data, i.e. is the difference between MPC^L

go via taxes and transfers if mutual funds purchase government debt.

[8]In the above formula, asset prices do not explicitly show up. However, they are implicit in Δ^{QE}, which denotes the change in the value of assets, i.e. the change in the price times quantity. In the full model, we will come back more explicitly to asset prices.

and MPC^I large? A substantial body of empirical studies has found MPCs out of fully liquid wealth to be very sizable. For example, Fagereng, Holm and Natvik (2018) estimate an average MPC of 63 percent in the first year, based on high-quality administrative data on Norwegian lottery participants.[9] The literature on MPCs out of less liquid sources of wealth is less extensive, but generally reports much smaller MPCs. Di Maggio, Kermani and Majlesi (2018) use Swedish data to estimate MPCs out of changes in stock market wealth, and estimate these to lie between 5 and 14 percent, much below typical estimates for the MPCs out of fully liquid wealth. Moreover, they report that –among the same individuals– MPCs out of fully liquid dividend payments are much higher. The empirical evidence is thus consistent with sizable direct effects.

Based on the above formula, we can obtain a back-of-the-envelope estimate for the direct effects of QE. This helps to get a sense of the quantitative importance of QE since the Great Recession. Between 2007 and 2017, checkable deposits held by households and non-profit organizations increased from about 1.5 to 6.3 percent of annual GDP.[10] Figure 1 suggests that this increase was largely driven by QE. Assuming $MPC^L = 0.63$ following Fagereng et al. (2018) and $MPC^I = 0.095$, the mid point of estimates provided by Di Maggio et al. (2018), this implies a direct effect of $(6.3 - 1.5) \cdot (0.63 - 0.095) = 2.57$ percent of GDP.

Thus, the data suggest that the direct effects of QE on GDP are substantial. However, the overall effect of QE depends also on the GE response to these direct effects, which includes the effects on goods and asset prices. We will use the model to evaluate the overall effects of QE.

3 QE in an incomplete-markets model

This section presents a fully-fledged general equilibrium model with incomplete markets and with heterogeneous agents who face borrowing constraints and who hold assets with different degrees of liquidity. We use the model to quantify the macro effects of QE and contrast them to the effects of conventional policy. QE is a purchase by the central bank of long-term debt held by mutual funds.[11]

3.1 The baseline model

The economy is populated by households, firms, banks, mutual funds, a treasury, and a central bank.

[9] Parker, Souleles, Johnson and McClelland (2013) report an average quarterly MPC between 50 and 90 percent for the U.S. during the Great Recession. For more discussion on the empirical evidence, see Kaplan et al. (2017).

[10] The amount of checkable deposits held by households and non-profits was \$219 billions in 2007 and \$1,219 billions in 2017. Nominal GDP was \$14,457 billions in 2007 and \$19,485 billions in 2017.

[11] It will become clear that the long-term debt can be interpreted either as issued by the fiscal authority or by households, or both.

Households. There is a continuum of infinitely-lived, ex-ante identical households, indexed by $i \in [0,1]$. Household i's preferences are represented by:

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t U\left(C_t(i), N_t(i)\right), \tag{1}$$

where $C_t(i)$ is a basket of goods consumed in period t, $N_t(i)$ denotes hours worked supplied on a competitive labor market, and $\beta \in (0,1)$ is the subjective discount factor. Moreover, $\mathbb{E}_t$ is the expectations operator conditional on information available in period t, and $U(C,N)$ is a utility function which is increasing and concave in consumption and decreasing in hours. The consumption basket is given by $C_t(i) \equiv \int_0^1 \left(C_t(i,j)^{\frac{\varepsilon_t-1}{\varepsilon_t}} dj\right)^{\frac{\varepsilon_t}{\varepsilon_t-1}}$, where $C_t(i,j)$ denotes the household's consumption of good j and $\varepsilon_t > 1$ is the exogenous elasticity of substitution between goods. Following the NK literature, variations in ε_t can be thought of as "cost push" shocks, since they affect mark-ups charged by firms. Household optimization implies that the price of the consumption basket is given by $P_t = \int_0^1 \left(P_t(j)^{1-\varepsilon_t} dj\right)^{\frac{1}{1-\varepsilon_t}}$, where $P_t(j)$ is the price of good j.

Households are subject to idiosyncratic unemployment risk. When unemployed, the household cannot supply labor, i.e. $N_t(i) = 0$, so it has no labor income. When employed, the household can freely choose the number of hours worked, earning a real wage rate w_t per hour. Unemployed households become employed with a probability p^{UE}, whereas employed households become unemployed with a probability p^{EU}. These transitions are exogenous and take place at the very end of each period. When unemployed, a household receives an unemployment benefit given by $\Theta_t(i) = \Theta^U \geq 0$. This benefit is provided by a government agency which runs a balanced budget by imposing an equal social insurance contribution Θ^E on the employed. That is, when employed the household receives a negative transfer given by $\Theta_t(i) = \Theta^E = -\frac{u}{u-1}\Theta^U \leq 0$, where $u = p^{EU}/(p^{EU} + p^{UE})$ is the unemployment rate.[12]

Households can hold deposits, denoted by $D_t(i)$ in real terms, which pay a nominal interest rate and are fully liquid, in the sense that there are no transaction costs involved. Deposits provide households with a means of self insurance against the idiosyncratic income risks associated with unemployment, helping them to cushion the decline in consumption when they lose their job. However, households must obey a borrowing constraint:

$$D_t(i) \geq \underline{D}, \tag{2}$$

where $-\underline{D}$ is a borrowing limit.[13]

[12] McKay and Reis (2016) provide an in-depth analysis of the stabilization role of social insurance in a NK model with heterogeneous agents. We choose to simplify along this dimension.

[13] In our model, the borrowing constraint remains constant over time. Guerrieri and Lorenzoni (2017) study the effects of shocks to borrowing limits.

Households can also own shares in mutual funds, which may generate higher returns but are less liquid. The evolution of a household's mutual-fund wealth, denoted $A_t(i)$, is given by:

$$A_t(i) = (1 + r_t^A)A_{t-1}(i) - X_t(i) \ , \tag{3}$$

where r_t^A is the real return on the funds, $X_t(i)$ is a withdrawal by the household from the fund. For simplicity, we assume withdrawals vary only by employment status. Specifically, the withdrawals of the employed and unemployed are denoted by two constants, X^E and X^U, respectively. This simplified setup allows for some partial insurance from the illiquid asset to employment risk. In Appendix B.2, we show that this outcome can be micro-founded as the result of an adjustment cost, and our results are robust to alternative modeling of withdrawals (Section 2).

In each period, household i chooses $C_t(i)$, $D_t(i)$, and $N_t(i)$ to maximize (1) subject to (2), (3), the restriction that it can only choose $N_t(i)$ when employed, and a budget constraint specified in real terms as:

$$C_t(i) + D_t(i) = w_t N_t(i) + \frac{R_{t-1}}{\Pi_t} D_{t-1}(i) + \Theta_t(i) + X_t(i) - T_t, \tag{4}$$

where R_{t-1} is the gross nominal interest rate on deposits from period $t-1$ to period t, $\Pi_t = P_t/P_{t-1}$ is the corresponding gross rate of inflation, and T_t is a lump-sum tax levied to finance government expenditures other than benefits.

Firms. Each consumption good is produced by a different firm. The structure of household preferences implies that firms are monopolistically competitive in the goods market. Firms operate a linear technology using labor only, i.e. their output is given by $Y_t(j) = Z_t N_t(j)$. Here, Z_t denotes Total Factor Productivity (TFP).

Firms also face a quadratic cost of price adjustment following Rotemberg (1982), given in real terms by $Adj_t(j) = \frac{\phi}{2}\left(\frac{P_t(j) - P_{t-1}(j)}{P_{t-1}(j)}\right)^2 Y_t$, where $\phi \geq 0$ is a parameter which governs the cost of price adjustment, and $Y_t = \int_0^1 Y_t(j)dj$ denotes aggregate output. We will compare economies with $\phi = 0$, a flexible price economy, and $\phi > 0$, a sticky price economy.

The dividends paid by firm j are given, in real terms, by $Div_t(j) = \frac{P_t(j)}{P_t} Y_t(j) - w_t N_t(j) - Adj_t(j)$ where in equilibrium it holds that $P_t(j) = P_t$. Therefore, aggregate dividends satisfy

$$Div_t = Y_t - w_t N_t - Adj_t, \tag{5}$$

where $Adj_t = \int_0^1 Adj_t(j)dj = \frac{\phi}{2}\left(\Pi_t - 1\right)^2 Y_t$. Firms maximize the present value of profits subject to their production function and the household's demand schedule, which leads to the following relation, commonly known as the New Keynesian Phillips Curve:

$$1 - \varepsilon_t + \varepsilon_t \frac{w_t}{Z_t} = \phi\left(\Pi_t - 1\right)\Pi_t - \phi \mathbb{E}_t \left[\Lambda_{t,t+1} \frac{Y_{t+1}}{Y_t} \left(\Pi_{t+1} - 1\right)\Pi_{t+1}\right], \tag{6}$$

where $\Lambda_{t,t+1}$ is the stochastic discount factor used by the firms.[14] We assume that the distribution of initial prices is the same across firms, so they behave symmetrically and we drop the index j from now on.

Mutual funds. There is a representative mutual fund which owns shares in aggregate equity of the firms (S_t) as well as long-term treasury debt (B_t^m). We model the latter following Woodford (2001). A unit of long-term debt pays ρ^k dollars in any period $t+k+1$ going forward, where $0 \leq \rho < 1$. In the steady state, the duration of long-term government debt is given by $\frac{1}{1-\beta\rho}$. Note that government debt is *fully liquid* to the mutual funds, as they do not face any trading frictions. The equities of the representative mutual fund are the mutual fund shares owned by the households.

Let $X_t \equiv \int_0^1 X_t(i)di = (1-u)X^E + uX^U$ be the total amount withdrawn by households from the mutual fund. The flow budget constraint of the mutual fund is given by:

$$X_t = (1 + \rho q_t^B)\frac{B_{t-1}^m}{\Pi_t} - q_t^B B_t^m + q_t^S \left(S_{t-1} - S_t\right) + S_{t-1} Div_t, \tag{7}$$

where $Div_t \equiv \int_0^1 Div_t(j)dj$ are aggregate dividends transferred from the firms to the fund, q_t^B is the price of government debt issued in period t, and q_t^S is the price of a firm equity share. The mutual fund allocates its budget over B_t^m and S_t, in order to maximize expected returns. This implies the following no-arbitrage relation:

$$\mathbb{E}_t \left[\frac{q_{t+1}^S + Div_{t+1}}{q_t^S}\right] = \mathbb{E}_t \left[\frac{\left(1 + \rho q_{t+1}^B\right)/\Pi_{t+1}}{q_t^B}\right]. \tag{8}$$

The aggregate volume of firm equity shares is normalized to $S_t = 1$. The realized rate of return of the mutual fund sector can then be expressed as:

$$r_t^A = \frac{(1 + \rho q_t^B) B_{t-1}^m / \Pi_t + q_t^S + Div_t}{q_{t-1}^B B_{t-1}^m + q_{t-1}^S} - 1. \tag{9}$$

Note that the mutual fund does not hold deposits on its balance sheets. In equilibrium, the return on deposits is dominated by the return on long-term government debt. The reason is that households value deposits for precautionary saving reasons, which drives down the real interest rate on deposits.

[14] Since we will linearize the model around a zero-inflation steady state, the precise specification of the stochastic discount factor is irrelevant for the results, as it drops out in the linearization.

Banks. There is a perfectly competitive banking sector. Banks can hold reserves at the central bank, denoted by M_t in real terms, which pay a nominal interest rate R_t, controlled by the central bank. In order to fund these assets, banks must create liabilities, i.e. deposits. No-arbitrage implies that reserves and deposits carry the same nominal interest rate R_t. In equilibrium banks therefore earn no profits. Consolidation of the balance sheet of the banking sector implies that:[15]

$$\int_0^1 D_t(i)di = M_t. \tag{10}$$

Treasury. Real government expenditures are exogenous and denoted by G_t. The treasury targets a constant real level of long-term debt, denoted by $B_t = B$, during each period. The budget constraint of the treasury is given by:

$$G_t = q_t^B B - (1 + \rho q_t^B)\frac{B}{\Pi_t} + T_t^{cb} + T_t, \tag{11}$$

where T_t^{cb} is a seigniorage transfer received from the central bank. Note that the lump-sum component of taxation (T_t) adjusts to balance the budget.

Central bank. The central bank targets the (gross) nominal interest rate on reserves (R_t) and the real amount of reserves (M_t), depending on the policy regime. The budget constraint of the central bank, in real terms, is given by:

$$T_t^{cb} + \frac{R_{t-1}}{\Pi_t}M_{t-1} + q_t^B B_t^{cb} = M_t + (1 + \rho q_t^B)\frac{B_{t-1}^{cb}}{\Pi_t}, \tag{12}$$

where B_t^{cb} denotes the central bank's holdings of long-term government debt.

We consider two versions of the model, each with a different conduct of monetary policy. In the first version, the central bank conducts conventional interest rate policy. In this case, the central bank targets the interest rate on reserves according to the following rule:

$$\widehat{R}_t = \widehat{R}_{t-1}^{\rho_R}\widehat{\Pi}_t^{\xi_\Pi^R}\widehat{Y}_t^{\xi_Y^R}z_t^R, \tag{13}$$

where hats denote variables relative to their steady-state values: $\widehat{Y}_t \equiv Y_t/\overline{Y}$, $\widehat{\Pi}_t \equiv \Pi_t/\overline{\Pi}$, and $\widehat{R}_t \equiv R_t/\overline{R}$ (note: $\overline{R}$, $\overline{\Pi}$, and $\overline{Y}$ are the steady-state values of R, Π, and Y, respectively). In the above policy rule, ξ_Π^R and ξ_Π^R are stabilization coefficients which determine the response of monetary

[15] It would also be straightforward to allow the banking sector to create additional deposits without holding reserves. However, this would not impact directly on our key mechanism, which requires QE to trigger the creation of *additional* deposits, as strongly suggested by Figure 1.

policy to fluctuations in output and inflation. ρ_R measures the persistence of interest rate policy changes, and z_t^R is an exogenous shock to the interest rate policy rule. Under conventional policy, the central bank does not own any government debt ($B_t^{cb} = 0$) and the real amount of reserves (and hence aggregate deposits) is held at a constant level ($M_t = \overline{M}$).[16]

In the second version of the model, the central bank conducts QE rather than interest rate policy. When QE is used, the nominal interest rate is pegged at $R_t = \overline{R}$, reflecting the reality that QE is typically used when the nominal interest rate cannot be moved. We further assume that if the central bank purchases government debt, it finances these purchases by issuing reserves:

$$q_t^B B_t^{cb} - (1 + \rho q_t^B) \frac{B_{t-1}^{cb}}{\Pi_t} = M_t - \frac{R_{t-1}}{\Pi_t} M_{t-1}. \tag{14}$$

In this case, QE targets the total amount of reserves according to the following rule:[17]

$$\widehat{M}_t = \widehat{\Pi}_t^{\xi_\Pi^{QE}} \widehat{Y}_t^{\xi_Y^{QE}} z_t^{QE}, \tag{15}$$

where $\widehat{M}_t = M_t / \overline{M}$ is the amount of real reserves relative to the steady-state level and z_t^{QE} is an exogenous shock to the QE rule, akin to conventional monetary policy shocks often considered in the NK literature. We will study this shock to better understand the workings of QE. In the above rule, ξ_Π^{QE} and ξ_Y^{QE} are policy coefficients which are, respectively, the elasticities of real reserves with respect to inflation and output.

An interesting special case of the QE rule sets both stabilization coefficients to zero, i.e. $\xi_\Pi^{QE} = \xi_Y^{QE} = 0$. In this case monetary policy directly targets a certain level of real reserves given by $M_t = \overline{M} z_t^{QE}$. We refer to this policy as *Real Reserve Targeting (RRT)*. This policy implies that, in the absence of QE shocks, the level of real reserves is constant and hence the nominal amount of reserves moves one for one with the price level; but unlike under conventional policy, B_t^{cb} is not constrained to be zero.

3.2 Equilibrium

The balance sheets of the various actors are summarized here (see also Table 3 in Appendix B):

- households hold mutual fund shares (A) and deposits (D), and the sum is household equity (note that some households may be borrowing if we allow $\underline{D} < 0$);

[16] We abstract from the zero lower bound on the net nominal interest rate ($R_t - 1$). However, we will assume that the interest rate is pegged at zero in the model version with QE. Regarding QE policy, we similarly do not impose a lower bound on B_t^{cb}, i.e. the central bank itself could in principle issue long-term debt.

[17] This rule can be reformulated as a rule for nominal reserves, being a function of the current and lagged price level, and nominal output.

- mutual funds hold firm shares (S) and long-term government bonds (B^m) and issue mutual fund shares (A);
- the fiscal authority has tax claims and issues long-term government debt ($B = B^m + B^{cb}$);
- the central bank may hold government debt ($B^{cb} > 0$) and issue real reserves M above the level $\bar{M}$;
- banks hold reserves (M) and issue deposits (D);
- firms earn profits and issue shares to mutual funds (S).

Given laws of motion for the exogenous states $\{\varepsilon_t,\ Z_t,\ G_t,\ z_t^{QE}\}$ and government policies $\{B,\ T_t,\ R_t,\ M_t,\ B_t^{cb},\ T_t^{cb}\}$, the competitive equilibrium is defined as a joint law of motion for households' choices $\{N_t(i), C_t(i), D_t(i), A_t(i)\}_{i\in[0,1]}$, mutual fund choices $\{B_t^m, S_t = 1\}$, aggregate quantities $\{Y_t, N_t, Div_t\}$ and prices $\{\Pi_t, w_t, q_t^B, q_t^S, r_t^A\}$, such that $\forall t$,

(i) Each household $i \in [0,1]$ maximizes (1) subject to the constraints (2), (3), and (4), with $N_t(i) = 0$ when he/she is unemployed;

(ii) Firms in total produce $Y_t = Z_t N_t$, pay out dividends according to (5), and set nominal prices such that the New Keynesian Phillips Curve (6) holds;

(iii) The mutual fund's budget constraint (7), and pricing conditions (8) and (9) hold, and the mutual fund's assets equal its liabilities:

$$\int_0^1 A_t(i)di = q_t^B B_t^m + q_t^S S_t.$$

(iv) The banks create deposits such that (10) holds;

(v) The treasury's and central bank's budget constraints, (11) and (12), hold;

(vi) The markets for deposits/reserves clear, i.e., equation (10) holds. Households' expectations about the distribution of assets are consistent with the actual distribution. Also, the markets for long-term government debt and labor clear, i.e.,[18]

$$B = B_t^{cb} + B_t^m; N_t = \int_0^1 N_t(i)di.$$

3.3 The transmission of QE in the model

We highlight a few properties of the model which help to understand the transmission of QE to the real economy, via asset markets. We also discuss an alternative way of interpreting the model,

[18]The goods market clearing is satisfied because of Walras law. To see this, we sum over all budget constraints from individual households, the mutual fund, the treasury, and the central bank; then, we use the balance sheet of the banks and the market clearing conditions for government debt and labor to reach $Y_t = \int_0^1 C_t(i)di + G_t + \phi\left(\Pi_t - 1\right)^2 Y_t$.

as well as a possible extension of the model, which connects our analysis to some other channels studied in the literature (though not all).

Let us first consider the effects of QE on asset markets, which depends on the response of mutual funds. Whenever a mutual fund receives deposits in exchange for long-term debt sold to the central bank, it will try to use these deposits to purchase other long-term debt. This response drives up the price of long-term debt, q_t^B, which implies a decline in the long-term interest rate and thus a stimulating effect. To see this more clearly, one can derive the following partial-equilibrium elasticity of q_t^B with respect to B_t^m:

$$\frac{dq_t^B/q_t^B}{dB_t^m/B_t^m} = \frac{1}{\rho - 1} < 0,$$

and note that B_t^m declines as the fund sells long-term debt to the central bank, so that q_t^B increases.[19]

Now, consider the effect of the change in asset markets on the real economy. To this end, it is useful to aggregate the budget constraint of the treasury and the households:

$$C_t + G_t + D_t = w_t N_t + X_t + q_t^B B - (1 + \rho q_t^B)\frac{B}{\Pi_t} + T_t^{cb} + \frac{R_{t-1}}{\Pi_t} D_{t-1}. \tag{16}$$

From this constraint it can be seen that an increase in the price of debt, q_t^B, increases the available liquid budget. Intuitively, the increase in q_t^B creates a redistribution of wealth from a sector which does not purchase physical goods (the mutual funds) to sectors that do purchase goods (the government and/or households). Keeping government expenditures (G_t) fixed, households can use the additional budget to purchase or to hold more liquid assets. The additional liquidity in turn increases their willingness to spend on goods. Thus, the increase in the price of long-term asset is a crucial component in the transmission of QE to aggregate demand.[20]

While the baseline assumptions on taxation and mutual fund withdrawals may appear stringent, results are in fact identical under a range of fiscal and financial market arrangements, including ones in which the withdrawal X_t varies endogenously over time. We show this formally in Appendix B.4. To understand this result, it is useful to consolidate the budget constraints of mutual funds (7), the treasury (11), and the central bank (12). This consolidation leads to:

$$X_t - T_t + G_t + \frac{R_{t-1}}{\Pi_t} M_{t-1} = Div_t + M_t.$$

From this equation it can be seen that QE has an impact on $X_t - T_t$, a liquid flow which enters directly into the households' budget constraints. However, households care only about the net flow,

[19] This partial-equilibrium elasticity is derived by taking a first-order approximation of Equation (7) around a steady state with zero inflation, keeping all variables other than q_t^B and B_t^m unchanged. In general equilibrium, the mutual fund's demand for long-term debt is also affected by changes in Π_t and Div_t.

[20] Alternatively, the government might increase expenditures, in which case QE directly boosts aggregate demand for goods. We later consider this possibility in the model.

as opposed to the financial flow X_t and the taxation flow T_t individually. Since T_t adjusts residually to satisfy the government budget constraint, different configurations for X_t yield identical results, as long as the gap $X_t^E - X_t^U$ remains constant (so that for everyone the change in withdrawal is exactly offset by lump-sum taxes/transfers).

The consolidated budget constraint (16) also allows for slight re-interpretation of the model. Once the households and the treasury are consolidated, it is *irrelevant* whether B represents household debt or treasury debt. The effects of QE are exactly the same, given a certain response of G_t to a QE intervention. Thus, while we have not explicitly modeled mortgage debt, the essence of the QE transmission channel does not depend on whether treasury debt or MBS are being purchased by the central bank. Related to this, the withdrawal X_t may be re-interpreted a change of household long-term debt position

Finally, note that the transmission of liquidity via an increase in the price of debt is accompanied by an increase in the value of the total stock of outstanding debt, qB. In Section 4.5, we discuss a version of the model in which the supply of B also moves. This is consistent with empirical evidence in Di Maggio et al. (2016), who show that Fed purchases of MBS led to an increase in the amount of mortgage debt extended to households. Alternatively, liquidity might be increased via a relaxation of borrowing constraints. Such constraints may move over time, for instance due to changes in house prices or via changes in regulatory policy, which we do not explore here.

4 The aggregate and redistributional effects of QE

We calibrate the model to the U.S. economy and set the length of a period to one quarter. We further normalize $\overline{z}^{QE} = \overline{Z} = 1$ and will discuss the calibration of $\overline{G}$ and $\overline{\varepsilon}$ below. The model is computationally tractable. We argue that this tractability preserves consistency with the micro data along key dimensions. In particular, there is a rich joint distribution of the two types of assets, which evolves endogenously over time.

4.1 Characterizing the Distribution of Wealth

The model is in principle computationally complex, as the economic state contains a time-varying joint distribution of liquid and partially liquid asset holdings. However, two insights allow us to reduce the complexity considerably, each of which may be of independent interest. To explain these, we anticipate a few elements of the calibration strategy.

The distribution of partially liquid wealth. The model is consistent with any initial distribution of partially liquid wealth, including the one observed in the data, given that withdrawals are fixed (conditional upon employment status). The aggregate amount of illiquid wealth is uniquely pinned down, however. In Appendix B.3, we describe in more detail the characteristics of the illiquid wealth distribution.

The distribution of liquid wealth. Having dealt with the distribution of $A_t(i)$, we now show how to keep track of the distribution of liquid wealth, $D_t(i)$, in a parsimonious way. To this end, we consider a calibration in which the amount of liquidity in the steady state is not too large. In that case, those who become unemployed exhaust their deposits within the first quarter of unemployment, and thus hit the liquidity constraint.

When calibrating the model to average deposit holdings in the US economy, this in fact turns out to be an outcome that is naturally obtained. Indeed, the deposit holdings of most US households do not exceed a few weeks of wage income, even among higher-income households.[21] Also, we will argue that this calibration strategy generates MPCs that are close to the data.

The fact that households hit the liquidity constraint upon job loss has some important consequences.[22] It implies that all employed households with the same employment duration behave identically, as do the newly unemployed with the same employment duration before job loss, and those who have been unemployed for more than one quarter.

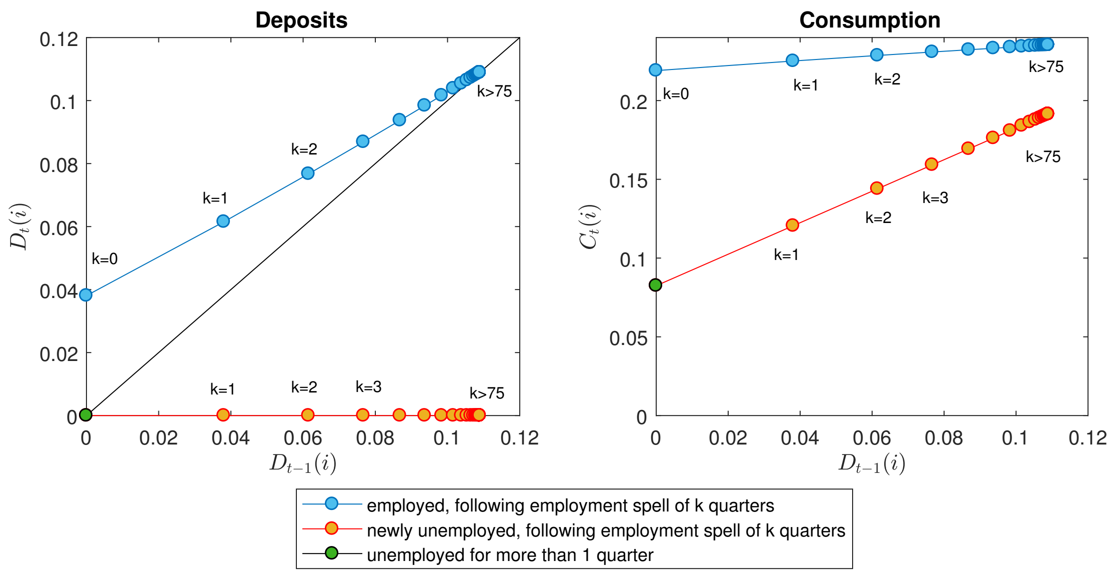
We exploit this outcome to solve the model as easily as a typical medium-scale DSGE model with a representative agent. In particular, we group agents who were employed in quarter $t-1$ into cohorts, indexed by the length of the employment spell in the previous quarter, denoted by $k \geq 0$. The cohort with $k = 0$ was previously unemployed, is currently employed, and enters the period with zero deposits. Hence they make identical decisions. Therefore, agents in cohort $k = 1$ all start with the same level of deposits, $D_{t-1}(i)$. Hence, conditional on their employment status, all agents within cohort $k = 1$ make the same decisions. Extending this logic, within any cohort $k \geq 1$ a fraction p^{EU} of the agents has become unemployed in the current quarter. They all behave identically and move out of the cohort in the next quarter. The remaining fraction of the cohort $1 - p^{EU}$ remains employed. Again, they all behave identically and move on to become cohort $k+1$ in the next quarter. Finally, turning to the households who were unemployed in quarter $t-1$, we note that all behave identically as they have depleted their deposits.

Figure 2 illustrates the steady-state choices of deposits and consumption of the different cohorts. Note that for larger values of the employment spell k, the levels of deposits and consumption converge. We use a total of K cohorts, and group all cohorts with $k \geq K$ into one bin. We thus need to

[21] In the data, a small fraction of households holds a large amount of deposits, for instance because they have set aside deposits in anticipation of a large upcoming durable purchase. Such a cash holding motive is outside the scope of our model. Moreover, Campbell and Hercowitz (2018) argue that even those households may have high MPCs out of liquid wealth, which is key for the mechanism in our model.

[22] This property is also exploited for tractability by Krusell, Mukoyama and Smith (2011) and Ravn and Sterk (2017), who assume zero aggregate liquidity, as well as by Challe and Ragot (2016) who assume that the employed are on a locally linear segment of the utility function. In our model, there is positive aggregate liquidity and a globally concave utility function.

Figure 2: Decision rules (steady state).



Notes: markers denote mass points of the liquid wealth distribution observed in the steady-state equilibrium. The black line is the 45-degree line.

keep track of K state variables characterizing the wealth distribution. In our quantitative exercises, we set $K = 75$. The precise value of the cutoff K is quantitatively irrelevant as long as it is not too small. To appreciate this point, note that from Figure 2 it can be seen that the behavior of cohorts beyond $k = 15$ is almost indistinguishable.

To solve the model, we apply a first-order perturbation method for dynamic analysis, using the popular `Dynare` software package. More details are provided in Appendix B.1.[23]

4.2 Calibration

Table 1 presents the parameter values. More details are also provided in Appendix C.1 and Appendix C.2. We assume the following utility function:

$$U(C, N) = \frac{C^{1-\sigma} - 1}{1 - \sigma} - \frac{\kappa_0}{1 + \kappa_1} N^{1+\kappa_1},$$

[23]We keep track of $K = 75$ state variables characterizing the wealth distribution. However, we obtained almost identical results with as little as $K = 20$. This is a much lower number than required by similar, perturbation-based solution methods. For example, the popular method of Reiter (2009) typically requires hundreds of state variables to obtain good accuracy. LeGrand and Ragot (2017) solve models by truncating idiosyncratic histories. In our application, even with a truncation cutoff lowered to $K = 20$, this would still imply 2^{20} state variables, i.e., more than a million.

where $\sigma > 0$ is the coefficient of relative risk aversion, which we set equal to one. Moreover, $\kappa_1 > 0$ is the inverse Frisch elasticity of labor supply, which is also set to one. Finally, $\kappa_0 > 0$ is a parameter scaling the disutility of labor, which we calibrate such that employed workers on average supply $\overline{N} = 1/3$ units of labor in the steady state.

We further calibrate the steady-state elasticity of substitution between goods as $\overline{\varepsilon} = 9$, which implies a steady-state markup of 12.5 percent, and $\beta = 0.99$, which corresponds to an annual subjective discount rate of 4 percent. We target an unemployment rate of $u = 0.045$ and an unemployment inflow rate of $p^{EU} = 0.044$, corresponding to a monthly inflow rate of about 1.5 percent, as measured in the Current Population Survey. The implied unemployment outflow rate is $p^{UE} = 0.934$. The unemployment benefit is targeted to be 25 percent of average wage income in the steady state, which implies that $\Theta^U = 0.25\frac{\overline{\varepsilon}-1}{\overline{\varepsilon}}\overline{N} = 0.074$.[24] The price adjustment cost parameter is set to $\phi = 47.1$, which corresponds to an average price duration of three quarters in the Calvo equivalent of the model.

To facilitate comparison of the two policies, we calibrate the model such that the steady states of the model version with QE and the version with conventional policy coincide. Specifically, we assume that in both cases the central bank targets zero inflation in the steady state, i.e. $\overline{\Pi} = 1$. The implied nominal steady-state interest rate is $\overline{R} = 1$.[25] We further set $\rho = 0.947$, which implies a duration of government debt of four years. The borrowing limit,$-\underline{D}$, is set to zero.[26]

The steady-state values of government expenditures, deposits, and government debt, i.e., $\overline{G}$, $\overline{D} = \overline{M}$, and B are chosen to hit the following targets. We target a ratio of government expenditures to output of 23 percent, in line with national accounts data, and a deposit-to-annual-output ratio of 7.5 percent, in line with data from the Flow Of Funds (FoF) accounts for households and non-profits. The real return on long-term government debt is targeted to be four percent annually. While not explicitly targeted, our model implies a ratio of the value of government debt to annual output, i.e. $\frac{\overline{q}\overline{B}}{4\overline{Y}}$ of 49.6 percent.

The parameters pinning down the illiquid wealth withdrawals for the employed and the unemployed, X^E and X^U, are calibrated as follows. We target the median amount of transaction accounts (i.e. deposits in the model) held by a household with median income, as a fraction of (pre-tax) median income. This ratio is about 26 percent in the SCF, averaged over the years 1989-2016, which is

[24] Statutory benefits are typically around 40 percent of labor income. However, the actual amount received by households is much lower due to limited eligibility and take-up. Chodorow-Reich and Karabarbounis (2016) argue that taking into account all these factors reduces the benefit to around 6 percent of income. Our calibration strikes a balance between their number and the statutory rate.

[25] Note that in the version with conventional policy, we abstract from the Zero Lower Bound on the nominal interest rate. In our comparison exercises, we thus ask whether QE is more or less effective than a hypothetical conventional policy that would not be subject to the ZLB. Alternatively, we could have calibrated the model version with conventional policy to be away from the ZLB, but this would make a clear comparison more difficult since the steady states of the two model versions would be different.

[26] We have solved a model with a positive borrowing limit. We obtained very similar results to our baseline, since we target the same steady-state real interest rate. Details of this version are available upon requests.

Table 1: Parameter values and steady-state targets.

Parameter	Description	Value	Notes
β	subjective discount factor	0.99	subjective annual discount rate: 4%
σ	coefficient of relative risk aversion	1	convention
κ_0	labor disutility parameter	11.4296	average labor supply employed: 1/3
κ_1	inverse Frisch elasticity	1	convention
p^{EU}	unemployment inflow rate	0.044	monthly rate: 1.5% (CPS)
p^{UE}	unemployment outflow rate	0.934	steady-state unemployment rate: 4.5%
Θ^U	unemployment benefit	0.0741	benefit 25% of avg. real wage
X^E	net mutual fund withdrawal: employed	0.0218	real interest rate: 0 %
X^U	net mutual fund withdrawal: unemployed	0.0852	median holdings liquid wealth (SCF), see text
B	government debt parameter	0.0399	long-term interest rate: 4%
$-\underline{D}$	borrowing limit	0	see text
$\bar{\varepsilon}$	elasticity of substitution varieties	9	markup: 12.5%
ϕ	price adjustment cost parameter	47.1	average price duration: 3 quarters
$\overline{G}$	real government expenditures	0.0732	expenditures-to-annual-output: 23%
ρ	decay government debt	0.9470	duration of government debt: 4 years
$\overline{D} = \overline{M}$	steady-state deposits (=reserves)	0.1009	deposits-to-annual-output (FoF): 7.5%
$\overline{\Pi}$	long-run inflation target	1	net inflation rate: 0%

slightly less than \$4,000.[27]

Finally, we assume that each of the stochastic driving forces $z \in \{\varepsilon, Z, G, z^{QE}\}$ follows an independent process of the form $\ln z_t = (1 - \lambda_z) \ln \overline{z} + \lambda_z \ln z_{t-1} + \nu_t^z$. Here, $\lambda_z \in [0, 1)$ is a persistence parameter and ν_t^z is an i.i.d. innovation, drawn from a Normal distribution with mean zero and a standard deviation given by $\sigma_z \geq 0$. We allow λ_z and σ_z to potentially differ across the four types of shocks, and we will discuss their values below.

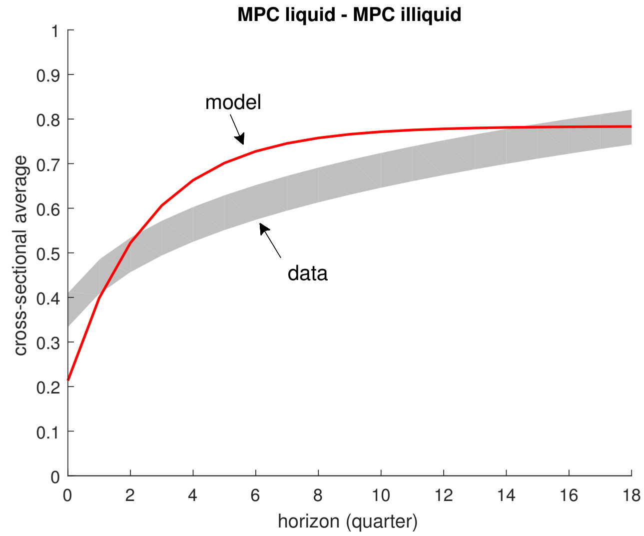
4.3 Model implications for micro-level consumption

We now explore the implications of the calibration for micro-level consumption behavior, and in particular for the gap in Marginal Propensities to Consume (MPCs) out of liquid and illiquid (partially liquid) wealth. This is important since the simple formula presented in Section 2 makes clear that this gap is a key determinant of the power of QE. We evaluate MPCs at different horizons, the importance of which has been emphasized by Auclert et al. (2018).

Figure 3 shows the average MPC gap, cumulated over time in both the model and the data, where the latter is computed based on estimates by Fagereng et al. (2018) and Di Maggio et al.

[27]Note that our strategy is conservative. Guerrieri and Lorenzoni (2017) use 2001 SCF and the median asset holdings is \$2,726. If we target the smaller amount of liquid assets, our tractability will be further strengthened, since unemployed agents are even more constrained.

Figure 3: Gap in Marginal Propensities to Consume out of liquid and illiquid wealth.



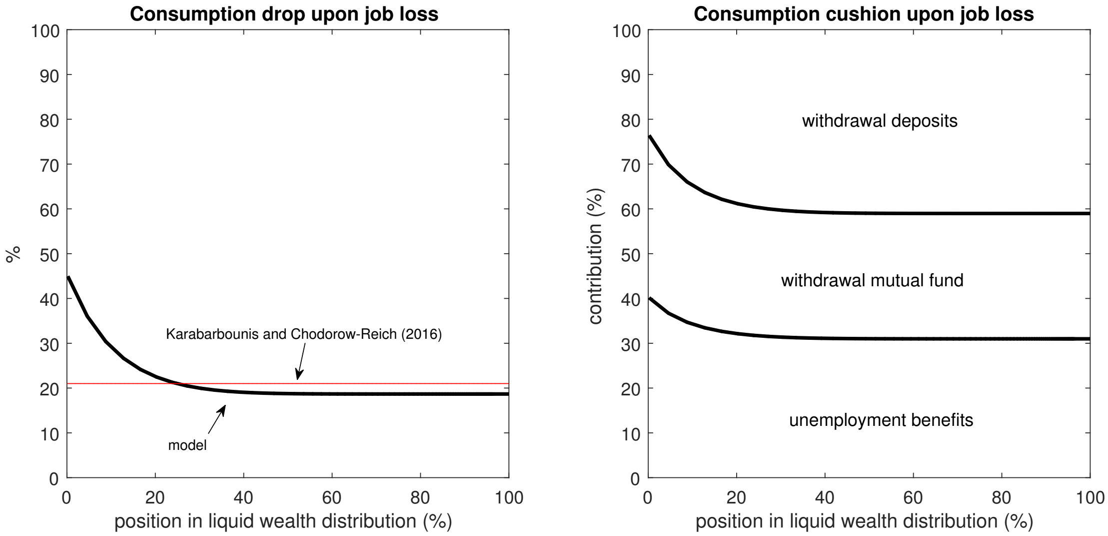
Notes: the figure shows the cumulative average MPCs across households. In the model, MPCs are computed as the response to a one-time surprise increase in additional wealth in deposits / mutual funds, keeping all prices constant. The range shown for "data" is computed as the dynamic MPC function for liquid assets provided by Fagereng et al (2018), minus a range of estimates for the MPC out of stock mutual funds in Table 3 of Di Maggio et al (2018). The latter is only provided at a one year horizon.

(2018). The gap in the model is large and increases over time. Because of labor supply effects, it does not converge to one. Relative to the data, the model initially somewhat undershoots, whereas for medium-run horizons there is some overshooting. At longer horizons, the MPC gap in the model and the data is very similar. Given that this MPC gap and its dynamic shape have not been targeted in the calibration procedure, we conclude that overall the model does a reasonable job in accounting for this key piece of empirical evidence.

One might also wonder about the ability of households to smooth consumption in the face of unemployment shocks. The left panel of Figure 4 plots the model-implied drop in consumption upon job loss, as a function of the household's position in the distribution of liquid wealth (deposits). The line is downward-sloping, as households with more liquid wealth are better able to cushion the consumption effect of becoming unemployed. The average consumption drop is 22 percent, which is very close to the empirical estimate of Chodorow-Reich and Karabarbounis (2016), who report a 21 percent drop based on data from the Consumer Expenditure Survey.

The right panel of Figure 4 shows the composition of the "consumption cushion" upon job loss. The cushion is defined as the difference between the drop in labor income and the drop in consumption upon job loss. Between 30 and 40 percent of the consumption cushion is financed by unemployment benefits, depending on the amount of liquid assets owned by the households. These benefits directly help households alleviate the fall in consumption. Around 30 percent of the consumption cushion is due to additional withdrawals from mutual funds after job loss. The

Figure 4: Consumption behavior upon job loss.



Notes: the black line in the left panel plots $100 \cdot [1 - C_t(i)/C_{t-1}(i)]$ for households who lost their jobs in the current quarter t. The right panel shows the contributions of the components of the "consumption cushion," for households who lost their job in the current quarter t. The consumption cushion is defined as $cush_t(i) \equiv w_{t-1}N_{t-1}(i) - [C_{t-1}(i) - C_t(i)]$. The contribution of unemployment benefit is computed as $(\Theta^U - \Theta^E)/cush_t(i)$, the contribution of liquidation of mutual funds as $\mu/cush_t(i)$, and the contribution of deposit withdrawal is computed as $[D_t(i) - D_{t-1}(i)]/cush_t(i)$. Both panels show outcomes in the deterministic steady state.

remainder of the cushion is due to the withdrawal of deposits.

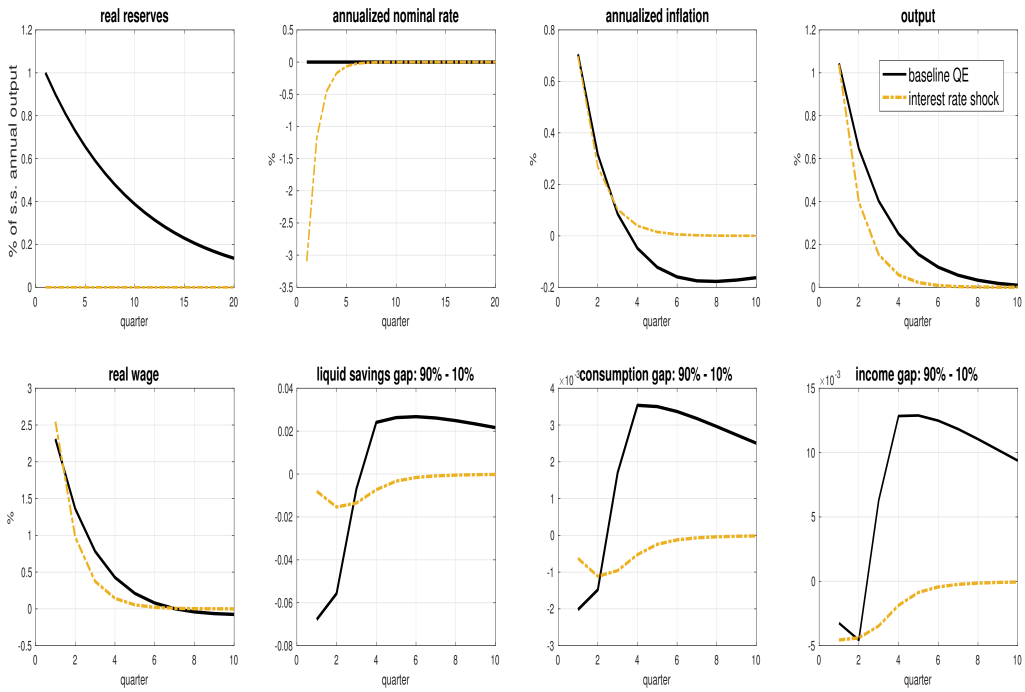
4.4 Equilibrium responses to a QE shock

Before applying the model to the Great Recession, we conduct a simple experiment which helps to understand how QE affects the macroeconomy. To this end, we consider an exogenous shock to QE, i.e. a positive innovation to z_t^{QE}. For transparency, we consider a version with Real Reserve Targeting (RRT, i.e. $\xi_\Pi^{QE} = \xi_Y^{QE} = 0$), so that there is no feedback from output and inflation to real reserves. The shock is scaled such that the purchase of long-term debt (and hence the increase in real reserves) is equivalent to one percent of *annual* steady-state output. We further assume a persistence coefficient of $\lambda_{z^{QE}} = 0.9$, which implies that the QE expansion has a half life of about 1.7 years.

As a comparison, we will also consider the effects of a "conventional" interest rate shock. In this version, we set $\xi_\Pi^R = 1.5$ and $\xi_Y^R = 0$, and otherwise identical parameter values, implying the same steady state, in order to facilitate comparison. In this latter version, we purposely abstract from liquidity effects, in order to maximally contrast the liquidity channel from the interest rate channel as mentioned in the introduction.[28] By doing so we highlight the value added of our analysis to the

[28]Realistically, conventional policy is also often implemented via Open Market Operations (OMO), i.e. purchases of short-term debt. Like QE, such interventions may create liquidity effects, which we abstract from here. The size of these effects would, however, depend on the extent to which short-term debt is illiquid. In our model, short-term debt

Figure 5: Expansionary QE and interest rate shock.



Notes: Responses plotted in deviations from the steady state. The shock is scaled such that real reserves increase by an amount equivalent to one percent of annual output on impact. The conventional monetary policy assumes a conventional Taylor coefficient $\xi_{\Pi}^{R} = 1.5$. The size (i.e., the standard deviation of one time shock to $\ln z_t^R$ is 1.8%) and persistence (i.e., $\rho_R = 0.9$) of the interest rate shock are scaled, so that the output and inflation response on impact are the same as in the case of QE. The savings/consumption/income gaps are measured by the difference of log of savings/consumption/income in the two groups.

New Keynesian literature, which typically abstracts from liquidity effects, placing the interest rate channel at the center of the analysis.

The black solid lines in Figure 5 plot the responses to the QE expansion in the baseline model. Immediately after the central bank starts purchasing government debt, output increases by 1.09 percent on impact and by 0.61 percent on average during the first year following the intervention.[29] Inflation also responds strongly to QE. One year after the intervention, the price level has increased by 1.16 percent. Real wages also increase substantially, reflecting the increase in labor demand which ensues from the increase in goods demand. As QE is rolled back, this increase dies out to almost zero after two years.

The responses of macroeconomic variables to a QE expansion are strong . To illustrate this point,

is fully liquid, and hence there would be no liquidity effects (in fact, with interest rates at zero, short-term debt and money (reserves) are fully equivalent, hence a swap of these assets will have no effects). Finally, note that the amount of assets purchases under traditional OMOs is typically much smaller than under QE, so to the extent that OMOs generate liquidity effects, these may be quantitatively less important.

[29] If we use the simple formula, the direct effect on output amounts to 0.63-0.095 = 0.54 percent.

Figure 5 also presents the responses to an interest rate shock in the "conventional policy" version of the model. Here, we scale the shock such that the output response is identical to the one obtained in the QE version of the model. Achieving this requires a very large reduction in the nominal interest rate, of about 300 basis points. Note also that, while the responses of output and the real wage are qualitatively similar in both versions of the model, inflation overshoots in the QE version, after about 4 quarters following the initial shock.

The bottom panels of Figure 5 show that QE also has non-trivial distributional consequences. In particular, the figure plots the 90-10 percentile range of the distributions of liquid savings, consumption and income. Following the QE expansion, all three measures of inequality initially decline, but quickly turn positive. Moreover, the effects are very persistent, decaying more slowly than the shock itself. Following an interest rate cut, the three measures also decline, which is consistent with empirical evidence in Coibion et al. (2017). The responses, however, are much smaller than those after a QE shock, even though the response of aggregate output is the same (by construction). Moreover, after an interest rate shock there is no subsequent reversal; by comparison, the responses of distributional variables to a QE expansion are both larger and switch signs after a few periods.

To better understand the aggregate effects of a QE shock, note that its key implication is that it increases the amount of liquid assets held by households. Since households have high marginal propensities to consume out of liquid wealth, aggregate demand increases. But given that households do not *directly* receive any of the additional liquidity from the central bank or the mutual funds, how are they able to finance both an increase in consumption and an increase in liquid asset holdings?

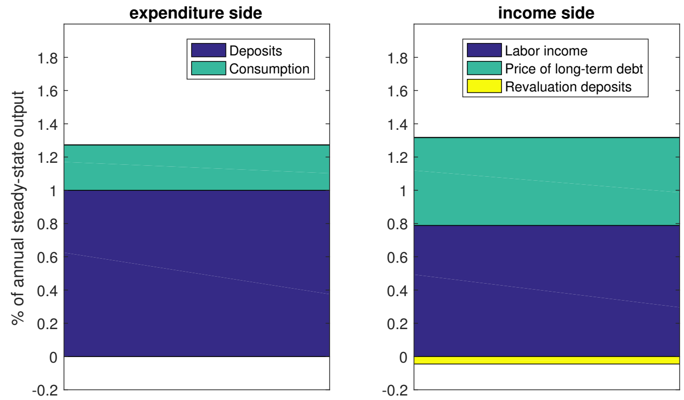
Figure 6 decomposes the change in the income and the expenditure side of Equation (16), the aggregated budget constraint of the households and the treasury (both of which contribute to aggregate demand), in the initial quarter following the shock. As anticipated, both consumption expenditures and deposit holdings increase (together by 1.3% of annual steady-state output).[30] Most of the change on the income side is due to the labor income component (about 62%), which increases by enough to finance a large chunk of the additional deposits. A significant part (about 39%) is due to the change in the price of long term debt, which is due to an increase in the price of new debt,[31] and to a downward revaluation of its stock of existing debt due to higher inflation. Finally, a downward revaluation of the households' deposits reduces their spending capacity, although this effect is small.

To understand the strong distributional effects of a QE shock, liquidity is also of key importance. The initial increase in liquidity is absorbed relatively evenly among the group of employed households, reducing inequality among households within this group, who all benefit from an in-

[30] Recall the nature of the experiment, all asset purchases happen in the initial quarter. Therefore, the initial increase in deposits is relatively large compared to the consumption response. After the initial period, however, the quantitative exit sets in and the deposit response takes the opposite sign, while the consumption response is still positive.

[31] On top of the previous comparison of QE and traditional OMO, since the price of long-term assets is more forward looking than that of short-term ones (i.e., $\rho = 0$) keeping everything else constant, the effect of QE is larger compared to traditional OMO.

Figure 6: Decomposition of the QE effects (response on impact).



Notes: The decomposition is based on the consolidated budget constraint of the households and the treasury, as in Equation (16). The component "Deposits" refers to D_t , "Consumption" is C_t, "Labor income" is $w_t N_t$, "Price of long-term debt" is $q_t^B B - (1 + \rho q_t^B)\frac{B}{\Pi_t}$ (where B is constant), and "Revaluation deposits" is $\frac{R_{t-1}}{\Pi_t} D_{t-1}$. The decomposition shown is the change in each of these components in the initial quarter following the shock. The components G_t and T_t^{cb} remain constant.

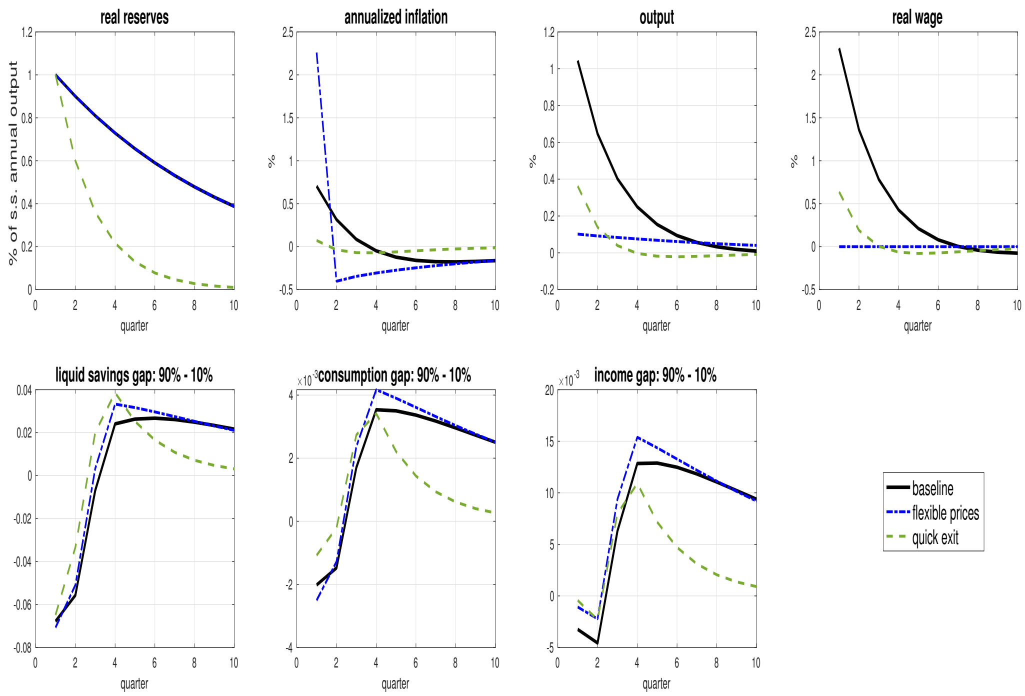
crease in income.[32] Subsequently, however, some employed households become unemployed and sell off their liquid assets to those who remain employed, who then become even richer. At the same time, previously unemployed households become employed. These households have not been able to benefit from the initial increase in income, and therefore did not accumulate more liquid assets. As a result, inequality in liquid assets increases, which feeds into consumption inequality and also income inequality, the latter via labor supply.

Finally, to understand the role of price stickiness and the speed of QE exit, we consider two alternative scenarios. First, we consider a version of the model with flexible prices (i.e. setting $\phi = 0$), illustrated by the blue dashed lines in Figure 7. In this case, the effect on output is much smaller, whereas there is a large spike in inflation on impact (2.5 percentage points). Intuitively, the increase in prices strongly dampens the increase in goods demand following the QE intervention. That is, indirect effects mostly offset the direct effects. Real wages remain constant under flexible prices. The fact that the QE shock still creates a small increase in output under flexible prices is associated with labor supply effects and re-distributions of wealth.

The green dashed lines in Figure 7 show the effects in the baseline model when the QE expansion is less persistent, setting $\lambda_{z^{QE}} = 0.6$, so that the exit is quicker. In that case, the initial expansion in output and inflation is much smaller. Intuitively, the contractionary effects associated with the quick unwinding of QE are immediately anticipated following the intervention, which dampens its effectiveness on impact. Thus, the overall power of a QE intervention depends crucially not only on the degree of price stickiness, but also on expectations regarding its persistence.

[32] Note that households at both the 90th and 10th percentiles of the three distributions are all employed.

Figure 7: Responses to an expansionary QE shock.



Notes: Responses plotted in deviations from the steady state. The shock is scaled such that real reserves increase by an amount equivalent to one percent of annual output on impact. The policy rule assumes $\xi_{\Pi}^{QE} = \xi_{Y}^{QE} = 0$ (Real Reserve Targeting). The baseline and flexible price responses assume a persistence coefficient of $\lambda_{zQE} = 0.9$, whereas the "quick exit" response assumes a persistence coefficient of $\lambda_{zQE} = 0.6$. The hours/savings/consumption/income gaps are measured as the differences of log of the 90th and the 10th percentiles of the consumption/income distributions.

4.5 Robustness

We now consider a number of alternative assumptions on mutual funds and policies.

Alternative assumptions on mutual funds. In the baseline model, households' mutual fund withdrawals are constant over time, and hence X_t does not respond to QE. In Appendix B.2, we show that this outcome can be microfounded by introducing a convex and pecuniary cost of mutual fund withdrawals, modeled as a tax.

In Section 3.3, we explained that QE policy has identical effects if one allows the aggregate withdrawal X_t to adjust, as long as the withdrawal net of tax $X_t - T_t$ is the same as in the baseline. Therefore, the result remains unchanged if we allow mutual funds to decide the payout policy (Appendix D.2). In Appendix D.1, we further consider an alternative version of the model in which the difference of individuals' mutual fund withdrawals do fluctuate over time. Specifically, we consider a version of the model in which the cost of mutual fund withdrawals is specified as a utility cost rather than a tax. This generates a richer distribution of mutual fund withdrawals, which is directly connected to the distribution of consumption. Individual mutual fund withdrawals then respond to QE, along with the distribution of consumption. We find that the responses of macroeconomic variables are also most similar to the baseline responses.

Alternative assumptions on fiscal policy. In the baseline model, government expenditures (G_t) do not respond to QE, while taxes (T_t) adjust to satisfy the government budget constraint. In Appendix D.3, we consider a model version in which instead T_t remains constant over time, while G_t adjusts to satisfy the government budget constraint. In this case, QE raises aggregate demand directly via government expenditures and we again find that QE has strong positive effects on output and inflation, as in the baseline model.[33] Finally, the consideration of varying government debt supply does not change the conclusion we had for QE, since households do not directly hold government debt.

Forward Guidance and Helicopter Drops. In Appendix E, we consider two other unconventional policy options. The first one is Forward Guidance, i.e. statements about future interest rate policy. We show that when the central bank also uses QE, there is no "Forward Guidance Puzzle", i.e. the policy only has small macroeconomic effects. We also consider Helicopter Drops, i.e. outright transfers to the households (via the fiscal authority), financed by the issuance of reserves. We find that the effects of Helicopter Drops on aggregate output are smaller but slightly more persistent than the effects of QE.

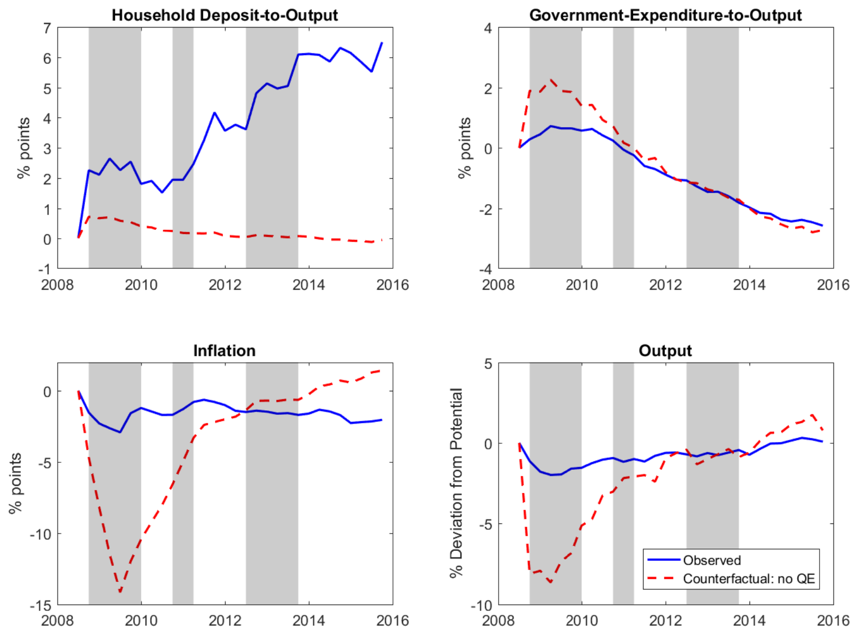
4.6 The macro effects of QE since the Great Recession

We now quantify the macro effects of QE on the U.S. economy since the Great Recession, when the nominal interest rate was (almost) at the zero lower bound, starting from 2008Q3. To this end, we structurally estimate the model, using data on the deviation of real output from its potential, the government-spending-to-output ratio, the deposits-to-output ratio, and *year-over-year* CPI inflation. To measure potential output, we use estimates from the Congressional Budget Office. The data, as differences relative to the 2008Q3 levels, are shown in Figure 8.

We estimate the version of the model with QE, and four shocks: cost push shocks, TFP shocks, government expenditure shocks, and QE shocks. We assume that all four shocks follow first-order autoregressive processes, as before, and we estimate the associated parameters. One might think of QE shocks as discretionary policy interventions. At the same time we also allow for systematic responses via the QE rule, and we estimate the stabilization coefficients on output and inflation.

[33] We finally consider a version in which all of X_t, T_t, and G_t are all kept at their steady-state levels, see the same appendix. In this version, QE generates similar responses of macro variables compared to the case where we allow G_t to vary. One can see from the consolidated budget constraint that QE now works purely through the adjustments of inflation and dividends. In this version, QE is essentially replacing long-term debt with short-term debt of the consolidated entity.

Figure 8: The impact of QE in the U.S. since the Great Recession.



Notes: data series and a counterfactual simulation without QE. For a description of the data series, see the main text and also Appendix A. In the counterfactual, we set $\widetilde{\xi}_Y^{QE} = \widetilde{\xi}_\Pi^{QE} = 0$ and shut down the (smoothed) QE shocks. Grey areas denote rounds of QE purchases by the Federal Reserve. Time series have been normalized around 2008Q3.

The remaining parameters are calibrated as described above. The model is estimated by Maximum Likelihood, using the Kalman filter combined with a perturbation method.

Table 4 in Appendix C.3 displays the estimated parameter values. The implied magnitude of QE shocks is substantial, with one standard deviation being 17% of the steady-state amount of reserves. At the same time, we also find a systematic component to the QE rule. In particular, we estimate the coefficient on inflation $\widetilde{\xi}_\Pi^{QE}$ to be significantly negative. When the annualized inflation falls by 1 percentage points, the central bank buys assets and creates reserves worth of 0.315 percent of annual GDP (roughly 0.51 trillions of US dollars in 2012). The point estimate for $\widetilde{\xi}_Y^{QE}$ is positive, though not as significant as $\widetilde{\xi}_\Pi^{QE}$. When output falls by 1%, the central bank creates reserves worth of 0.24 percent of annual GDP (roughly 0.39 trillions of US dollars in 2012). Thus, QE since the Great Recession appears best described as a mix of systematic and discretionary interventions.

With the estimated model at hand, we quantify the effects of active QE on the macro economy. We do so by simulating a counterfactual in which we both set $\widetilde{\xi}_Y^{QE} = \widetilde{\xi}_\Pi^{QE} = 0$ and shut down the (smoothed) QE shocks. We keep the rest of the smoothed shocks (i.e., best estimates for the shocks given the whole set of observations). In this case, real reserves and deposits remain fixed at their steady-state levels throughout the sample period.

Figure 8 shows the results of this counterfactual. The difference between the two lines in the upper left panel captures the Fed's asset purchases, which resulted in large-scale deposit creation. The lower right panel shows that QE had a large positive impact on aggregate output. Without active QE interventions, the recession would have been much deeper. For example, the fall in output relative to potential would have been about 7.5%, compared to about 1.7% in 2008Q4. Note that this effect is much larger than the direct effect computed in Section 2. Thus, in the initial years the direct effects of QE were amplified by general equilibrium effects.

However, after the recession the QE effects gradually fade out, and later on even switch sign, even if the policy itself had not been rolled back. Over longer horizons, general equilibrium effects turn from an amplifying into a dampening factor, as prices have had more time to adjust.[34] Indeed, the effects during the second and third rounds of QE are considerably smaller than that during the first round, as the later rounds had been anticipated by the private sector. The counterfactual analysis implies that inflation could have been higher than the data since the end of 2012. In 2015Q4, inflation would have been 3.5 percentage points higher if the central bank never purchased any assets.

5 An extension with capital

We have shown the effect of QE in a New-Keynesian model with heterogeneous agents but without capital. Now, we introduce physical capital and investment adjustment costs into the baseline model and explores robustness of the results. We also discuss the additional channels created by the introduction of capital.

We assume that physical capital is owned by the mutual fund and rented out to firms. Similar to long-term government bonds, capital is not directly traded among households. The production function of goods-producing firm j is now given by:

$$Y_t(j) = Z_t K_t(j)^\alpha N_t(j)^{1-\alpha}, \tag{17}$$

where $\alpha \in [0, 1]$. The aggregate capital stock evolves according to:

$$K_{t+1} = (1-\delta)K_t + I_t, \tag{18}$$

where $\delta \in (0, 1]$ is the depreciation rate and I_t is aggregate investment into physical capital.

There are also a large number of identical capital producers, who construct new capital using input of final output and subject to adjustment costs. They sell new capital to mutual funds at the price q_t^K. For simplicity, we assume that the mutual funds own capital producers, and the objective

[34] Figure 8 also shows a large positive effect of QE on inflation, although this effect was relatively short-lived and switched sign during 2013. The latter result reflects the overshooting of inflation also visible in Figure 7. In these responses, the effect on inflation dies out much faster than the effect on output.

of a capital producer is to choose investment I_t to maximize the expected present value of profits:[35]

$$\max \mathbb{E}_t \sum_{s=t}^{\infty} \beta^{s-t} \left\{ q_s^k I_s - \left[1 + \Omega \left(\frac{I_s}{I_{s-1}} \right) \right] I_s \right\}.$$

Here, $\Omega(I_t/I_{t-1})I_t$ is an investment adjustment cost function which satisfies $\Omega(1) = \Omega'(1) = 0$, so that the steady state is unaffected by the presence of the adjustment cost. Additionally, we assume that the adjustment cost function is convex: $\Omega''(x) > 0$. From profit maximization it follows that the price of capital goods equals the marginal cost of investment goods production:

$$q_t^k = 1 + \Omega \left(\frac{I_t}{I_{t-1}} \right) + \frac{I_t}{I_{t-1}} \Omega' \left(\frac{I_t}{I_{t-1}} \right) - \beta \mathbb{E}_t \left[\left(\frac{I_{t+1}}{I_t} \right)^2 \Omega' \left(\frac{I_{t+1}}{I_t} \right) \right]. \tag{19}$$

Profits or dividends from capital production $Div_t^k = q_s^k I_s - [1 + \Omega \left(\frac{I_s}{I_{s-1}} \right) I_s]$ are redistributed lump sum to mutual funds.

We now return to the goods producers. Let r_t^K denote the (net) rental rate of capital. The cost minimization problem of firm j, given a certain level of output $Y_t(j)$, is given by:

$$\min_{N_t(j), K_t(j)} w_t N_t(j) + r_t^K K_t(j)$$

$$\text{s.t. } Y_t(j) = Z_t K_t(j)^{\alpha} N_t(j)^{1-\alpha}.$$

Let $mc_t(j)$ be the Lagrange multiplier on the production function constraint. The decisions on hiring labor and capital lead to the expression for marginal cost, as well as the labor-capital ratio

$$mc_t(j) = \frac{1}{Z_t} \left(\frac{r_t^K}{\alpha} \right)^{\alpha} \left(\frac{w_t}{1-\alpha} \right)^{1-\alpha}. \tag{20}$$

$$\frac{r_t^K}{w_t} = \frac{\alpha}{1-\alpha} \frac{N_t(j)}{K_t(j)} \tag{21}$$

As before, firms maximize the expected present value of profits subject to the households' demand schedule and their production function. This delivers the same New Keynesian Phillips Curve in (6).

The budget constraint of the mutual fund now becomes:

$$X_t = (1 + \rho q_t^B) \frac{B_{t-1}^m}{\Pi_t} - q_t^B B_t^m + Div_t + Div_t^k - q_t^k I_t + r_t^K K_{t-1}. \tag{22}$$

The mutual fund maximizes expected returns, leading to the following additional no-arbitrage rela-

[35] Alternatively, one can assume that households own the capital producers. In that case, the discount factor used in the objective function is different. But we find no significant quantitative difference so that we use this simple setup.

tionship:

$$\mathbb{E}_t\left[\frac{(1+\rho q^B_{t+1})/\Pi_{t+1}}{q^B_t}\right] = \mathbb{E}_t\left[\frac{r^k_{t+1}+(1-\delta)q^k_{t+1}}{q^k_t}\right]. \tag{23}$$

As a summary, the model with capital adjusts some of the equilibrium conditions in the baseline case. We replace the production function, marginal cost expression, and MF budget constraint by (17), (20), and (22), respectively, and add equations (21), (19), and (23) and obtain three extra variables: r^K_t, q^k_t, and K_t , together with I_t from the capital evolution (18). The goods market clearing condition (after aggregating all budget constraints) is thus modified to

$$Y_t = C_t + \left[1+\Omega\left(\frac{I_t}{I_{t-1}}\right)\right] I_t + G_t + \phi\left(\Pi_t - 1\right)^2 Y_t.$$

Calibration. We choose a different level of steady-state aggregate productivity Z such that the aggregate output in this economy is the same as in the baseline economy. The depreciation rate of investment is set to $\delta = 0.025$, and $\alpha = 0.30$ so that investment-output ratio is around 19% (the average over the period between 1984Q1 and 2008Q2). The adjustment cost function is specified as

$$\Omega(x) = \frac{\zeta_0}{\zeta_1}(x-1)^{\zeta_1}$$

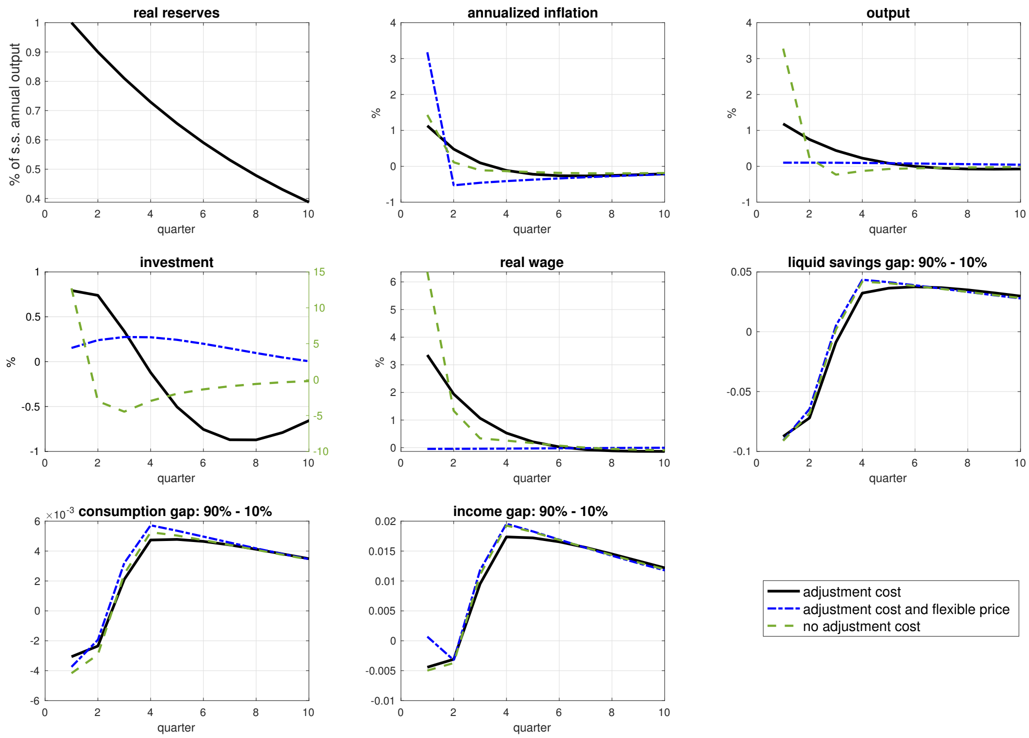
It is well known that the parameters ζ_0 and ζ_1 control the shape of the investment response to capital prices. We choose $\zeta_1 = 2$ so that inverse elasticity of net investment to the price of capital, captured by $(x-1)\Omega''(x)/\Omega'(x)$, is 1, in line with empirical evidence. Then, we compare two scenarios, one without adjustment costs ($\zeta_0 = 0$), and one with some adjustment costs ($\zeta_0 = 0.2$).

We keep the calibration targets the same as before, except that we also target the same wage as in the baseline, for the sake of comparability. To achieve this, we give up on the deposit-to-output ratio as a calibration restriction. In the calibrated model with capital, this ratio turns out to be 6.3%, which is slightly below the 7.5% in the baseline.

Responses to a QE shock, We feed the economy with capital the same expansionary QE policy shock as before. Figure 5 reveals that macro variables respond in a similar way to the version without capital when the economy features adjustment costs of capital ($\zeta_0 = 0.2$). When the economy has no adjustment cost on capital investment ($\zeta_0 = 0$), aggregate investment responds significantly to QE and falls immediately as the unwinding of QE starts. Output as a result follows a similar pattern. As is typically the case, dynamics appear more realistic with adjustment costs and therefore we will focus on this case from now on.

Following the expansion, output and inflation increase. Importantly, however, the output expansion is now driven not only by consumption but also investment. Investment increases despite the fact that the consumption boom triggered by QE has some crowding-out effect on investment.

Figure 9: Responses to an QE shock with capital.



Notes: Responses plotted in deviations from the steady state. The shock is scaled such that real reserves increase by an amount equivalent to one percent of annual output on impact. The policy rule assumes $\xi_{\Pi}^{QE} = \xi_{Y}^{QE} = 0$ (Real Reserve Targeting). The baseline and flexible price responses assume a persistence coefficient of $\lambda_{zQE} = 0.9$, whereas the "quick exit" response assumes a persistence coefficient of $\lambda_{zQE} = 0.6$. The hours/savings/consumption/income gaps are measured as the difference of log of the 90th and the 10th percentiles of the consumption/income distribution. The investment panel plots the case with no adjustment cost according to the right scale.

The investment expansion is due to both direct and indirect channels. The expansion of QE directly induces the mutual fund to replace government bonds sold to the central bank with new capital investment. This directly stimulates aggregate investment and output. Indirectly, the increase in aggregate goods demand triggers and increase in the demand for investment goods.

Note that as the central bank gradually unwinds QE following the initial period, i.e., selling government debt back to the market, investment contracts in the medium run (note: repeated QE shocks would delay this contraction.). Sims and Wu (2020) find a similar overshooting of investment because of the substitutions between consumption and investment, although in a different environment. When the exit is anticipated to happen quickly, the intervention has much smaller output effects, as in the version without capital. Moreover, under flexible prices a quick increase in prices mutes most of the output effects, similar to the version without capital.

Finally, the impact on inequality is also similar to that in the baseline, as the introduction of capital has only a limited impact on the dynamics of wages and inflation.

6 Concluding remarks

We found that QE can be an effective tool to stabilize the economy and that it has greatly dampened the Great Recession. However, it may not be desirable to replace a conventional interest rate policy with QE, as the latter tends to create strong side effects on inequality. In future work, it would be interesting to extend the model. For example, it would be straightforward to endogenize unemployment, which may strengthen the redistributive effects. It may also worth studying the money multiplier after QE in a setting with frictions in the banking sector.

We conclude with a note on conventional monetary policy for future research. We have assumed that the central bank directly controls the short-term interest rate. In practice, most central banks implement interest rate policy through open market operations. That is, they lower short-term interest rates by purchasing short-term T-bills. To the extent that deposits are more liquid than T-bills, the stimulative and redistributive effects of QE emphasized may apply to conventional policy as well.

References

Auclert, Adrien, "Monetary Policy and the Redistribution Channel," 2016. Working Paper.

— **, Matthew Rognlie, and Ludwig Straub**, "The Intertemporal New Keynesian Cross," 2018. Working Paper.

Barro, Robert J., "Are Government Bonds Net Wealth?," *Journal of Political Economy*, 1974, *82* (6), 1095–1117.

Batty, Michael, Jesse Bricker, Joseph Briggs, Elizabeth Holmquist, Susan McIntosh, Kevin Moore, Eric Nielsen, Sarah Reber, Molly Shatto, Kamila Sommer, Tom Sweeney, and Alice Henriques Volzg, "Introducing the Distributional Financial Accounts of the United States," 2015. Federal Reserve Board Discussion Paper.

Bhattarai, Saroj and Chris Neely, "A Survey of the Empirical Literature on U.S. Unconventional Monetary Policy," 2016. Federal Reserve Bank of St. Louis Working Paper 021-A.

Bilbiie, Florin O. and Xavier Ragot, "Optimal Monetary Policy and Liquidity with Heterogeneous Households," 2016. Working Paper.

Caballero, Ricardo J. and Emmanuel Farhi, "The Safety Trap," *Review of Economic Studies*, 2017, *85* (1), 223–274.

Campbell, Jeffrey R., "Quantitative Easing in Joseph's Egypt with Keynesian Producers," 2014. Chicago Fed Working Paper 2014-15.

— **and Zvi Hercowitz**, "Liquidity Constraints of the Middle Class," 2018. Working Paper.

Carpenter, Seth, Selva Demiralp, Jane Ihrig, and Elizabeth Klee, "Analyzing Federal Reserve asset purchases: From whom does the Fed buy?," *Journal of Banking and Finance*, 2015, *52*, 230–244.

Challe, Edouard and Xavier Ragot, "Precautionary Saving over the Business Cycle," *Economic Journal*, 2016, *126* (590).

Chen, Han, Vasco Curdia, and Andrea Ferrero, "The Macroeconomic Effects of Large-scale Asset Purchase Programmes," *Economic Journal*, 2012, *November*, 289–315.

Chodorow-Reich, Gabriel and Loukas Karabarbounis, "The Cyclicality of the Opportunity Cost of Unemployment," *Journal of Political Economy*, 2016, *124* (6), 1563–1618.

Coibion, Olivier, Yuriy Gorodnichenkoand LorenzKueng, and JohnSilvia, "Innocent Bystanders? Monetary policy and inequality?," *Journal of Monetary Economics*, 2017, *88*, 70–79.

Cui, Wei, "Monetary-fiscal Interactions with Endogenous Liquidity Frictions," *European Economic Review*, 2016, *87*, 1–25.

Debortoli, Davide and Jordi Galí, "Monetary Policy with Heterogeneous Agents: Insights from TANK models," 2017. Working Paper.

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

—, Marc Giannoni, and Christina Patterson, “The Forward Guidance Puzzle,” 2012. Federal Reserve Bank of New York Staff Reports 574.

Di Maggio, Marco, Amir Kermani, and Christopher Palmer, “How Quantitative Easing Works: Evidence on the Refinancing Channel,” *Review of Economic Studies*, 2016. forthcoming.

—, —, and Kaveh Majlesi, “Stock Market Returns and Consumption,” *Journal of Finance*, 2018. forthcoming.

Doepke, Matthias and Martin Schneider, “Inflation and the Redistribution of Nominal Wealth,” *Journal of Political Economy*, 2006, *114* (6), 1069–1097.

Fagereng, Andreas, Martin B. Holm, and Gisle J. Natvik, “MPC Heterogeneity and Household Balance Sheets,” 2018. Working Paper.

Galí, Jordi, *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*, Princeton University Press, 2008.

Gertler, Mark and Peter Karadi, “QE 1 vs. 2 vs. 3... A Framework for Analyzing Large Scale Asset Purchases as a Monetary Policy Tool,” *International Journal of Central Banking*, 2012, *9* (S1), 5–53. Working Paper.

Goldstein, Itay, Jonathan Witmer, and Jinh Yang, “Following the Money: Evidence for the Portfolio Balance Channel of Quantitative Easing,” 2015. Bank Of Canada Staff Working Paper.

Gornemann, Nils, Keith Kuester, and Makoto Nakajima, “Doves for the Rich, Hawks for the Poor? Distributional Consequences of Monetary Policy,” 2016. Working Paper 12-21, Federal Reserve Bank of Philadelphia.

Guerrieri, Veronica and Guido Lorenzoni, “Credit Crises, Precautionary Savings, and the Liquidity Trap,” *Quarterly Journal of Economics*, 2017, *132* (3), 1427–1467.

Hagedorn, Marcus, Jinfeng Luo, Kurt Mitman, and Iourii Manovskii, “Forward Guidance,” 2017. Working Paper.

Harrison, Richard, “Optimal Quantitative Easing,” 2017. Bank of England Working Paper 678.

Holmquist, Elizabeth, “Household and Nonprofit Balance Sheets in the Financial Accounts of the United States,” 2019. FEDS Notes 2019-01-04, Fed Board.

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante, “Monetary Policy According to HANK,” *American Economic Review*, 2017, *108* (3), 697–743.

Krishnamurthy, Arvind and Annette Vissing-Jorgenson, “The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy,” *Brookings Papers on Economic Activity*, 2011, pp. 215–265. Fall.

Krusell, Per, Toshihiko Mukoyama, and Anthony A. Smith, “Asset Pricing in a Huggett Economy,” *Journal of Economic Theory*, 2011, *3*, 812–844.

Lagos, Ricardo and Randall Wright, “A Unified Framework for Monetary Theory and Policy Analysis,” *Journal of Political Economy*, 2005, *113* (3), 463–484.

LeGrand, Francois and Xavier Ragot, “Optimal Fiscal Policy with Heterogeneous Agents and Aggregate Shocks,” 2017. Working Paper.

Luetticke, Ralph, “Transmission of Monetary Policy with Heterogeneity in Household Portfolios,” 2015. Working Paper.

McKay, Alisdair and Ricardo Reis, “The Role of Automatic Stabilizers in the U.S. Business Cycle,” *Econometrica*, 2016, *84* (1), 141–194.

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

Parker, Jonathan A, Nicholas S Souleles, David S Johnson, and Robert McClelland, “Consumer Spending and the Economic Stimulus Payments of 2008,” *American Economic Review*, 2013, *103* (6), 2530–53.

Ravn, Morten and Vincent Sterk, “Job uncertainty and deep Recessions,” *Journal of Monetary Economics*, 2017, *90*, 125–141.

Ravn, Morten O. and Vincent Sterk, “Macroeconomic Fluctuations with HANK and SAM: an Analytical Approach,” 2016. Working Paper.

Ray, Walker, “Monetary Policy and the Limits to Arbitrage: Insights from a New Keynesian Preferred Habitat Model,” 2018. Working Paper.

Reiter, Michael, “Solving Heterogeneous-agent Models by Projection and Perturbation,” *Journal of Economic Dynamics and Control*, 2009, *33* (33), 649–665.

Rocheteau, Guillaume, Pierre-Olivier Weill, and Tsz-Nga Wong, “An Heterogeneous-Agent New-Monetarist Model with an Application to Unemployment,” 2019. NBER Working Paper 25220.

Rotemberg, Julio J., “Sticky Prices in the United States,” *Journal of Political Economic*, 1982, *90* (6), 1187–1211.

Sims, Ecir and Cynthia Wu, *Journal of Monetary Economics*, 2020. forthcoming.

Sterk, Vincent and Silvana Tenreyro, “The Transmission of Monetary Policy through Redistributions and Durable Purchases,” *Journal of Monetary Economics*, 2018. forthcoming.

Wallace, Neil, “A Modigliani-Miller Theorem for Open-Market Operations,” *American Economic Review*, 1981, *71*, 267–274.

Weale, Martin and Tomasz Wieladek, “What are the Macroeconomic Effects of Asset Purchases?,” *Journal of Monetary Economics*, 2016, *79*, 70–93.

Wen, Yi, “When and How to Exit Quantitative Easing?,” *Federal Reserve Bank of St. Louis Review*, 2014, *96* (3), 243–265.

Woodford, Michael, “Fiscal Requirements for Price Stability,” *Journal of Money, Credit and Banking*, 2001, *33*, 669–728.

_, *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press, 2003.

_, “Methods of Policy Accommodation at the Interest-Rate Lower Bound,” in “The Changing Policy Landscape, Jackson Hole Conference,” Federal Reserve Bank of Kansas City, 2012.

Appendix

Supplemental material (online Appendix) for "Quantitative Easing with Heterogenous Agents" by Wei Cui and Vincent Sterk

A Data and additional empirical results

In this section of the appendix, we provide further empirical background to the transmission of QE. In A.1, we compare data from the Flow of Funds to those from the Survey of Consumer Finances (SCF). In A.2, we estimate the impact of QE announcements on liquidity, in particular deposits held by households. Finally, in A.3 we expand on the data used in the estimation of the model.

A.1 Data on household liquidity

Flow of Funds versus SCF. The data shown in Figure 1 in the main text are taken from the Flow of Funds (FoF) data. An important advantage of the Flow of Funds data is that they are constructed from administrative sources, that they aggregate up to the macro level and that their frequency is relatively high (quarterly). However, a possible concern regarding these data is that the sector "Households and Non-Profit Organizations" is constructed as a residual. For this reason, we draw a comparison to data from the Survey of Consumer Finances (SCF). A disadvantage of the SCF is the relatively low frequency (once every 3 years) and the limited number of survey participants. But the data do give a direct insight into households' finances.

Table 2: Checkable deposits and currency on the household balance sheet.

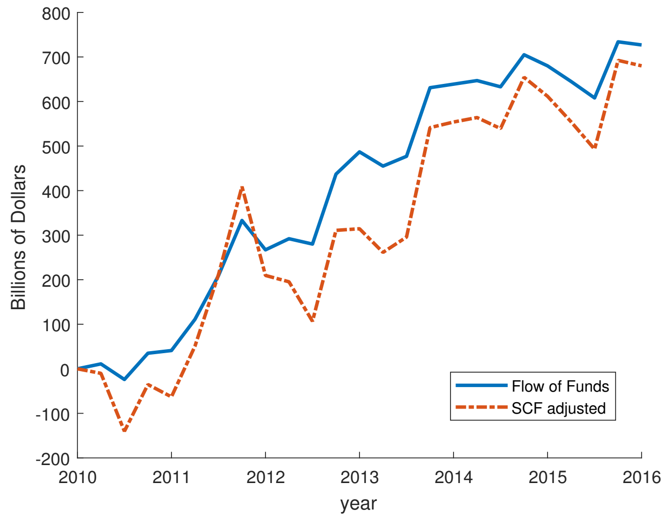
year	ratio SCF to FoF	year	ratio SCF to FoF
1989	0.70	*2007*	8.42
1992	0.49	*2010*	2.68
1995	0.54	*2013*	1.62
1998	0.93	*2016*	1.60
2001	1.51	*1989-2016*	2.06
2004	2.10		

Source: Batty et al. (2015), Table 1.

To compare household liquidity in the FoF to the SCF we draw on Batty et al. (2015), who document the ratio of households' checkable deposits and currency in the FoF relative to the SCF, see Table 2. The table shows a highly unusual spike in 2007. From 2010 onward, however, when most of the increase in household deposits in the FoF took place, the ratio is relatively stable. Based on the SCF data, we constructed an adjusted time series for household deposits. To do so, we multiplied the FoF series by the ratio shown in Table 2, linearly interpolated between SCF releases. Figure 10 shows the increase in the original flow of funds series, as well as in the SCF adjusted series. In both series, there is a similarly large increase in household deposits. We conclude that the increase in household deposits since QE as observed in the FoF is consistent with the SCF.

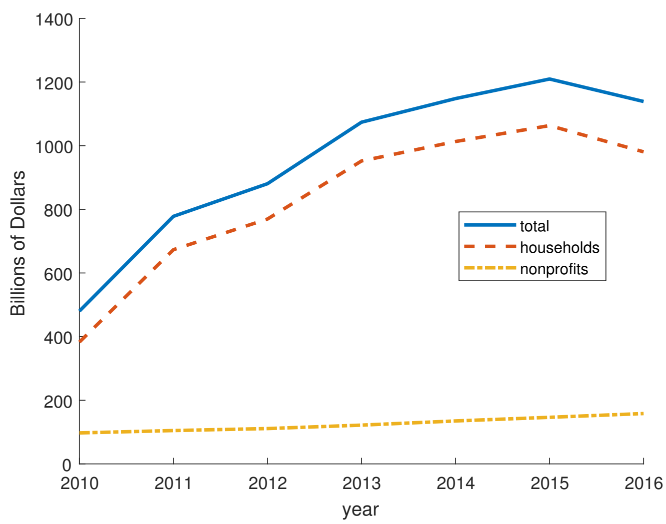
Nonprofits. Another potential concern with the SCF data is that it also combines households and non-profit organizations. Figure 11 shows data from Holmquist (2019), who splits up the total series into a household component and a nonprofit component. As can be seen from the figure, the relative contribution of non-profits has been small and fairly constant over time.

Figure 10: Household deposits: FoF versus SCF.



Source: Flow of Funds accounts, Survey of Consumer Finances, and Batty et al. (2015).

Figure 11: Checkable deposits and currency in the Flow of Funds (annual).



Source: Holmquist (2019).

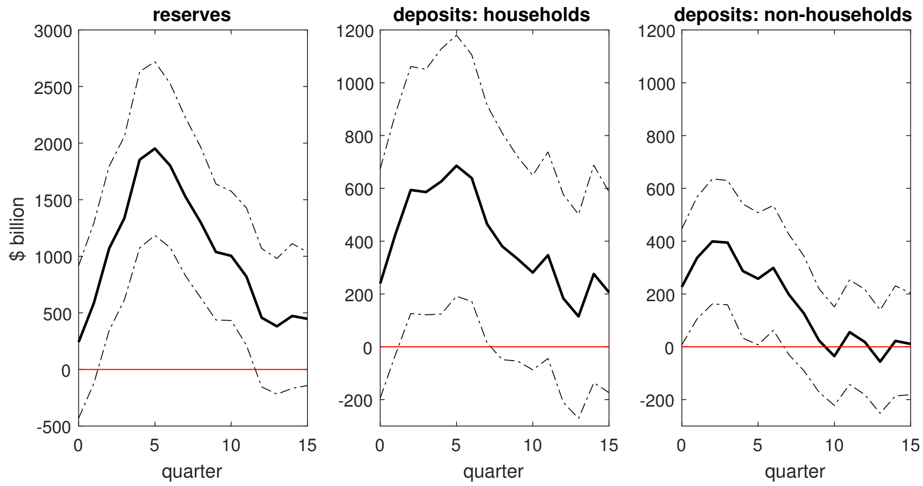
A.2 Liquidity following QE announcements

In the main text, we analyzed raw data on reserves and deposits during QE episode. We now estimate the effects of QE on reserves and deposits by running a local projection exercise. In particular, we estimate the following equation:

$$Y_t = \beta \mathbf{1}^{QE}_{t-k} + \gamma X_t + \varepsilon_t,$$

where Y_t is one of the three variables: reserves, deposits held by households, or deposits held by non-households. Moreover, $\mathbf{1}^{QE}_t$ is a dummy which equals one if there was a QE announcement (i.e., 2008q4, 2010q3 and 2012q3), and X_t is a matrix of control variables which includes unemployment rate, CPI inflation

Figure 12: Reserves and deposits responses following QE announcement.



Source: Federal Reserve Board, Flow of Funds accounts. In the top panel, grey areas denote rounds of QE purchases by the Federal Reserve. In the bottom panel, responses estimated using a local projection controlling for unemployment, inflation, GDP growth and the Federal Funds Rate. Dash-dotted lines denote 90% confidence bands.

and the GDP growth rate between quarter $t-4$ and t, the federal funds rate in quarter t, as well as a constant. Finally, ε_t is a residual. We run this regression separately for each lag period k. The sample runs from 1990q1 until 2018q4.

Figure 12 shows that QE announcements were followed by a surge in reserves and deposits. The latter ended up being held mostly by households (middle panel).[36] A smaller fraction was held by non-households (right panel), a category which consists mainly of non-financial firms.[37]

A.3 Data used in the calibration and estimation

For households' deposits data, we use "Households and Nonprofit Organizations; Checkable Deposits and Currency" from the U.S. Flow-of-Funds accounts. For consumption data, we use "Personal Consumption Expenditures" from U.S. BEA (Bureau of Economic Analysis) minus "Personal Consumption Expenditures: Durable Goods." For government expenditures data, we use "Government Consumption Expenditures and Gross Investment" from BEA. Output is defined as the sum of consumption and government expenditures. For inflation, we use the growth rate of "Consumer Price Index for All Urban Consumers". The above four series are obtained from 1985Q1 to 2018Q2, and we use the sample averages of each series to calibrate the model in the steady state.

For the estimation exercise, we only use the sub-sample period 2008Q3-2015Q4 because the nominal interest rate (i.e., the Fed Funds rate) is at (almost) zero during this period. We use the deposits-to-output ratio, the government-expenditures-to-output ratio, and inflation. For the output deviation, we do not use detrended output because the sample is too short. Instead, we first obtain the real potential GDP estimated by the U.S. Congressional Budget Office (CBO); the output deviation is then the difference between observed

[36]The fact that in the initial period there is uncertain in increase of QE, might be expected since we are looking at announcements of QE programs (which are not immediately implemented).

[37]In the model , there is no motive for firms to hold liquidity. However, the corporate finance literature has provided empirical evidence that increased liquidity in firms may stimulate investment and hiring. This would be another complementary transmission channel of QE which we do not consider in this paper.

real GDP in natural log terms and the CBO estimated potential real GDP in natural log terms. To simplify the estimation exercise, we normalize all variables in 2008Q3 to zero.

For the Survey of Consumer Finance (SCF), we use 2016 SCF Chartbook.[38] Specifically, we use "Median value of before-tax family income for families with holdings" Table on Page 7 and "Median value of transaction accounts for families with holdings" Table on Page 151.

B Tractability and computation

In this part, we explain a few technical details that enable the model to be tractable. Most of the discussion is focused on keeping track of the asset distribution. Before going to the detail, the following table summarizes the balance sheets of all agents in the economy discussed in the main text.

Table 3: Balance sheets.

households		**mutual funds**	
mutual fund shares (A)	household equity	firm shares (S)	mutual fund shares (A)
deposits (D)		lt govt. debt (B^m)	

fiscal authority		**central bank**	
claims to surpluses	lt govt. debt ($B = B^m + B^{cb}$)	lt govt. debt (B^{cb})	extra reserves ($M - \bar{M}$)
banks		**firms**	
reserves (M)	deposits (D)	claims to profits	firm shares (S)

Notes: left-hand sides denote assets, whereas right-hand sides denote liabilities/equities.

B.1 The distribution of liquid wealth and model solution

Let us now discuss in detail how we solve the model. In the baseline model, we only need to keep track of the distribution of liquid wealth (deposits) among households. In the presence of aggregate shocks, this distribution fluctuates over time, which is relevant to the state of the economy. In the calibrated model, it turns out that the liquid wealth distribution consists of only mass points. This happens as households who become unemployed spend all their liquid wealth in the initial quarter of unemployment, hitting the no-borrowing constraint within the first quarter of unemployment. Therefore, all the unemployed choose $D_t(i) = 0$.

It follows that any household which transitions from unemployment to employment holds exactly zero deposits. As a result, all employed households with the same employment duration behave identically (see also the discussion in Section 4.1). Moreover, all households which have been unemployed for more than one quarter consume simply their current net income, whereas the newly unemployed households consume their current income plus their liquid wealth (which in turn depends on their previous employment duration).

Let us introduce some notation indicating various "cohorts" of employed and unemployed households. Let a superscript E denote the employed, EU the newly unemployed, and UU those who have been unemployed for at least one quarter. Further, let k denote the employment duration of a household up until the current period (i.e. excluding the current period). For example $C_t^E(k)$ with $k = 0$ denotes the consumption level of a currently employed household who was unemployed in the previous quarter and $C_t^{EU}(k)$ with $k = 3$ denotes a newly unemployed household, who had completed an employment spell of 3 quarters upon job loss.

[38] https://www.federalreserve.gov/econres/files/BulletinCharts.pdf.

We can now characterize the household's choices with the following system of equations. For employed households we have the following equations:

$$C_t^E(k) + D_t^E(k) = w_t N_t^E(k) + \Theta_t^E + X_t^E - T_t, \quad k = 0, \tag{24}$$

$$C_t^E(k) + D_t^E(k) = w_t N_t^E(k) + \frac{R_{t-1}}{\Pi_t} D_{t-1}^E(k-1) + \Theta_t^E + X_t^E - T_t, \quad \forall k \geq 1, \tag{25}$$

$$\left[C_t^E(k)\right]^{-\sigma} = \beta \mathbb{E}_t \left[\frac{R_t}{\Pi_{t+1}} \left[\left(1 - p^{EU}\right) \left(C_{t+1}^E(k+1)\right)^{-\sigma} + p^{EU} \left[C_{t+1}^{EU}(k+1)\right]^{-\sigma}\right]\right], \quad \forall k \geq 0, \tag{26}$$

$$w_t \left[C_t^E(k)\right]^{-\sigma} = \kappa_0 \left[N_t^E(k)\right]^{\kappa_1}, \quad \forall k \geq 0. \tag{27}$$

For the newly unemployed households (EU) and the remaining unemployed households (UU) we have:

$$C_t^{EU}(k) + D_t^{EU}(k) = \frac{R_{t-1}}{\Pi_t} D_{t-1}^E(k-1) + \Theta_t^U + X_t^U - T_t, \quad \forall k \geq 1, \tag{28}$$

$$D_t^{EU}(k) = 0, \quad \forall k \geq 1, \tag{29}$$

$$C_t^{UU} + D_t^{UU} = \Theta_t^U + X_t^U - T_t, \tag{30}$$

$$D_t^{UU} = 0. \tag{31}$$

The above system contains three blocks of equations. Equations (24), (25), (28), and (30) are budget constraints. Moreover, (26), (29), and (31) characterize the optimal choices for deposits (using the fact that the unemployed are at the no-borrowing constraint, whereas the employed are on the Euler equation for deposits), and (27) is the first-order optimality condition for labor supply of the employed households.

In practice, we truncate the above system at a certain employment duration, i.e. we let $k = 0, 1, 2, 3.., K$, which renders the state-space finite dimensional. As can be seen from Figure 2, under our calibration, households converge fairly quickly to a maximum amount of assets. In our application, we set $K = 75$ and verify that results are insensitive to the truncation threshold. Setting the threshold as low as $K = 20$ delivers very similar results. We close the system by setting for the final cohort of employed households:

$$C_t^E(K) + D_t^E(K) = w_t N_t^E(K) + \frac{R_{t-1}}{\Pi_t} D_{t-1}^E(K) + \Theta_t^E + X_t^E - T_t,$$

$$\left[C_t^E(K)\right]^{-\sigma} = \beta \mathbb{E}_t \left[\frac{R_t}{\Pi_{t+1}} \left[\left(1 - p^{EU}\right) \left[C_{t+1}^E(K)\right]^{-\sigma} + p^{EU} \left[C_{t+1}^{EU}(K)\right]^{-\sigma}\right]\right],$$

$$w_t \left[C_t^E(K)\right]^{-\sigma} = \kappa_0 \left[N^E(K)\right]^{\kappa_1}.$$

These equations impose that beyond an employment duration of $k = 75$ quarters (i.e., more than 18 years), all households behave identically. This is not a very restrictive cutoff, since households already behave practically identically beyond an employment duration of 10 to 20 quarters, see Figure 2. We solve the above system jointly with the remaining model equations, given by:

$$
\begin{aligned}
1-\varepsilon_t+\varepsilon_t \frac{w_t}{Z_t} &= \phi\left(\Pi_t-1\right) \Pi_t-\phi \mathbb{E}_t\left[\beta \frac{Y_{t+1}}{Y_t}\left(\Pi_{t+1}-1\right) \Pi_{t+1}\right], \\
X_t &= Div_t+(1+\rho q_t^B) \frac{B_{t-1}^m}{\Pi_t}-q_t^B B_t^m, \\
X_t &= u X_t^U+(1-u) X_t^E, \\
X_t^U &= X^U \\
X_t^E &= X^E
\end{aligned}
$$

$$
\begin{aligned}
G_t &= q_t^B B-(1+\rho q_t^B) \frac{B}{\Pi_t}+T_t^{cb}+T_t+\Psi, \\
T_t^{cb}+\frac{R_{t-1}}{\Pi_t} M_{t-1}+q_t^B B_t^{cb} &= M_t+(1+\rho q_t^B) \frac{B_{t-1}^{cb}}{\Pi_t}, \\
B_t^{cb}+B_t^m &= B, \\
\sum_{k=0}^K \psi^E(k) D_t^E(k) &= M_t, \\
\sum_{k=0}^K \psi^E(k) N_t^E(k) &= N_t, \\
Y_t-G_t+\phi Y_t\left(\Pi_t-1\right)^2 &= \sum_{k=0}^K \psi^E(k) C_t^E(k)+\sum_{k=0}^K \psi^{EU}(k) D_t^{EU}(k)+\psi^{UU} C_t^{UU}, \\
Y_t &= Z_t N_t,
\end{aligned}
$$

in addition to the policy equations for either QE or conventional policy and the exogenous evolution of $(\varepsilon_t, Z_{t,}, G_t, z_t^{QE})$, as stated in the main text. In the above equations, $\psi^E(k)$, $\psi^{EU}(k)$, and ψ^{UU} are population share parameters, which satisfy $\psi^{UU}=u(1-p^{UE})-p^{EU}(1-u)$, $\psi^E(k)=p^{UE}(1-p^{EU})^k u$ for $k<K$, $\psi^E(K)=1-\sum_{k=0}^{K-1} \psi^E(k)-u$, $\psi^{EU}(k)=p^{UE}(1-p^{EU})^k p^{EU} u$ for $k<K$, and $\psi^{EU}(K)=p^{EU} \psi^E(K)$. For equilibrium dynamics, we use a first-order perturbation method to solve for the joint system.

Finally, we discuss how to verify easily that in equilibrium the unemployed hit the no-borrowing constraint in deposits. This is the case if:

$$
\left[C_t^{EU}(K)\right]^{-\sigma}>\beta \mathbb{E}_t\left[\frac{R_t}{\Pi_{t+1}}\left[p^{UE}(C_{t+1}^E(0))^{-\sigma}+\left(1-p^{UE}\right)(C_{t+1}^{UU})^{-\sigma}\right]\right].
$$

This equation implies that the newly unemployed with the longest previous employment spell do not want to save, i.e., they are at the constraint. If this condition holds, then the same is true for all the other unemployed households, since these are less wealthy, which implies that $C_t^{UU} \leq C_t^{EU}(K)$ and $C_t^{EU}(k) \leq C_t^{EU}(K)$. See also Figure 2 for an illustration of this point. We verify that the above equation holds in the steady state.[39]

B.2 Mutual fund withdrawals with adjustment costs

In the baseline model, withdraws are exogenous to households. We show how to micro-found this with adjustment costs. We can add an adjustment cost $\Psi_t(i)$ to the budget constraint of the household ((4)). The

[39]Under a local perturbation, the constraint then also holds outside the steady state.

adjustment cost is given by $\Psi_t(i) = \omega_t(i)\Psi(X_t(i))$, where $\Psi(\cdot) \geq 0$ is a convex function with $\Psi(0) = 0$; $\omega_t(i) = 1$ for the employed and $\omega_t(i) = \omega^U \in [0, 1)$ for the unemployed. The idea behind the latter is that the unemployed might face a less steep tax schedule than the employed. In order to preserve the equilibrium conditions in B1, we treat the withdrawal costs as taxes[40] so that the goods-market clearing condition remains the same.

We now discuss in more detail the household's choices regarding partially liquid wealth stored in mutual funds. The first-order optimality condition for the household's decision for the withdrawal $X_t(i)$ can be written as:

$$U_{C,t}(i) = \omega_t(i)\Psi'(X_t(i))U_{C,t}(i) + \lambda_t(i)$$

where $U_{C,t}(i)$ is the marginal utility of consumption and $\lambda_t(i) \geq 0$ is the shadow value of mutual fund wealth, i.e. the Lagrange multiplier on the evolution of $A_t(i)$ in (3). As discussed in the main text, there is no lower bound on partially liquid wealth since the natural limit in the model implies that $A_t(i) \geq -\infty$. This in turn implies that the shadow value of illiquid wealth is zero, i.e. $\lambda_t(i) = 0$.[41] Therefore, households' decisions become independent of the distribution of $A_t(i)$. Similarly, the mutual fund's choices do not depend on this distribution. We can therefore drop $A_t(i)$ as a state in the computation.

Specifically in our case, the first-order condition for $X_t(i)$ reduces to:

$$1 = \omega_t(i)\Psi'(X_t(i)).$$

We can now solve for the withdrawal $X_t(i)$ directly from the above equation. It follows that the employed all withdraw a constant amount $X_t(i) = X^E$, whereas the unemployed all withdraw $X_t(i) = X^U \leq X^E$. In the calibration, we treat X^E and X^U directly as parameters. Thereby we avoid having to make assumptions on the precise functional form of adjustment cost function. The first-order condition for $A_t(i)$ can be expressed as

$$U_{C,t}(i)\lambda_t(i) = \beta\mathbb{E}_t\left[U_{C,t+1}(1 + r^A_{t+1})\lambda_{t+1}(i)\right].$$

Given $\lambda_t(i) = 1 - \Psi'_t(X_t(i)) = 0$ at any time t, the left-and the right-hand side collapse to zero. The expected return $\mathbb{E}_t r^A_{t+1}$ is then determined purely via the mutual funds.

Remark: The two assumptions, the lack of a borrowing limit on partially liquid wealth and the assumption that the withdrawal cost function only depends on the withdrawal amount X, together imply that $X_t(i)$ are constants given the households' employment status. Violating either assumption will generate a time-varying distribution of $X_t(i)$ that we have to keep track of. The first assumption is not strong, as it reduces the extent of financial frictions relative to a model with a lower limit on $A_t(i)$. Moreover, as explained in Appendix B.3, the model is consistent with any initial distribution of $A_t(i)$, including ones with bounded support. Over time, unemployed households eat into their partially liquid wealth, but the speed at which they do so in the calibrated model is low, due to the adjustment cost. Therefore, lower bound of the support of distribution of $A_t(i)$ reduces only gradually. In finite time, the support of the distribution remains bounded as long as the initial distribution has bounded support. The second assumption is relaxed in Appendix D.

Remark: This micro-foundation implies that the system of equations that governs the dynamics of the economy is the same except that we should add the total adjustment cost tax $\Psi_t = \int \Psi_t(i)di$ to the government budget constraint and replace X_t^E and X_t^U in the household budget constraints before in Section B1 by

[40]Realistically, many of the costs triggered by mutual fund withdrawals are associated with taxation. In particular, withdrawals may trigger capital gains taxes or –in case of retirement accounts– early withdrawal penalties. Such costs arguably do not reflect a loss of real resources. Also, back-end fees charged by mutual funds upon withdrawal might best be thought of as a transfer, since the transaction itself requires few resources. The main purpose of back-end fees is to reduce the likelihood of large and sudden net outflows from the fund, which tend to complicate its investment strategy.

[41]Theoretically, there are other solutions without $\lambda_t(i) = 0$ all the time. In our quantitative exercises, this does not happen. One can verify the transversality condition numerically, as the speed of $A_t(i)$ going to positive infinity or negative infinity is slow. See the discussion below.

$X_t^E - \omega^E \Psi(X_t^E)$ and $X_t^U - \omega^U \Psi(X_t^U)$.

B.3 The distribution of partially liquid wealth

We now discuss in more detail the properties of the distribution of $A_t(i)$. As noted in the main text, the average level of wealth is pinned down uniquely in the model as $A_t \equiv \int_0^1 A_t(i)di = q_t^B B_t^m + Div_t$. Given that the right-hand side variables are uniquely pinned down in the steady state, so is A_t.

Next, we note that the *distribution* of $A_t(i)$ is not uniquely pinned down in the steady state of the model. To see why, first note that the mutual fund's decisions do not depend on the distribution of $A_t(i)$. Second, recall that households withdraw X^E from the fund when employed and X^U when unemployed. Therefore, the decisions of households also do not depend on $A_t(i)$, as noted above. It follows that the steady state of the model is consistent with *any* distribution of partially liquid wealth, as long as its mean equals $A = q^B B^m + Div$.

At the same time, given an initial distribution for $A_t(i)$ in the initial period $t = 0$, the evolution of the distribution for any subsequent period $t = 1, 2, ...$ is pinned down uniquely. Since the return on mutual funds r_t^A responds to economic shocks, so will the distribution of $A_t(i)$. But even without aggregate shocks, the distribution will evolve, as it is non-stationary. To see why, consider a steady state with $r^A > 0$. The cross-sectional variance of illiquid wealth evolves as:

$$Var(A_t(i)) \geq (1 + r^A)^2 Var(A_{t-1}(i)) + Var(X_t(i)) > Var(A_{t-1}(i)),$$

where the first inequality follows from the fact that the covariance between $A_{t-1}(i)$ and $X_t(i)$ may be negative due to the persistence of employment and unemployment spells. It now follows that the wealth distribution is non-stationary and that its cross-sectional variance is ever increasing. As noted in the main text, this non-stationarity is due to the "saving by holding" property of the model, which is in line with cross-sectional evidence on saving rates and the fact that wealth inequality has been steadily increasing for decades.

B.4 Identical equilibria under different fiscal-financial arrangements: proof

Here, we prove the claim in the main text that there is a range of financial market arrangements and fiscal arrangements which generate identical effects of QE. That is, keeping the path of fiscal spending $\{G_t\}$, monetary policy $\{R_t, M_t\}$, and the bond at hands of mutual funds $\{B_t^m = B_t - B_t^{cb}\}$ the same, the claim is that the equilibrium allocations are identical to variations in the laws of motions for X_t, T_t, and B_t, provided that the gap $X_t^E - X_t^U$ is constant.

First, let us denote a different combination of withdrawals and lump-sum taxes (or transfers) as $\tilde{X}_t(i)$ and $\tilde{T}_t(i)$ in a new economy, while $X_t(i)$ and T_t are used for an original economy. Since $X_t^E - X_t^U = \tilde{X}_t^E - \tilde{X}_t^U$, every individual i's withdrawal is increased by the same amount of ΔX_t. Let us postulate that the tax policy $\tilde{T}_t$ as

$$\tilde{T}_t = \tilde{X}_t(i) - X_t(i) - T_t = \Delta X_t - T_t$$

and we will prove that $\tilde{T}_t$ holds in equilibrium. It is straightforward to have

$$X_t(i) - T_t = \tilde{X}_t(i) - \tilde{T}_t.$$

Second, we can define withdrawal net of tax as $X_t^T(i) \equiv X_t(i) - T_t$ which is exogenous to the household. The budget constraint then becomes

$$C_t(i) + D_t(i) = w_t N_t(i) + \frac{R_{t-1}}{\Pi_t} D_{t-1}(i) + \Theta_t(i) + X_t^T(i)$$

Therefore, $X_t^T(i) = \tilde{X}_t^T(i)$ under the two arrangements. This means that households' budget constraint does not change (or, their choices stay the same if prices stay the same).

Third, it only remains to prove that $\tilde{T}_t$ satisfies the government budget constraint. Then, indeed all prices stay the same as in the original economy, and allocation in the new economy is the same as that in the original one (since the equilibrium in the original economy is unique according to standard proofs of incomplete market models). Instead of looking at the government budget constraint directly (see the Remark), we can look at the consolidated budget constraint of the mutual funds, the fiscal authority, and the monetary authority in the new economy:

$$\tilde{X}_t - \tilde{T}_t + G_t + \frac{R_{t-1}}{\Pi_t} M_{t-1} = Div_t + M_t.$$

After plugging in the expression for $\tilde{T}_t$, we have

$$X_t - T_t + G_t + \frac{R_{t-1}}{\Pi_t} M_{t-1} = Div_t + M_t,$$

which has to be true because this is the consolidated budget constraint in the original economy. That is, treating $X_t^T(i) = X_t(i) - T_t$ (and thus $X_t - T_t$) as one variable, the constraints that matter for the economy (i.e., household constraints, the consolidated budget constraints, and finally the goods market clearing condition) stay the same, and the new economy has exactly the same allocation in the original economy.

Remark: Although allocations are invariant in the proof above, bond prices move, which also explains why taxes move. To see this, we can write the mutual funds' budget constraint in the new economy

$$\tilde{X}_t = Div_t + (1 + \rho \tilde{q}_t^B) \frac{B_{t-1}^m}{\Pi_t} - \tilde{q}_t^B B_t^m,$$

which means that a different financial market arrangement $\tilde{X}_t$ must imply a different outcome of bond price $\tilde{q}_t^B$. If QE under the alternative arrangement goes more through the channel of withdrawals, then $\tilde{q}_t^B$ has a smaller response to QE compared to q_t^B. Notice that because of the lump-sum rebate from the monetary authority to the fiscal authority, the consolidated budget constraint of the government (including the fiscal and the monetary authority) can be written as

$$G_t + (1 + \rho \tilde{q}_t^B) \frac{B_{t-1}^m}{\Pi_t} + \frac{R_{t-1}}{\Pi_t} M_{t-1} = \tilde{q}_t^B B_t^m + \tilde{T}_t + M_t,$$

where the left-hand side is the spending of the consolidated government, while the right-hand side is the total revenues raised from taxes, issuing reserves, and long-term government bonds held by mutual funds. We thus can see that $\tilde{T}_t$ we propose neutralizes the changes from X_t to $\tilde{X}_t$ since

$$\begin{aligned}
\tilde{T}_t + \tilde{q}_t^B B_t^m - (1 + \rho \tilde{q}_t^B) \frac{B_{t-1}^m}{\Pi_t} &= \tilde{T}_t - \tilde{X}_t + Div_t = \int (T_t - X_t(i))\, di + Div \\
&= \int \left(\tilde{T}_t - \tilde{X}_t(i)\right) di + Div_t \\
&= T_t - X_t + Div_t \\
&= T_t + q_t^B B_t^m - (1 + \rho q_t^B) \frac{B_{t-1}^m}{\Pi_t}
\end{aligned}$$

C Steady state, calibration, and estimation

Here, we show to solve for the steady-state economy in a systematic way, which is useful for the calibration exercise. The calibration strategy is shown after we discuss how to solve for the steady state efficiently.

A more "black-box" alternative is to solve the entire system of steady-state equations all at once using a numerical solution routine. The procedure below, however, makes it easier to hit certain calibration targets.

C.1 Solving for the steady-state equilibrium

We first show how to solve for the steady-state equilibrium, given q, w, Π, X^U, G, B, M, Θ^U, and Θ^E. We solve for the resulting tax policy T and interest rate policy R, together with the net outflow X^E. Suppose we have an initial guess of (T, R, X^E).

For the unemployed agents without any savings, the budget constraint implies:

$$C^{UU} = \Theta^U + X^U - T.$$

There are K cohorts of employed agents. The labor supply decision of the kth cohort satisfies

$$\frac{w}{C^E(k)} = \kappa_0 N^E(k),$$

with $\sigma = 1$, which means that the labor income is $wN^E(k) = \frac{w^2}{C^E(k)\kappa_0}$. To this end, we first solve the consumption and saving choice.

For the Kth cohort, the Euler equation for deposits is

$$\frac{1}{C^E(K)} = \beta R \left[p^{EU} \frac{1}{C^{UU} + D^E(K)R} + (1 - p^{EU}) \frac{1}{C^E(K)} \right],$$

and the budget constraint is

$$C^E(K) = \frac{w^2}{C^E(K)\kappa_0} + D^E(K)\,(R-1) + \tilde{\Theta}^E,$$

where $\tilde{\Theta}^E \equiv \Theta^E + X^E - T$. The above two equations pin down $C^E(K)$ and $D^E(K)$.[42]

Let us now guess $C^E(1)$. For the $i = 0$th cohort, the Euler equation and the budget constraint are

$$C^E(0) = \beta^{-1} R^{-1} \left[p^{EU} \frac{1}{C^{UU} + D^E(0)R} + (1 - p^{EU}) \frac{1}{C^E(1)} \right]^{-1}$$

$$C^E(0) = \frac{w^2}{C^E(0)\kappa_0} - D^E(0) + \tilde{\Theta}^E.$$

Since we know $C^E(1)$, the two equations solve the two unknowns $C^E(0)$ and $D^E(0)$.

[42] There is even an analytical solution as $C^E(K)$ can be solved from a quadratic equation. To see this, rearrange the Euler equation

$$\frac{1 - \beta R\left(1 - p^{EU}\right)}{C^E(K)} = \frac{\beta R p^{EU}}{C^{UU} + RD^E(K)} \rightarrow D^E(K) = \frac{\beta p^{EU}}{1 - \beta R(1 - p^{EU})} C^E(K) - \frac{C^{UU}}{R},$$

which can be used to express the budget constraint as a quadratic equation of $C^E(K)$:

$$\left[1 - \frac{(R-1)\beta p^{EU}}{1 - \beta R(1 - p^{EU})}\right] \left[C^E(K)\right]^2 + \left(\frac{R-1}{R} C^{UU} - \tilde{\Theta}^E\right) C^E(K) - \frac{w^2}{\kappa_0} = 0.$$

Since $\left[1 - \frac{(R-1)\beta p^{EU}}{1 - \beta R(1 - p^{EU})}\right] > 0$ and $-\frac{w^2}{\kappa_0} < 0$, we use the positive root for $C^E(K)$ (omitted).

For any $k = 1, 2, ..., K-1$ cohort we then obtain from the budget constraint

$$C^E(k) = \frac{w^2}{C^E(k)\kappa_0} + RD^E(k-1) - D^E(k) + \tilde{\Theta}^E,$$

or $D^E(k) = \frac{w^2}{C^E(i)\kappa_0} + RD^E(k-1) + \tilde{\Theta}^E - C^E(k)$. From the Euler equation, we obtain

$$\frac{1}{C^E(k)} = \beta R \left[p^{EU} \frac{1}{C^{UU} + D^E(k)R} + (1 - p^{EU}) \frac{1}{C^E(k+1)} \right],$$

or $C^E(k+1) = \left[\frac{1}{\beta R(1-p^{EU})C^E(k)} - \frac{p^{EU}}{(1-p^{EU})[C^{UU}+D^E(k)R]} \right]^{-1}$. Therefore, given $C^E(1)$, we can calculate $D^E(1)$, $C^E(2)$, $D^E(2)$, ..., $C^E(K-1)$ and $D^E(K-1)$. We look for $C^E(1)$ such that $D^E(K-1) = D^E(K) - \epsilon$ where ϵ is an arbitrary small and positive number (i.e., the amount of savings converge to the fixed point). This is effectively a shooting algorithm, which generates optimal consumption and saving choices. Aggregate labor supply is then given by

$$N = \sum_{k=0}^{K} \psi^E(k) N^E(k) = \sum_{k=0}^{K} \psi^E(k) \frac{w}{\kappa_0 C^E(k)}. \tag{32}$$

Now, we turn to the government side. Notice that the debt held by the central bank and the total debt held by the mutual fund are thus

$$B^{cb} = \frac{M(1-R)}{q^B - (1+\rho q)},$$
$$B^m = B - B^{cb}.$$

Finally, after obtaining all equilibrium objects with a given (T, R, X^E), we check the following three equations. We know that from the budget constraint of the mutual funds:

$$uX^E + (1-u)X^U = N - wN + (1 + \rho q - q)B^m. \tag{33}$$

The market clearing for reserves is given by:

$$M = \sum_{k=0}^{K} \psi^E(k) D^E(k). \tag{34}$$

The goods market clearing is given by:

$$\sum_{k=0}^{K} \psi^E(k) C^E(k) + \sum_{k=0}^{K} \psi^{EU}(k) D^{EU}(k) + \psi^{UU} C^{UU} + G = AN. \tag{35}$$

These three equations above solve the three unknowns (T, R, X^E). That is, if these three equations do not hold, we change our initial guess of (T, R, X^E) and iterate the computation.

C.2 Calibration

The above strategy for calculating the steady-state equilibrium objects will be used in the following calibration exercise.

The average labor supply is targeted (1/3 in our calibration), so N is known. Without loss of generality, we normalize the steady-state TFP to $Z = 1$, so that $Y = N$. Recall that the wage rate is $w = (\varepsilon - 1)/\varepsilon$, and we thus know the average labor income in the model. The unemployment benefit is calibrated to be a fraction (0.25 in our calibration) of average labor income, so Θ^U is known. Because of the budget-neutral unemployment insurance, Θ^E is also known. We target the return of long-term government debt which generates the steady-state q. We target the reserves-to-output ratio $M/4Y$ (or M/Y), the government-expenditure-to-output ratio G/Y, the real interest rate r, and the median deposits-to-income ratio. With these targets, we directly obtain G and M. The following discussion shows how we calibrate κ_0 and X^U .

First, we have an initial guess of (κ_0, X^U). The lump-sum tax T, or equivalently the level of government debt B, is set such that the model hits the median deposits-to-income ratio. To see this, notice that the consolidated fiscal and monetary budget constraint implies that

$$G + (R - 1)\, M + (1 + \rho q^B - q^B)B^m = T. \tag{36}$$

For any given T, we obtain $B^m = [T - (R - 1)M - G] / (1 + \rho q^B - q^B)$; using (33) gives X^E.

Second, we now have (T, R, X^E), and we can follow the strategy specified before to calculate other steady-state equilibrium objects. In addition, since the level of total debt satisfies $B = B^m + B^{cb}$, we can also view that B is calibrated to hit the median deposits-to-income ratio, where B is given by

$$B = B^m + B^{cb} = B^m + M\frac{(1 - R)}{q^B - (1 + \rho q^B)}.$$

Finally, to hit the steady-state labor supply N and the real interest rate $r = R/\overline{\Pi} = R$, the initial guessed values of κ_0 and X^U are adjusted such that the labor market clearing condition (32) and the reserve market clearing condition (34) are satisfied.[43] Notice that for a different (κ_0, X^U) one needs to re-calculate T, R, X^E, and therefore B.

Remark: If the model has adjustment costs, then the ratio of the total adjustment cost Ψ to the total withdrawals X will be used to target percentage adjustment cost (e.g., 10%). Notice that we do not need to specify the adjustment cost function because we can replace X^E and X^U by $\tilde{X}^E = X^E - \omega^E\Psi(X^E)$ and $X^U - \omega^E\Psi(X^E)$ and add total adjustment cost Ψ to the right-hand side of ((36)).

C.3 Estimation result

The estimation procedure used in the main text is standard. We use the (log) linear representation of the model around the deterministic steady state (which is determined by parameters discussed in the main text). Then, we express the model implied deviation of real output from its potential, the government-spending-to-output ratio, the deposits-to-output ratio, and *year-over-year* CPI inflation. We then write the model in Kalman filter form, and we use the observed data to estimate the parameters governing the exogenous processes and the realization of the innovations in the exogenous processes via Maximum Likelihood.

The estimated parameters are the following. The smoothed shocks (the best estimates for the shocks given the whole set of observations) are used to generate Figure 8.

D Robustness and extensions

In this part, we consider different assumptions on liquidation of mutual fund wealth, together with various assumptions on the behavior of fiscal policy.

[43]Equation (35) is not used here as we have used the consolidated fiscal and monetary budget constraint in calibration. The households' budget constraints and the consolidated government budget constraint imply the goods market clearing condition (35).

Table 4: Estimated parameter values.

Parameter	Description	value	std. error	t-statistic
λ_{ε}	persistence cost push shock	0.979	0.036	27.203
λ_A	persistence TFP shock	0.854	0.217	3.939
λ_G	persistence G shock	0.996	0.006	189.626
$\lambda_{z^{QE}}$	persistence QE shock	0.714	0.071	10.107
σ_{ε}	st.dev. cost push innovation	0.033	0.037	0.890
σ_A	st.dev. TFP innovation	0.009	0.001	6.433
σ_G	st.dev. G innovation	0.007	0.001	7.647
$\sigma_{z^{QE}}$	st.dev. QE innovation	0.170	0.038	4.508
$\widetilde{\xi}_Y^{QE}$	QE coef. output	0.240	0.090	2.662
$\widetilde{\xi}_{\Pi}^{QE}$	QE coef. inflation	-0.315	0.091	3.459

Notes: parameters have been estimated using Maximum Likelihood. See the main text and Appendix A for a description of the data series and the sample period.

D.1 Richer distribution of withdrawals.

We consider an alternative setup under which we obtain a richer distribution of withdrawals. This distribution connects to the distributions of consumption and liquid wealth, which move endogenously over time.

In order to achieve this, we specify the adjustment cost as a utility cost rather than as a tax in the budget constraint. The first-order condition would become (again using $\lambda_t(i) = 0$ according to AppendixB.2):

$$U_{C,t}(i) = \Psi'(X_t(i)).$$

In this case, there would be a richer distribution of withdrawals which is connected to the time-varying distribution of consumption. But still, withdrawals do not depend on $A_t(i)$, and hence tractability is preserved. Note that in this case, one does have to take a stand on the adjustment cost function, however.

To proceed, we specify the following adjustment cost function, this time specified in units of utility:

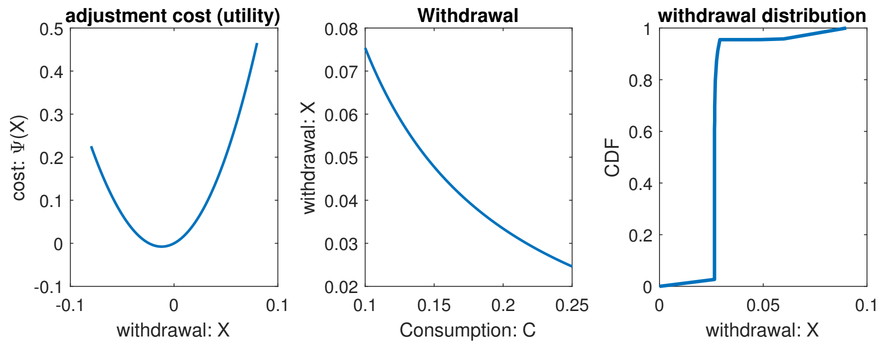
$$\Psi(X) = \gamma_0 + \gamma_1 X + \frac{(\gamma_2 + X)^{1+\gamma_3}}{1+\gamma_3},$$

and note that in this case the function is the same for the employed and the unemployed, as it is no longer a tax. We fix $\gamma_3 = 5$ (and experiment with different parameters) then re-calibrate the model parameters, targeting the same steady-state statistics as in the baseline. We obtain the following parameter values for the adjustment cost function: $\gamma_1 = -45.1752$ and $\gamma_2 = 2.1548$, which replaces the two calibrated withdrawal levels in the baseline calibration. We then set $\gamma_0 = -\frac{\gamma_2^{1+\gamma_3}}{1+\gamma_3}$ in order to normalize $\Psi(0) = 0$. In the steady state, mutual fund withdrawals are between 0.026 for the long-term employed (who have the highest consumption levels) and 0.09 for the long-term unemployed (who have the lowest consumption levels).

The left panel of Figure 13 plots the adjustment cost function, whereas the middle panel plots the relation between consumption and withdrawals. The right panel plots the distribution of withdrawals. There is now a distribution of withdrawals which is directly related to the distribution of consumption. As anticipated, households with higher levels of consumption withdraw less from the mutual fund. Importantly, withdrawals move over time, as consumption fluctuates.

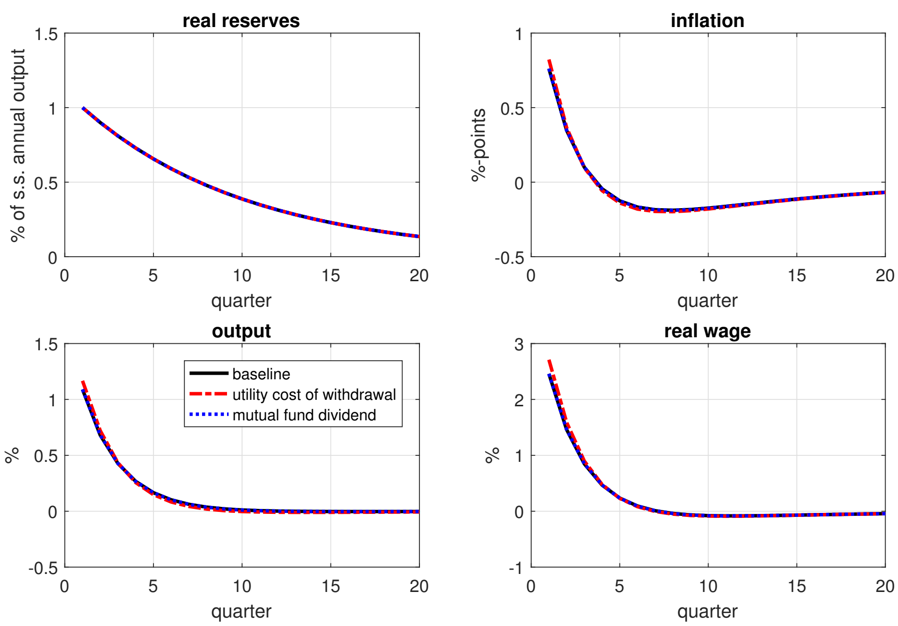
We now compare the macroeconomic effects of a QE shock in the model with the alternative adjustment cost to the responses in the baseline model. Figure 14 shows that responses are most similar in the two

Figure 13: Model with alternative adjustment cost: steady state.



versions of the model.

Figure 14: Responses to a QE shock: robustness w.r.t. mutual funds.



D.2 Mutual fund dividends

In the versions of the model discussed thus far, the mutual fund withdrawal, X_t, is either exogenous or chosen by the households. . We now consider another version of the model in which instead the mutual fund decides on X_t, now best thought of as a liquid dividend payout. Specifically, the manager of the representative mutual fund now decides actively on the payout policy X_t, while the adjustment cost for the households are fixed. For comparability, the gap between the withdrawals $X_t^U - X_t^E = \mu$ is kept the same as in the baseline, but now the fund's payout policy will affect the levels of X_t^U and X_t^E. With this alternative assumption, part of the liquidity created by QE can be *directly* transferred to households via mutual funds.

The mutual fund manager maximizes $\sum_{t=0}^{\infty} \beta^t X_t$ subject to a sequence of (7) by choosing X_t and B_t^m in each period. The first-order condition for government bond holdings and deposits imply that:

$$q_t^B = \beta \mathbb{E}_t \frac{1 + \rho q_{t+1}^B}{\Pi_{t+1}}. \tag{37}$$

Equation (37) implies that the manager will adjust the net outflow X_t freely, and QE policy is likely to induce her/him to increase the outflow from the fund directly, which pushes up aggregate consumption demand from households.

It turns out that the macroeconomic responses in this version are precisely the same as in the baseline, see Figure 14. It can be shown that in this version, X increases somewhat upon impact whereas T falls (not plotted). Importantly, the response of $X - T$ is precisely the same as in the baseline, generating the same reactions of inflation, output, and real wage. This results follows from the result proved in Appendix B.4

D.3 Fiscal policy

We now check for robustness with respect to assumptions made on fiscal policy, by comparing the following three versions of the model:

1. The baseline with a fixed mutual-fund payout X_t;
2. ("constant T, Flexible G"). A version with both a fixed mutual-fund payout and a fixed tax (i.e., T is kept the same as in the steady state). In this version, government expenditures (G) adjust to balance the government budget;
3. ("B rule"). The baseline, but with a rule for real government debt given by $B/\overline{B} = (Y/\overline{Y})^{\varsigma}$, where $\overline{B}$ and $\overline{Y}$ are the steady-state levels of respectively government debt and output. Taxes adjust to balance the government budget. Government expenditures are kept fixed. We set $\varsigma = -0.5$.

Figure 15 shows the responses to a QE shock. As a general observation, note that a positive QE shock stimulates aggregate output and inflation in all cases.

In the baseline, T falls on impact, followed by an increase. Intuitively, in the initial period the mutual funds try to replace debt sold to the Fed with new debt. This reduces borrowing costs for the government, allowing for lower taxes. In subsequent periods, QE is gradually reversed. This means that the demand for government debt by the mutual funds falls below steady state, which increases borrowing costs for the government. As a result, taxes must rise.

In the second version ("Constant T, flexible G"), we keep the taxes constant and let government expenditures adjust. There is a short-lived rise in G, followed by a contraction. The stimulative effect on output and inflation is greater than in the baseline on impact, but is shorter-lived. Finally, version 3 ("B rule") is the baseline but with a rule for government debt. Responses are similar to the baseline. The effect on output is somewhat more muted early on, but also more persistent.

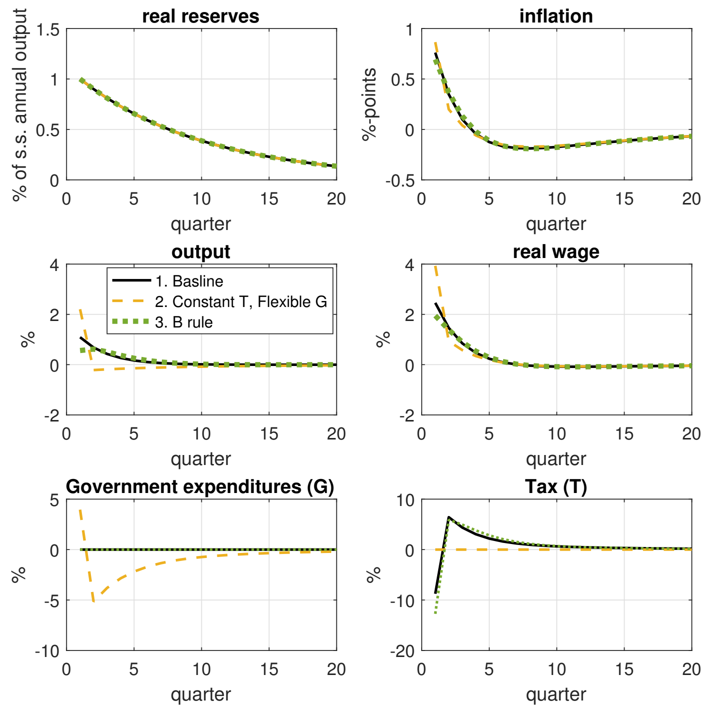
E Other unconventional policies

We consider two popular proposals of unconventional policies: Helicopter Drops and Forward Guidance.

E.1 Helicopter drops (HD)

In this part, we compare QE to an alternative policy measures, known as a helicopter drop. Helicopter money is a theoretical and unconventional monetary policy tool that central banks use to stimulate economies.

Figure 15: Responses to an expansionary QE shock: robustness w.r.t. fiscal policy.



Notes: see text and notes on Figure 7.

The term helicopter money is attributed to Milton Friedman, while former Federal Reserve Chairman Ben Bernanke later popularized the notion. Helicopter money involves the central bank supplying large amounts of money to the public, as if the money was being scattered from a helicopter.

A money-financed tax cut is essentially equivalent to Friedman's "helicopter drop" of money. Helicopter drop in our model is then an expansionary fiscal policy through a lower tax T, that is financed by an increase in reserves and thus deposits. We therefore impose the following helicopter policy restriction

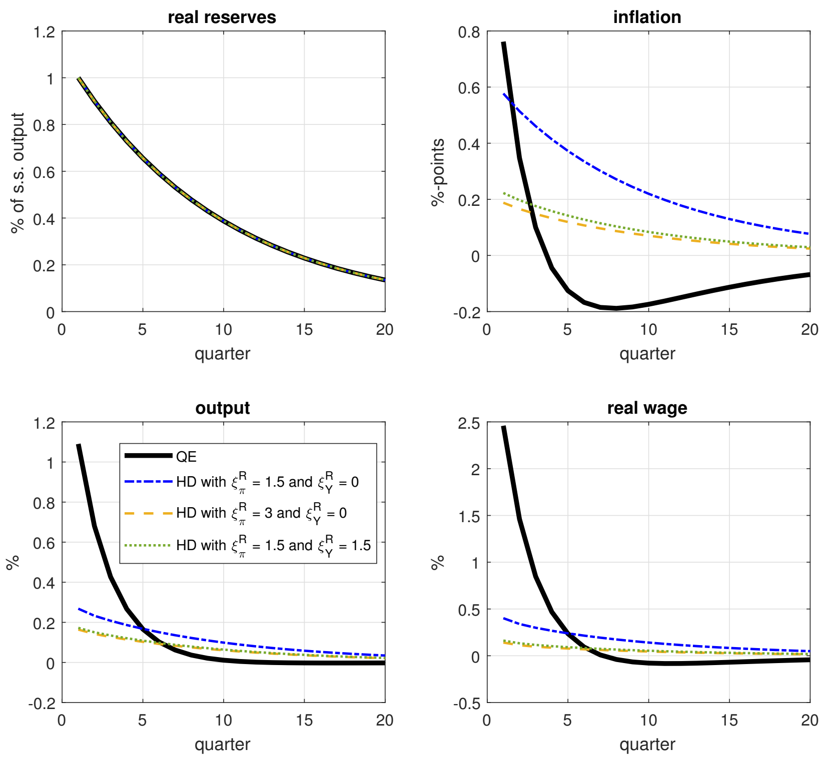
$$B_t^{cb} = 0.$$

Since the government targets a constant level of debt, the only way to finance the tax cut is through money/reserve finance in the initial period, besides the revaluation of government debt by variations in inflation, due to general equilibrium effects.

It turns out that the economic dynamics exhibit indeterminacy if we let the central bank target a certain interest rate rule. Therefore, we let the central bank actively set interest rates, i.e., ξ_Π^R and ξ_Y^R are not constrained at zero. One can view this type of policy as a conventional monetary policy without the constraint on the amount of reserves but with the constraint of the amount of government debt held by the central bank.

Figure 16 compares HD policy with QE. Both types have the same path for real reserves, but HD policy has a much smaller initial impact on the aggregate while generating only slightly more persistent effect. Interestingly, inflation never drops below zero, which means that the price level is permanently higher in the long run, and the nominal quantity of reserves/deposits is also permanently higher, reflecting the quantity theory of money by Friedman. To understand the smaller impact of HD, note that one can view reserves as another form of debt owed by the consolidated government of the treasury and the central bank. The tax cut

Figure 16: Responses to a Helicopter Drop.



Notes: responses to a Helicopter Drop shock compared to quantitative easing.

is initially funded by more reserves, which implies that a tax increase in the future is needed to pay back the debt. QE, on the other hand, substitutes reserves for long-term government debt, and does not directly imply a significant or any tax increase in the future (although there are tax implications via equilibrium effects on the government budget constraint).

Although more work on HD policy could certainly be done, our preliminary conclusion is that determinacy is far from guaranteed under this policy, and also that its aggregate impact might be relatively small, compared to QE.

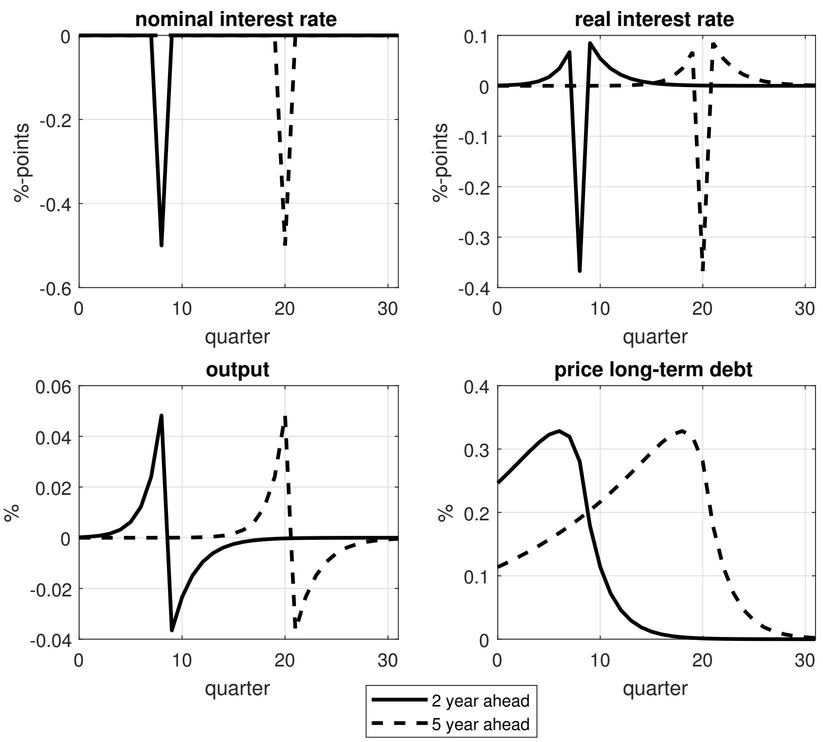
E.2 Forward guidance

Among unconventional policies, another alternative to QE is Forward Guidance: an announcement about monetary policy in the future. In the standard NK model without QE, Forward Guidance is an extremely effective policy once the zero lower bound on the nominal interest rate binds. In fact, macroeconomic responses to Forward Guidance turn out to be so enormous that they might call into question the basic tenets of the NK model, see Del Negro et al. (2012).

To address this puzzle, McKay et al. (2016) revisit the effects of Forward Guidance in an incomplete-markets NK model (without QE). They show that the output response to a five-year-ahead announcement is dampened substantially, relative to a representative-agent version of the model. Nonetheless, the effects of Forward Guidance remain large in comparison to empirical evidence, as presented for instance in Del Negro et al. (2012). Hagedorn, Luo, Mitman and Manovskii (2017) consider an incomplete-markets NK model with a target for nominal expenditure growth and show that the effects of Forward Guidance are much smaller.

We explore the effects of Forward Guidance in our model, with a QE policy on the part of the central bank. For transparency, we assume that a QE rule with Real Reserve Targeting is in place, i.e. we set $\xi_{\Pi}^{QE} = \xi_{Y}^{QE} = 0$. We then consider a pre-announced decline in the nominal interest rate of 50 basis points (corresponding to about 2 percentage points on an annualized basis) which lasts for one quarter. During all other periods, the net nominal interest rate remains fixed at zero. We consider a Forward Guidance announcement two years ahead, and another one five years ahead.

Figure 17: Responses to a Forward Guidance shock.



Notes: responses to a forward guidance shock, reducing the quarterly nominal interest rate by 50 basis points for one quarter, and announced 2 or 5 years ahead. Responses were computed in the model version with QE, setting $\xi_{\Pi}^{QE} = \xi_{Y}^{QE} = 0$, and letting the nominal interest rate R_t vary with the forward guidance shock, starting from $R_t = \overline{R}$.

Figure 17 shows the effects of the two Forward Guidance shocks. The figure shows that once the nominal interest rate is actually reduced, there is a strong decline in the real interest rate. During the quarters leading up to the implementation there is a small expansion in output, followed by a minor contraction after the implementation. Importantly, the output increase in the initial period of the announcement is almost negligible.[44] The impact response of the real interest rate (and hence inflation) is also extremely small. Moreover, the initial responses are declining in the announcement horizon. Finally, the lower right panel of Figure 17 shows that the Forward Guidance shock does have a substantial initial effect on the price of long-run treasury debt.

We thus conclude that once we account for incomplete markets *and* QE policy, the effects of forward guidance on output and inflation are no longer puzzlingly large. Rather, they are close to negligible. An implication of this finding is that, in comparison, QE stands out as the more effective stabilization policy, at least when the nominal interest rate is immutable in the short run.

[44]The output response is less than 0.0002 percent. Putting this number in perspective, McKay et al. (2016) report an initial output increase of 0.25 (0.1) percent under complete (incomplete) markets, in response to a forward guidance shock to the real interest rate of 50 basis points, 20 quarters ahead.