

Evaluating Central Banks' Tool Kit: Past, Present, and Future*

Eric Sims
Notre Dame and NBER

Jing Cynthia Wu
Notre Dame and NBER

First draft: February 28, 2019

Current draft: April 19, 2019

Abstract

We develop a structural DSGE model to systematically study the principal tools of unconventional monetary policy – quantitative easing (QE), forward guidance, and negative interest rate policy (NIRP) – as well as the interactions between them. To generate the same output response, the requisite NIRP and forward guidance interventions are twice as large as a conventional policy shock, which seems implausible in practice. In contrast, QE via an endogenous feedback rule can alleviate the constraints on conventional policy posed by the zero lower bound. Quantitatively, QE1-QE3 can account for two thirds of the observed decline in the “shadow” Federal Funds rate. In spite of its usefulness, QE does not come without cost. A large balance sheet has consequences for different normalization plans, the efficacy of NIRP, and the effective lower bound on the policy rate.

Keywords: zero lower bound, unconventional monetary policy, quantitative easing, negative interest rate policy, forward guidance, quantitative tightening, DSGE, Great Recession, effective lower bound

*We are grateful to Todd Clark, Drew Creal, and Rob Lester, as well as seminar participants at the Federal Reserve Banks of Dallas and Cleveland, for helpful comments. Correspondence: esims1@nd.edu, cynthia.wu@nd.edu.

1 Introduction

In response to the Financial Crisis and ensuing Great Recession of 2007-2009, the Fed and other central banks pushed policy rates to zero (or, in some cases, slightly below zero). With conventional policy unavailable, central banks launched a sequence of experimental unconventional policy interventions. The principal tools of unconventional policy include large scale asset purchases (or quantitative easing, QE), forward guidance (FG), and negative interest rate policy (NIRP). Even ten years subsequent to the Crisis, many advanced economies are still plagued by near-zero policy rates. Further, an emerging consensus is that global forces and demographic trends have lowered the “natural” or “neutral” rate of interest; see Laubach and Williams (2003) and Del Negro, Giannone, Giannoni and Tambalotti (2017). This means the Great Recession is likely not a one-off event and that heretofore unconventional tools will become a more conventional part of central banks’ tool kit. The time is therefore ripe for a comprehensive, quantitative review of these tools in a unified modeling framework.

Our paper makes a number of important contributions to the literature. First, rather than introducing them in a piecemeal fashion as in the existing literature, we develop the first quantitative DSGE model that builds in all three different types of unconventional tools. This allows us to study similarities, differences, and potential interactions between them. Second, we draw attention to two distinct channels by which NIRP can transmit to the economy. On the one hand, NIRP in our model can be stimulative via a forward guidance type channel, but on the other hand, it erodes the net worth of financial intermediaries, which works in the opposite direction. Third, we propose a novel and intuitive endogenous specification for QE that is similar to a conventional Taylor rule. We find that endogenous QE can successfully neutralize the adverse consequences of a binding ZLB. Despite its usefulness, QE does not come without a cost. A fourth contribution of our paper is to address questions relating to the size of a central bank’s balance sheet, including consequences of different normalization plans as well as the interaction between the size of the balance sheet and NIRP. Fifth, we

provide a first attempt at endogenizing an effective lower bound (ELB) for policy rates. The ELB depends on the overall size of a central bank’s balance sheet. Finally, we introduce a novel way to model forward guidance, where our modeling approach ameliorates the so-called forward guidance “puzzle” (Del Negro, Giannoni and Patterson 2015).

Though it shares similarities with canonical medium-scale models (e.g. Christiano, Eichenbaum and Evans 2005, and Smets and Wouters 2007), our model differs from standard models in some important ways. Firms are required to issue long term bonds to finance part of investment in new physical capital, similarly to Carlstrom, Fuerst and Paustian (2017). Asset markets are segmented in that households can only indirectly access long term bonds via holding short term debt (i.e. deposits) with financial intermediaries. Financial intermediaries are introduced in a way similar to Gertler and Karadi (2011, 2013). A costly enforcement problem results in an endogenous leverage constraint on intermediaries. This results in excess returns and time-varying interest rate spreads, which allows QE type policies to have real effects. New to the literature, we also formally model the central bank’s balance sheet with interest-bearing reserves. Along with a constraint on minimum reserve balances at intermediaries, this allows for the interest rate on reserves to potentially differ from the deposit rate, thereby allowing us to explore the effects of negative interest rate policy.

We consider conventional monetary policy to entail adherence to a Taylor (1993) type rule.¹ During normal times, the central bank sets the interest rate on reserves (what we hereafter refer to as the policy rate) consistent with the rate implied by the Taylor rule. If the reserve requirement is non-binding, then the deposit rate equals the policy rate.² QE in our model entails the central bank purchasing (or selling) long maturity private or government bonds, financed via the creation (or destruction) of reserves.³ QE policies can be

¹For a study of Ramsey optimal instrument rules in a model sharing similarities with ours, see Mau (2019).

²Several related papers (e.g. Gertler and Karadi 2011, 2013 and Carlstrom, Fuerst and Paustian 2017) do not model reserves at all and instead think of conventional policy as directly setting the deposit rate via a Taylor rule.

³Clarida (2012) distinguishes between different types of large scale asset purchases, calling purchases of

undertaken whether conventional policy is constrained by a zero lower bound or not, though they are more effective when short term rates are fixed. Forward guidance in our model involves cutting the desired short term policy rate when the actual policy rate is constrained by zero, which signals a commitment to lower short term rates after the constraint has lifted. We introduce into our model a reduced form parameter capturing the perceived credibility of forward guidance announcements.

Consistent with the experience of many countries that have experimented with NIRP, our model assumes that the policy rate can go negative, while the deposit rate (the economically relevant short term interest rate) cannot. To implement NIRP, the central bank can require intermediaries to hold more reserves than they would like. On the one hand, NIRP affects the economy much like forward guidance – it signals the intent to lower economically relevant short term rates in the future, which in turn can affect long term rates in the present. Because it involves a tangible action rather than mere words, it may be more credible than conventional forward guidance. A countervailing mechanism at work in the model is that NIRP reduces intermediary net worth, which, other factors held constant, works to tighten credit conditions. If the central bank’s balance sheet is small (i.e. intermediaries hold few reserves), this mechanism is relatively weak and NIRP is close to fully credible forward guidance. This is not necessarily true when the size of the central bank’s balance sheet is large.

We begin by comparing exogenous changes in both conventional and unconventional policies. In our model, a 100 basis point shock to the policy rate results in output rising by about 0.5 percent at peak. To generate a similar output expansion when conventional policy is constrained by the ZLB for two years in expectation, the central bank must expand its balance sheet by about 4 percent relative to steady state output. Generating the same output increase via FG or NIRP at the ZLB requires a cut in the desired policy rate of more

long term government bonds “quantitative easing” and purchases of privately-issued debt “credit easing.” Our model features both government and private sector long maturity bonds. We will use the term “quantitative easing” as it has been used in the financial press to refer to any type of large scale asset purchase (either government or non).

than twice the magnitude of a conventional policy shock. With a pre-QE size of the Fed's balance sheet, NIRP is less stimulative than a fully-credible FG, but more than a partially-credible one. The stimulative effects of NIRP become smaller the larger is the central bank's balance sheet.

Significant stimulus from NIRP requires very large cuts in policy rates, which for a variety of political and financial reasons are likely not feasible. Such large cuts into negative territory have also not been attempted in practice. Forward guidance may be plagued by credibility issues. Quantitative easing, in contrast, is relatively effective in our model. It has also been the most popular of unconventional tools among central banks. Most prominently, the Bank of Japan, the Fed, and the European Central Bank have each expanded their balance sheets via QE operations to roughly 5 trillion dollars, which amounts to anywhere between 25 and 100 percent of their corresponding GDPs.

Having analyzed the effects of exogenous changes in unconventional policy tools, we then focus on endogenous quantitative easing in attempt to speak to the Great Recession experience. To our knowledge, we are the first paper to model QE endogenously in the context of a detailed structural model. In particular, we posit a Taylor-type rule for quantitative easing when policy is constrained by the ZLB. The simple feedback rule for QE when the ZLB binds almost perfectly replicates the dynamics of output and other key variables compared to when the ZLB does not bind and conventional policy is active. On its own, a binding ZLB significantly amplifies cyclical fluctuations in output. Endogenous QE quite effectively dampens this excess volatility.

Endogenous quantitative easing serves as a highly effective substitute for conventional policy. This result holds conditional on different shocks, but more interestingly, also in an unconditional simulation with several adverse demand shocks meant to mimic the experience of the US during the Great Recession. In this simulation exercise, the endogenous QE rule involves expanding the central bank's balance sheet to about 25 percent of GDP, which corresponds to the Fed's balance sheet after QE3. This balance sheet expansion provides

stimulus roughly equivalent to a decline in the policy rate of about 2 percent, which is closely in line with the empirical estimates of the shadow rate from Wu and Xia (2016).

Although endogenous QE can serve as an effective antidote to the ZLB, its use does not come without costs. Many central banks around the world have accumulated a large balance sheet over the course of their QE operations, and they are now confronting questions related to the cost associated with unwinding now that the ZLB period is over. We quantitatively study different normalization plans, sometimes referred to as quantitative tightening (QT). A smooth balance sheet normalization is preferred to either an immediate normalization after a ZLB period has ended or to carrying a significantly larger balance sheet going forward. Interestingly, the normalization plan after the ZLB is expected to be over can impact how the economy fares during the ZLB. Specifically, if agents expect the central bank to maintain a large balance sheet after the ZLB has ended instead of smoothing returning to normal, the economy could experience a deeper recession even with the central bank accumulating a larger quantity of bonds.

We also study interactions between the size of a central bank’s balance sheet and the use of other unconventional tools. Other things being equal, NIRP is less effective the larger a balance sheet a central bank carries. Fixing other parameters but increasing the size of the central bank’s steady state balance sheet to 38% of GDP (which matches the Euro area as of the end of 2018), NIRP actually becomes mildly contractionary. Furthermore, there exists a cutoff negative policy rate below which financial intermediaries would voluntarily choose to exit. We refer to this cutoff rate as the effective lower bound (ELB) and find that it can be quite close to zero for an economy in which the central bank carries a large balance sheet.

1.1 Literature

The last decade has witnessed an explosion in work studying unconventional monetary policy. Kuttner (2018) provides an accessible overview. There are both reduced-form empirical papers and quantitative theoretical analyses; our paper is most similar to the latter. To date,

the majority of this literature has typically studied unconventional policies in isolation. We are one of the first to propose a theoretical framework in which the three principal types of unconventional policies may be simultaneously studied.

The principal modeling frameworks to study QE are the preferred-habitat model and DSGE models. Hamilton and Wu (2012) and Greenwood and Vayanos (2014) rely on the preferred-habitat model of Vayanos and Vila (2009) to study the empirical effects of the maturity structure of publicly held debt on the term structure. Walker (2019) incorporates a preferred habit framework into a New Keynesian model to study QE. In a similar vein, Gertler and Karadi (2011, 2013) and Carlstrom, Fuerst and Paustian (2017) build DSGE models with segmented asset markets and financial frictions to quantify the effects of QE on output and other economic aggregates. The structure of financial intermediaries in our model is similar to these papers. We differ in that we model the central bank as financing its operations through the issuance of interest-bearing reserves, which appears to be the operating framework on which the Federal Reserve has settled moving forward. The inclusion of reserves in our model is also central to studying negative interest rate policy.⁴ Moreover, to our knowledge, we are among the first papers to endogenize QE policy, which we consider more policy relevant than focusing on exogenous QE.

Del Negro, Giannoni and Patterson (2015) focus on forward guidance. They point out that standard monetary DSGE models predict implausibly large effects of central bank promises concerning the path of future policy rates. The effects of forward guidance become stronger as the projected path of policy rates goes farther off into the future, which they refer to as the “forward guidance puzzle.” McKay, Nakamura and Steinsson (2016) propose a possible resolution of the puzzle relying on heterogeneity and incomplete asset markets. Campbell, Evans, Fisher and Justiniano (2012) empirically study the responses of

⁴ Krishnamurthy and Vissing-Jorgensen (2011) use an event study methodology to quantify the effects of the different waves of QE in the United States on a variety of different interest rates. Gagnon, Raskin, Remache and Sack (2011) is one of the first papers to empirically study the effects of large scale asset purchases. They find that these purchases were successful in driving down long term interest rates primarily through lower risk and term premia. Bauer and Rudebusch (2014), in contrast, emphasize how large scale asset purchases can affect long term yields by signaling the expected path of short term rates.

asset prices and interest rates to Federal Reserve communications both before and during the Great Recession. We model forward guidance in a novel way as a shock to an underlying desired policy rate that affects the actual policy rate after a ZLB episode is over. As we later discuss, in an admittedly rather mechanical way, this ameliorates many of the puzzling theoretical results concerning the efficacy of forward guidance. We also introduce a reduced-form parameter to capture the degree of commitment involved in the implementation of forward guidance.

Several recent papers study NIRP. Abo-Zaid and Garín (2016) argue that in a New Keynesian model with credit frictions and money demand the optimal nominal interest rate may be negative. Lemke and Vladu (2016) and Kortela (2016) modify a shadow-rate term structure model to allow for a negative lower bound on the short end of the yield curve. Wu and Xia (2016) extend the shadow rate term structure model such that agents form non-trivial expectations about future negative interest rate policy, which emphasizes a forward guidance aspect to NIRP. Our model is closely related to this mechanism, though their term structure model is mute about the responses of macro aggregates to NIRP. Ulate (2018) and Eggertsson, Juelsrud, Summers and Wold (2019) both show empirically that interest rate pass through becomes weaker when policy rates go negative. They also both construct quantitative DSGE models with a bank-lending channel to study the pass through from policy rates to lending rates. Both papers argue that negative policy rates erode bank profits, which reduces pass-through from policy rates. A similar channel is at play in our model. Different than these papers, we stress the importance of the size of the central bank's balance sheet for the strength of this channel. Moreover, we compare NIRP with forward guidance and emphasize that NIRP can be thought of as forward guidance with commitment.

On balance, our conclusions align with a number of recent papers that argue that unconventional policies are an (almost) perfect substitute for conventional monetary policy and that the ZLB (or ELB) on policy rates is ultimately not much of a hindrance to effective stabilization policy. See, for example, Swanson and Williams (2014), Wu and Xia (2016),

Wu and Zhang (2017, 2019), Garín, Lester and Sims (2019), Debortoli, Galí and Gambetti (2016), Mouabbi and Sahuc (2017), and Swanson (2018a,b). While some of these and other papers do discuss different types of unconventional policies together, our analysis nevertheless has some advantages. In contrast to Wu and Xia (2016) and Swanson (2018b), who take a reduced form empirical approach, our model is structural. Compared to Wu and Xia (2016) and Wu and Zhang (2017, 2019), who produce a summary statistic for the joint effects of different types of unconventional policies, we are able to study such policies separately or in any combination. This facilitates an understanding of underlying mechanisms and allows us to compare and contrast the effectiveness of different types of unconventional policy interventions.

The remainder of the paper is organized as follows. [Section 2](#) lays out the model, and [Section 3](#) describes how our model can accommodate various monetary policy tools. [Section 4](#) compares alternative exogenous monetary policy shocks. [Section 5](#) studies endogenous quantitative easing and recreates the experience of the Great Recession. [Section 6](#) discusses future issues facing central banks related to the size of their balance sheets. [Section 7](#) offers concluding thoughts.

2 Model

In this section, we flesh out the key ingredients of our model. The principal actors in the model are households, labor unions, several types of production firms, financial intermediaries, a fiscal authority, and a central bank.

Although it shares many familiar ingredients with canonical medium-scale DSGE models (i.e. Christiano, Eichenbaum and Evans 2005, and Smets and Wouters 2007), several features of our model are less common. First, as in Carlstrom, Fuerst and Paustian (2017), we assume that production firms must issue perpetual bonds (Woodford 2001) to finance part of their new investment. These bonds allow for a succinct presentation and can easily be

mapped into a long term zero coupon bond for quantitative analysis. Second, we formally model financial intermediaries. As in Gertler and Karadi (2011, 2013), intermediaries finance themselves with net worth and short term debt (what we call deposits) and hold long term bonds issued by firms and the government as well as reserves issued by the central bank. Markets are segmented in the sense that households can only save via deposits. Because of a costly enforcement problem, intermediaries face an endogenous leverage constraint that results in excess returns. This feature, in conjunction with our assumption that firms must issue long term bonds to finance investment, generates an “investment wedge” that allows QE type policies to have real economic effects. Third, our framework is also unique in that we model the central bank as financing its operations through the issuance of interest-bearing reserves, which appears to be the operating framework on which the Federal Reserve has settled moving forward. Combined with a constraint on minimum reserve balances at financial intermediaries, this allows us to study negative interest rate policy (NIRP) when the deposit rate is constrained by the zero lower bound. Model details are in Appendix A.

2.1 Long Term Bonds

A representative wholesale firm and the fiscal authority (both discussed below) issue long term bonds to finance their activities. We follow Woodford (2001) in modeling these bonds as perpetuities with decaying coupon payments. Let $\kappa \in [0, 1]$ denote the decay parameter for coupon payments.⁵ A one unit bond issue in period t for Q_t dollars obligates the issuer to a coupon payment of one dollar in $t + 1$, κ dollars in $t + 2$, κ^2 dollars in $t + 3$, and so on.

Let $CF_{m,t}$ denote the new nominal issuance of these bonds by the representative wholesale firm (which we index by m). The total coupon liability due in period t for the firm based on past issuances is:

$$F_{m,t-1} = CF_{m,t-1} + \kappa CF_{m,t-2} + \kappa^2 CF_{m,t-3} + \dots \quad (2.1)$$

⁵For simplicity, we assume that κ is the same for both private and government bonds, although this need not be the case.

An attractive feature of these perpetual bonds is that one need not keep track of the entire sequence of past issues. Rather, one need only keep track of the period t and $t - 1$ total coupon liabilities. Iterating (2.1) forward one period, one observes:

$$CF_{m,t} = F_{m,t} - \kappa F_{m,t-1} \quad (2.2)$$

New bond issuances in period t trade at market price Q_t . Given the decaying coupon structure, bonds issued in period $t - j$ must trade at $\kappa^j Q_t$, for $j \geq 0$. This means that one only needs to keep track of the bond price for the current issue rather than the entire sequence of prices for past issues. In particular, the value of a bond portfolio consisting of all the outstanding private bonds in period t may be written:

$$Q_t F_{m,t} = Q_t CF_{m,t} + \kappa Q_t CF_{m,t-1} + \kappa^2 Q_t CF_{m,t-2} + \dots \quad (2.3)$$

Nominal government bonds have an identical structure to privately issued bonds. $CB_{G,t}$ denotes new issuances in period t , $B_{G,t}$ denotes the total nominal liability on current and past issuances, and $Q_{B,t}$ is the market price.

2.2 Financial Intermediaries

Financial intermediaries are structured similarly to Gertler and Karadi (2011, 2013). Each period there is a fixed mass of intermediaries indexed by i . Intermediaries finance themselves with net worth, $N_{i,t}$, and deposits taken from households, $D_{i,t}$. Each period, a fraction $1 - \sigma$, with $\sigma \in [0, 1]$, stochastically exit and return their net worth to their household owner. They are replaced by an equal number of new intermediaries that begin with real start up funds of X given to them by their household owner.

Intermediaries hold privately issued bonds, $F_{i,t}$; government issued nominal bonds, $B_{i,t}$; and interest-bearing reserves, $RE_{i,t}$, which are held on account with the central bank. The inclusion of reserves is different than Gertler and Karadi (2013) and ends up being useful

in thinking about negative interest rate policy. The balance sheet condition of a typical intermediary is:

$$Q_t F_{i,t} + Q_{B,t} B_{i,t} + RE_{i,t} = D_{i,t} + N_{i,t} \quad (2.4)$$

A financial intermediary accumulates net worth until stochastically exiting. Net worth for surviving intermediaries evolves according to:

$$\begin{aligned} N_{i,t} = & (R_t^F - R_{t-1}^d) Q_{t-1} F_{i,t-1} + (R_t^B - R_{t-1}^d) Q_{B,t-1} B_{i,t-1} \\ & + (R_{t-1}^{re} - R_{t-1}^d) RE_{i,t-1} + R_{t-1}^d N_{i,t-1} \end{aligned} \quad (2.5)$$

R_{t-1}^{re} is the (gross) interest rate on reserves, which is set by the central bank and known at $t - 1$. R_{t-1}^d is the deposit rate, which is determined in equilibrium. The first three terms are the excess returns from holding private bonds, government bonds, and reserves relative to the cost of funding via deposits. The last term measures cost savings from financing with net worth as opposed to deposits. R_t^F and R_t^B are the realized holding period returns on private and government issued bonds and satisfy:

$$R_t^F = \frac{1 + \kappa Q_t}{Q_{t-1}} \quad (2.6)$$

$$R_t^B = \frac{1 + \kappa Q_{B,t}}{Q_{B,t-1}} \quad (2.7)$$

The objective of an intermediary is to maximize its expected terminal net worth where discounting is by the stochastic discount factor of the household, $\Lambda_{t,t+1}$. Consider the problem of an intermediary continuing after period t . There is a $1 - \sigma$ probability that it will exit after $t + 1$, a $(1 - \sigma)\sigma$ probability that it will exit after $t + 2$, and so on. Accordingly, its objective is

$$V_{i,t} = \max (1 - \sigma) \mathbb{E}_t \sum_{j=1}^{\infty} \sigma^{j-1} \Lambda_{t,t+j} n_{i,t+j} \quad (2.8)$$

where $n_{i,t} = N_{i,t}/P_t$ is real net worth, with P_t the price of final output.

A financial intermediary faces two constraints. One is a costly enforcement problem as in Gertler and Karadi (2011, 2013). A financial intermediary can choose to abscond with some assets at the end of a period rather than continuing as an intermediary. If an intermediary does this, depositors can recover a fraction of the intermediary's assets, with the intermediary retaining the rest. For depositors to be willing to lend to intermediaries, it must not be optimal for the intermediary to divert funds in this way, which we refer to as going into bankruptcy. Accordingly:

$$V_{i,t} \geq \theta_t (Q_t f_{i,t} + \Delta Q_{B,t} b_{i,t}) \quad (2.9)$$

In (2.9), the right hand side of the inequality represents the (real) funds that a financial intermediary can keep should it choose to enter bankruptcy, while the left hand side is the value of continuing as an intermediary.⁶ Should it choose to divert, an intermediary can keep a stochastic fraction of its private bonds, θ_t . It can keep the fraction $\theta_t \Delta$ of government bonds, where $0 \leq \Delta \leq 1$. This means that it is easier for the intermediary to divert private bonds than government bonds. We choose to make θ_t stochastic (and exogenous). θ_t may be thought of as a type of liquidity shock – when θ_t increases, depositors can recover a smaller fraction of an intermediary's assets in the event of bankruptcy, which in turn makes them less willing to lend funds. This will have the effect of driving interest rate spreads up, which is a hallmark of liquidity crises. We assume that the third type of asset held by intermediaries – reserves – is fully recoverable by depositors in the event of bankruptcy.

The second constraint faced by intermediaries is a reserve requirement. Relative to the extant literature, this is new in our model. We assume that intermediaries are required to

⁶ $f_{i,t} = F_{i,t}/P_t$ and $b_{i,t} = B_{i,t}/P_t$ are real private and government bond holdings, respectively.

hold a minimum level of reserves that is set by the central bank. The reserve requirement is time-varying and proportional to an intermediary's deposits. Letting $re_{i,t} = RE_{i,t}/P_t$ and $d_{i,t} = D_{i,t}/P_t$ denote real reserve holdings and real deposits, the constraint is:

$$re_{i,t} \geq \varsigma_t d_{i,t} \quad (2.10)$$

For most of the analysis, the reserve requirement will be non-binding. (2.10) is included so as to potentially allow for a negative interest rate on reserves. When this constraint binds for all intermediaries, $\varsigma_t = \frac{re_t}{d_t}$.

All financial intermediaries will behave in the same way with identical optimality conditions. These are

$$\mathbb{E}_t \Lambda_{t,t+1} \Omega_{t+1} \Pi_{t+1}^{-1} (R_{t+1}^F - R_t^d) = \frac{\lambda_t}{1 + \lambda_t} \theta_t \quad (2.11)$$

$$\mathbb{E}_t \Lambda_{t,t+1} \Omega_{t+1} \Pi_{t+1}^{-1} (R_{t+1}^B - R_t^d) = \frac{\lambda_t}{1 + \lambda_t} \theta_t \Delta \quad (2.12)$$

$$\mathbb{E}_t \Lambda_{t,t+1} \Omega_{t+1} \Pi_{t+1}^{-1} (R_t^{re} - R_t^d) = -\frac{\omega_t}{1 + \lambda_t} \quad (2.13)$$

where

$$\Omega_t = 1 - \sigma + \sigma \theta_t \phi_t \quad (2.14)$$

$$\phi_t = \frac{1 + \lambda_t}{\theta_t} \mathbb{E}_t [\Lambda_{t,t+1} \Omega_{t+1} \Pi_{t+1}^{-1}] R_t^d - \frac{\omega_t re_t}{n_t \theta_t} \quad (2.15)$$

(2.11) - (2.13) are the key equilibrium conditions in the model. $\lambda_t \geq 0$ is the multiplier on the costly enforcement constraint, (2.9), while ω_t is the multiplier on the reserve requirement, (2.10). If neither constraint binds, then to first order expected returns on all three types of assets must equal the cost of funds (i.e. the deposit rate). If the costly enforcement constraint binds, then there will be excess returns of private and public bonds over the deposit rate. $\Delta < 1$ means that excess returns on government bonds will be lower than excess returns on private bonds. The interest rate on reserves can never exceed the deposit rate, but it could

be less than the deposit rate when the reserve requirement is binding. (2.14) is an auxiliary variable introduced to simplify the analysis. See derivations in Appendix A.1.

We show in Appendix A.1 that the value of an intermediary satisfies

$$V_{i,t} = \theta_t \phi_t n_{i,t}. \quad (2.16)$$

When the constraint in (2.9) binds,

$$\phi_t = \frac{Q_t f_{i,t} + \Delta Q_{B,t} b_{i,t}}{n_{i,t}} \quad (2.17)$$

which is an endogenous leverage ratio, whose equilibrium condition is in (2.15). The constraint makes the financial intermediary less levered than it would find optimal. This endogenous leverage constraint is ultimately what can give rise to excess returns.

One can show from (2.15) that:

$$\theta_t \phi_t \geq 1 + \lambda_t - \frac{\omega_t r e_t}{n_t} \quad (2.18)$$

If neither (2.9) nor (2.10) bind, then $\lambda_{t+j} = \omega_{t+j} = 0$ for all j , which implies that $\theta_t \phi_t = 1$.⁷ Intuitively, this means that net worth is as valuable to a household as to an intermediary. In this case, returns on all assets are equal. Hence, whether an intermediary invests in F_t , B_t , D_t or RE_t is irrelevant.

The two different constraints have competing effects on the value of an intermediary. Other things being equal, when the costly enforcement constraint, (2.9), binds, then $\lambda_t > 0$ and $\theta_t \phi_t$ is larger. In this case, there exist excess returns on holding long term assets (private and government bonds). Hence, net worth is more valuable inside an intermediary as opposed to a household (who cannot hold these assets and hence cannot take advantage of these excess returns). In contrast, when the reserve requirement, (2.10), binds, $\theta_t \phi_t$ gets

⁷In this circumstance, $\Omega_t = 1$. From the household's first order condition, $\mathbb{E}_t \Lambda_{t,t+1} \Pi_{t+1}^{-1} R_t^d = 1$.

smaller, other things being equal. A binding reserve requirement pushes the interest rate on reserves below the deposit rate, which means that intermediaries suffer a loss for each dollar of reserves they hold.

2.3 Households

There exist a continuum of households of unit measure. These households behave identically and for what follows we omit household indexes. Our setup follows Gertler and Karadi (2013). Within each household there are two types of members – workers and intermediaries. At any point in time a fixed fraction of household members are workers and a fixed fraction are intermediaries. As discussed above in [Subsection 2.2](#), intermediaries stochastically exit and become workers with probability $1 - \sigma$ and are replaced with an equal number of workers. These new intermediaries are given a fixed amount of startup net worth.

Lifetime utility for a household is given by:

$$\mathbb{E}_t \sum_{j=0}^{\infty} \beta^j \left\{ \ln (C_{t+j} - bC_{t+j-1}) - \frac{\chi L_{t+j}^{1+\eta}}{1+\eta} \right\} \quad (2.19)$$

where $\beta \in (0, 1)$ is a discount factor and $b \in [0, 1)$ is a measure of internal habit formation. $\chi > 0$ is a scaling parameter and η is the inverse Frisch elasticity. C_t is consumption and L_t is labor supply.

A household faces the following nominal budget constraint:

$$P_t C_t + D_t - D_{t-1} \leq MRS_t L_t + DIV_t - P_t X - P_t T_t + (R_{t-1}^d - 1) D_{t-1} \quad (2.20)$$

In [\(2.20\)](#), P_t is the price of goods and D_{t-1} is the (nominal) stock of deposits with which a household enters a period. R_t^d is the gross nominal interest rate on deposits. MRS_t is the nominal remuneration a household receives from supplying labor to labor unions (discussed below). DIV_t denotes nominal dividends from ownership in all non-financial firms as well as net worth from exiting intermediaries. X is a (real) transfer paid out to new intermediaries

as startup net worth. T_t is a lump sum tax paid to a government.

The first order conditions for the household are

$$\mu_t = \frac{1}{C_t - bC_{t-1}} - b\beta \mathbb{E}_t \frac{1}{C_{t+1} - bC_t} \quad (2.21)$$

$$\Lambda_{t,t+1} = \frac{\beta\mu_{t+1}}{\mu_t} \quad (2.22)$$

$$\chi L_t^\eta = \mu_t mrs_t \quad (2.23)$$

$$1 = R_t^d \mathbb{E}_t \Lambda_{t,t+1} \Pi_{t+1}^{-1} \quad (2.24)$$

(2.21) defines the marginal utility of consumption, μ_t . The stochastic discount factor, $\Lambda_{t,t+1}$, is given by (2.22). (2.23) is a standard labor supply condition, where $mrs_t = MRS_t/P_t$. The first order condition for deposits is (2.24), where $\Pi_t = P_t/P_{t-1}$ is the gross inflation rate.

2.4 Labor Market

There are two layers to the labor market. A unit measure of labor unions, indexed by $h \in [0, 1]$, purchase labor from households at MRS_t and repackage it for sale to a representative labor packer. $L_{d,t}(h)$ is labor sold to the labor packer, and $L_{d,t}(h) = L_t(h)$ for each union. The labor packer combines differentiated labor into final labor available for production, $L_{d,t}$, via a CES technology with elasticity of substitution $\epsilon_w > 1$. The demand curve facing each union is

$$L_{d,t}(h) = \left(\frac{W_t(h)}{W_t} \right)^{-\epsilon_w} L_{d,t} \quad (2.25)$$

$W_t(h)$ is the wage paid for union h 's labor and W_t is the aggregate wage, which satisfies:

$$W_t^{1-\epsilon_w} = \int_0^1 W_t(h)^{1-\epsilon_w} dh \quad (2.26)$$

Labor unions are subject to a Calvo-style nominal rigidity. Each period, there is a $1 - \phi_w$ probability that a union may adjust its wage, with $\phi_w \in [0, 1]$. Non-updated wages may be

indexed to lagged inflation at $\gamma_w \in [0, 1]$. Unions maximize the present discounted value of flow profits where discounting is by the household's stochastic discount factor. All updating unions choose the same reset wage, W_t^* . Written in real terms, $w_t^* = W_t^*/P_t$, it satisfies:

$$w_t^* = \frac{\epsilon_w}{\epsilon_w - 1} \frac{f_{1,t}}{f_{2,t}} \quad (2.27)$$

$$f_{1,t} = mrs_t w_t^{\epsilon_w} L_{d,t} + \phi_w \mathbb{E}_t \Lambda_{t,t+1} \left(\frac{\Pi_{t+1}}{\Pi_t^{\gamma_w}} \right)^{\epsilon_w} f_{1,t+1} \quad (2.28)$$

$$f_{2,t} = w_t^{\epsilon_w} L_{d,t} + \phi_w \mathbb{E}_t \Lambda_{t,t+1} \left(\frac{\Pi_{t+1}}{\Pi_t^{\gamma_w}} \right)^{\epsilon_w - 1} f_{2,t+1} \quad (2.29)$$

where $w_t = W_t/P_t$ is the aggregate real wage from (2.26). See derivations in Appendix A.2.1.

2.5 Production

There are several layers on the production side of the economy. A representative wholesale firm transforms capital and labor into output, $Y_{m,t}$. New physical capital, \widehat{I}_t , is produced by a competitive capital goods producer. Wholesale output is sold to a continuum of retail firms. Retail firms are indexed by $f \in [0, 1]$. These retail firms simply repackage wholesale output via $Y_t(f) = Y_{m,t}(f)$ and then sell their output to a competitive final goods firm. The final output good, Y_t , is a CES aggregate of retail outputs with elasticity of substitution $\epsilon_p > 1$. Retailers thus face the demand curve:

$$Y_t(f) = \left(\frac{P_t(f)}{P_t} \right)^{-\epsilon_p} Y_t \quad (2.30)$$

In (2.30), $P_t(f)$ is the price of retail output. The price of the final output good, P_t , satisfies:

$$P_t^{1-\epsilon_p} = \int_0^1 P_t(f)^{1-\epsilon_p} df \quad (2.31)$$

Retail firms purchase output from the representative wholesale firm at $P_{m,t}$, which retailers take as given. Retailers are subject to a Calvo-style nominal rigidity. Each period, there is a $1 - \phi_p$ probability they can adjust their price. Non-updated prices may be indexed to lagged inflation at $\gamma_p \in [0, 1]$. Future profits are discounted via the household's stochastic discount factor. All updating retailers adjust to the same reset price, P_t^* . Written in real terms, $p_t^* = P_t^*/P_t$, it satisfies:

$$p_t^* = \frac{\epsilon_p}{\epsilon_p - 1} \frac{x_{1,t}}{x_{2,t}} \quad (2.32)$$

$$x_{1,t} = p_{m,t} Y_t + \phi_p \mathbb{E}_t \Lambda_{t,t+1} \left(\frac{\Pi_{t+1}}{\Pi_t^{\gamma_p}} \right)^{\epsilon_p} x_{1,t+1} \quad (2.33)$$

$$x_{2,t} = Y_t + \phi_p \mathbb{E}_t \Lambda_{t,t+1} \left(\frac{\Pi_{t+1}}{\Pi_t^{\gamma_p}} \right)^{\epsilon_p - 1} x_{2,t+1} \quad (2.34)$$

where $p_{m,t} = \frac{P_{m,t}}{P_t}$. See details in Appendix A.2.2.

The representative wholesale firm produces output according to a Cobb-Douglas technology:

$$Y_{m,t} = A_t (u_t K_t)^\alpha L_{d,t}^{1-\alpha} \quad (2.35)$$

$Y_{m,t}$ is flow output and $L_{d,t}$ is labor input. $0 < \alpha < 1$ is the exponent on capital services in the production function. A_t is an exogenous productivity variable that obeys an exogenous stochastic process. u_t is capital utilization. K_t is the stock of physical capital, which the firm owns. The cost of utilization is faster depreciation, where $\delta(u_t)$ maps utilization into depreciation. Physical capital accumulates according to a standard law of motion:

$$K_{t+1} = \widehat{I}_t + (1 - \delta(u_t)) K_t \quad (2.36)$$

Similar to Carlstrom, Fuerst and Paustian (2017), we assume that the wholesale firm must issue perpetual bonds to finance the purchase of new physical capital. Different than them, we only require the firm to finance a constant fraction, $\psi \in [0, 1]$, of investment, not the entirety. This gives rise to a “loan in advance constraint” of the form:

$$\psi P_t^k \widehat{I}_t \leq Q_t C F_{m,t} = Q_t (F_{m,t} - \kappa F_{m,t-1}), \quad (2.37)$$

where P_t^k is the price at which the wholesale firm purchases new physical capital.

The wholesale firm hires labor in a competitive spot market at nominal wage W_t . Its nominal dividend is

$$DIV_{m,t} = P_{m,t} A_t (u_t K_t)^\alpha L_{d,t}^{1-\alpha} - W_t L_{d,t} - P_t^k \widehat{I}_t - F_{m,t-1} + Q_t (F_{m,t} - \kappa F_{m,t-1}) \quad (2.38)$$

The firm maximizes the present discounted value of real dividends, where discounting is by the stochastic discount factor of households. The first order conditions are

$$w_t = (1 - \alpha) p_{m,t} A_t (u_t K_t)^\alpha L_{d,t}^{-\alpha} \quad (2.39)$$

$$p_t^k M_{1,t} \delta'(u_t) = \alpha p_{m,t} (u_t K_t)^{\alpha-1} L_{d,t}^{1-\alpha} \quad (2.40)$$

$$p_t^k M_{1,t} = \mathbb{E}_t \Lambda_{t,t+1} [\alpha p_{m,t+1} A_{t+1} K_{t+1}^{\alpha-1} u_{t+1}^\alpha L_{d,t+1}^{1-\alpha} + (1 - \delta(u_{t+1})) p_{t+1}^k M_{1,t+1}] \quad (2.41)$$

$$Q_t M_{2,t} = \mathbb{E}_t \Lambda_{t,t+1} \Pi_{t+1}^{-1} [1 + \kappa Q_{t+1} M_{2,t+1}] \quad (2.42)$$

$$\frac{M_{1,t} - 1}{M_{2,t} - 1} = \psi \quad (2.43)$$

$w_t = W_t/P_t$ is the real wage, $p_{m,t} = P_{m,t}/P_t$ is the relative price of wholesale output, and $p_t^k = P_t^k/P_t$ is the relative price of new capital. (2.39) is the standard static first order condition for labor demand. $M_{1,t}$ is one plus the product of ψ with the multiplier on the constraint that firms must issue bonds to finance investment, (2.37), while $M_{2,t}$ is simply one plus the multiplier on the constraint. (2.40) is the first order condition for utilization. (2.41) and (2.42) are optimality conditions for capital and bonds, respectively. If the constraint did not bind, then $M_{1,t} = M_{2,t} = 1$ and (2.41)-(2.42) would reduce to standard asset pricing conditions. $M_{1,t}$ serves as an endogenous “investment wedge” and $M_{2,t}$ can be thought of as a “financial wedge.” These wedges distort the standard asset pricing decisions and fluctuations in these wedges are the mechanism through which QE type policies transmit to

the real economy. See details in Appendix [A.3](#).

A representative capital producer generates new physical capital via:

$$\hat{I}_t = \left[1 - S \left(\frac{I_t}{I_{t-1}} \right) \right] I_t \quad (2.44)$$

I_t is unconsumed final output. $S(\cdot)$ is an adjustment cost. Profits are discounted by the household's stochastic discount factor. The optimality condition is

$$1 = p_t^k \left[1 - S \left(\frac{I_t}{I_{t-1}} \right) - S' \left(\frac{I_t}{I_{t-1}} \right) \frac{I_t}{I_{t-1}} \right] + \mathbb{E}_t \Lambda_{t,t+1} p_{t+1}^k S' \left(\frac{I_{t+1}}{I_t} \right) \left(\frac{I_{t+1}}{I_t} \right)^2 \quad (2.45)$$

The derivation is in Appendix [A.4](#).

2.6 Fiscal Authority

A fiscal authority consumes an exogenous and stochastic amount of final output, G_t . It finances this expenditure by levying lump sum taxes on the household, with a transfer from the central bank, and by issuing (nominal) bonds, $B_{G,t}$.

Because of the frictions facing financial intermediaries, Ricardian Equivalence does not hold and the mix of lump sum tax and bond finance is relevant. As a simplification, we assume that there is a fixed quantity of outstanding real government bonds, i.e. $b_{G,t} = \bar{b}_G$, where nominal government bonds grow with the price level, i.e. $B_{G,t} = P_t \bar{b}_G$. Lump sum taxes endogenously adjust so that the government's budget constraint holds each period. Let $T_{cb,t}$ denote the total (real) transfer received from the central bank each period. The government's budget constraint is:

$$P_t G_t + P_{t-1} \bar{b}_G = P_t T_t + P_t T_{cb,t} + Q_{B,t} P_t \bar{b}_G (1 - \kappa \Pi_t^{-1}) \quad (2.46)$$

2.7 Central Bank

The central bank can hold private or government bonds on the asset side of its balance sheet. It finances these holdings by creating reserves. The central bank holds no equity and any operating surplus is returned to the fiscal authority via a lump sum transfer each period. The nominal balance sheet condition is:

$$Q_t F_{cb,t} + Q_{B,t} B_{cb,t} = R E_t \quad (2.47)$$

A full characterization of monetary policy requires specification of rules for quantities of bonds held and reserves, as well as a rule for the interest rate on reserves (which we shall refer to as the policy rate). We defer a discussion of these issues until the next section.

2.8 Market-Clearing

Privately issued bonds and government issued bonds must be held by either financial intermediaries or the central bank. In real terms, the bond market-clearing conditions are:

$$f_{m,t} = \sum_i f_{i,t} + f_{cb,t} \quad (2.48)$$

$$\bar{b}_G = \sum_i b_{i,t} + b_{cb,t} \quad (2.49)$$

The exogenous stochastic variables θ_t , A_t , and G_t obey standard AR(1) processes. Other aggregation results are standard and, along with the exogenous processes, are presented in more depth in [Appendix A.5](#).

3 (Un)conventional monetary policy tools

In this section, we formally discuss how our model can accommodate both conventional as well as unconventional monetary policy. To date, the literature has tended to study unconventional policy tools in isolation. We are the first to build a model in which all three types of policies can simultaneously be studied.

Before discussing unconventional policies, we must first specify what is meant by conventional policy. We define conventional policy as the adjustment of a short term interest rate, R_t^{TR} , and characterize it via an endogenous feedback rule similar to Taylor (1993):

$$\ln R_t^{TR} = (1 - \rho_r) \ln R^{TR} + \rho_r \ln R_{t-1}^{TR} + (1 - \rho_r) [\phi_\pi (\ln \Pi_t - \ln \Pi) + \phi_y (\ln Y_t - \ln Y_{t-1})] + s_r \varepsilon_{r,t} \quad (3.1)$$

where R^{TR} and Π are steady state values of the this policy rate and the inflation target, and $0 < \rho_r < 1$, ϕ_π , and ϕ_y are non-negative parameters. We restrict attention to parameter values giving rise to a determinate equilibrium, i.e. we assume $\phi_\pi > 1$. The policy rate adjusts in response to deviations of inflation from target and output growth from trend (which in our model is zero).⁸

During “normal” times, we assume that (i) the central bank sets the interest rate on reserves equal to the underlying policy rate, and (ii) the reserve requirement, (2.10), is non-binding, so that

$$R_t^d = R_t^{re} = R_t^{TR}. \quad (3.2)$$

⁸In the model there is no analytical expression for “potential” output and with the myriad nominal and financial frictions it is not even theoretically clear what concept of potential ought to be used. Responding to output growth may be both theoretically desirable (e.g. Walsh 2003 and Orphanides and Williams 2006) and empirically more in-line with observed central bank reaction functions (e.g. Coibion and Gorodnichenko 2011).

As we consider the economy in the “cashless limit,” adjustment of the policy rate via (3.1) during normal times does not involve central bank purchases or sales of any kind of asset. We can simply think of central bank holdings of private and government bonds as exogenously fixed. Given general equilibrium changes in bond prices arising due to changes in the policy rate, the central bank must adjust reserves so that its balance sheet, (2.47), holds. One could alternatively ignore the central bank’s balance sheet altogether and instead think of the deposit rate, R_t^d , as the policy instrument in normal times (as in, for example, Gertler and Karadi 2011, 2013).

Unconventional policies of the sort described and analyzed below have by-and-large been attempted during an environment in which short term rates were constrained from below by zero. To implement a conventional ZLB constraint, we suppose that the reserve requirement is non-binding, so that the deposit rate and interest rate on reserves are equal, and that the net interest rate on reserves must be non-negative:

$$R_t^{re} = R_t^d = \max \{1, R_t^{TR}\} \quad (3.3)$$

3.1 Quantitative easing

Perhaps the most prominent unconventional policy tool implement by central banks throughout the world is quantitative easing (QE). It was first introduced by the Bank of Japan in the early 2000s. In the aftermath of the Great Recession and subsequent ZLB period, the United States, Euro Area, and United Kingdom (among others) all adopted this policy tool. In the United States, by the end of QE3 the Federal Reserve had expanded its balance sheet to 4.5 trillion dollars (or approximately 25 percent of GDP). By the end of 2018, the ECB’s balance sheet had grown to 4.7 trillion Euros, which was about 38 percent of GDP in the Euro Area, and the Bank of Japan held 552 trillion in Yen, an amount in excess of 100 percent of Japanese GDP.

As in Gertler and Karadi (2011, 2013) and Carlstrom, Fuerst and Paustian (2017), we

interpret quantitative easing as central bank purchases of private or government issued bonds. These purchases are financed via the creation of interest-bearing reserves held by financial intermediaries. This is a realistic description of how QE policies played out in practice in the United States subsequent to the crisis. This also distinguishes our model from both Carlstrom, Fuerst and Paustian (2017) and Gertler and Karadi (2011, 2013).⁹

QE policies can have real effects to the extent to which financial intermediaries are constrained via the costly enforcement problem, (2.9). When intermediaries are constrained, bond purchases financed via the creation of reserves ease this constraint. In this case, central bank bond purchases do not simply crowd out intermediary bond purchases, in such a way that the total demand for bonds increases. This results in higher bond prices, which in turn eases the loan-in-advance constraint facing the wholesale firm. This results in higher investment and aggregate demand. If the constraint on intermediaries is non-binding, or if the wholesale firm does not need to finance investment via debt (i.e. $\psi = 0$ in (2.37)), then bond purchases have no economic effects. Provided $\Delta < 1$ and both of these constraints bind, central bank purchases of private bonds will have a stronger effect on excess returns compared to government bond purchases, and will hence be more stimulative.

We consider both exogenous and endogenous QE policies. For exogenous policy, we simply assume that central bank bond holdings follow exogenous AR(1) processes:

$$f_{cb,t} = (1 - \rho_f)f_{cb} + \rho_f f_{cb,t-1} + s_f \varepsilon_{f,t} \quad (3.4)$$

$$b_{cb,t} = (1 - \rho_b)b_{cb} + \rho_b b_{cb,t-1} + s_b \varepsilon_{b,t} \quad (3.5)$$

where f_{cb} and b_{cb} denote steady state (real) central bank private and government bond holdings, ρ_f and ρ_b are parameters constrained to lie between zero and one, and shocks are drawn from standard normal distributions (with s_f and s_b denoting standard deviations of

⁹Carlstrom, Fuerst and Paustian (2017) do not model the central bank's balance sheet at all, and Gertler and Karadi (2011, 2013) assume that the central bank finances purchases by issuing short term debt directly to households (though Gertler and Karadi 2013 discuss the similarity of such a setup to one with interest-bearing reserves).

the shocks). Endogenous QE policies are considered in [Section 5](#). Bond purchases may be undertaken either in “normal” times, when the policy rate is set according to [\(3.1\)](#), or during periods in which the policy rate is constrained by zero as in [\(3.3\)](#).

3.2 Forward guidance

Forward guidance entails promises from a central bank about the future path of its policy rate (Campbell, Evans, Fisher and Justiniano 2012 and Del Negro, Giannoni and Patterson 2015). Some degree of forward guidance has been used as a communications tool by central banks around the world for some time, though it has gained more attention recently due to the ZLB constraint on short term policy rates. Language used in central bank communications has varied widely across both space and time. For example, the Federal Reserve started by communicating its expected duration of extremely low policy rates, with announcements to the effect of “we expect conditions to warrant exceptionally low levels of the federal funds rate for/through...” Initially the forecast of the duration of low rates was ambiguous, with phrases such as “for some time” or for “an extended period.” Later the Fed turned to more specific guidance regarding the duration of low policy rates. As the duration of the ZLB period wore on, the Fed also resorted to more explicitly contingent language, for example promising low rates until a target unemployment rate of 6.5 percent was achieved. Mario Draghi of the ECB used words to influence policy with his famous “whatever it takes” remark in 2012.

We model forward guidance as a shock to the underlying desired policy rate in [\(3.1\)](#) at a time when the ZLB constraint, [\(3.3\)](#), is binding. Taking the expected duration of the ZLB, H , as given, a negative shock to the desired policy rate portends a lower interest rate on reserves (and hence a lower deposit rate, assuming the two are equal) *after* the ZLB is over. As a reduced form way to capture the credibility of forward guidance announcements, we scale the shock to the desired policy rate in a forward guidance experiment by $\gamma \in [0, 1]$. $\gamma = 1$ entails perfect credibility, whereas $\gamma = 0$ means that forward guidance announcements are not

believed by agents in the model. To the extent to which forward guidance announcements are credible, such shocks can impact long term yields in the present via the logic of the expectations hypothesis.

Our modeling of forward guidance is admittedly different from the extant literature, which typically captures forward guidance as an anticipated shock to a policy rule after a ZLB period is expected to be over. The difference in timing is important – a one unit shock to the desired policy rate at t has a much smaller effect on the economy than an announced shock of the same size to the policy rate in $t + H$. This is because of decay due to interest smoothing (i.e. $0 < \rho_r < 1$) – the change in the expected path of future deposit rates is smaller in our experiment compared to a more standard forward guidance shock. In a somewhat mechanical way, therefore, our new timing assumption leaves our model immune from the so-called “forward guidance puzzle,” which refers to the fact that forward guidance becomes more powerful as it targets policy rates in the more distant future (McKay, Nakamura and Steinsson 2016).¹⁰

3.3 Negative interest rate policy

Although it has not been attempted by the Federal Reserve, a number of central banks around the world have experimented with negative interest rate policy (NIRP) in the last several years – the Euro Area and Japan being prominent examples. While recent experience suggests that interbank interest rates can go negative, deposit rates in countries experimenting with NIRP have not. Accordingly, we implement NIRP in our model by assuming that the ZLB constraint applies only to the deposit rate and not to the interest rate on reserves.¹¹

¹⁰To be perfectly transparent, in our model a shock which lowers the deposit rate by a given amount H periods into the future has a *smaller* stimulative impact than a shock which lowers the deposit rate by the same amount $H + 1$ periods into the future. While it is therefore afflicted by the “forward guidance puzzle” so-defined, we should note that the puzzle is *less severe* in our model than in a version of our model with unconstrained financial intermediaries.

¹¹The conventional logic for the ZLB is that deposits and cash are substitutes, and cash pays a zero nominal return, so rates on deposits cannot go below zero or else households will hoard cash. Our model is cashless, so it is not immediately clear why such logic would carry over. We could instead allow the household to hold a non-interest bearing asset called cash, and could generate a non-zero demand for it by assuming

The inclusion of interest-bearing reserves in our model therefore allows us to incorporate NIRP into a model that can also be used to study other unconventional policy tools.

Formally, we assume that the central bank sets the interest rate on reserves equal to the desired underlying policy rate, R_t^{TR} . For the purposes of the presentation here, there is no constraint on how negative the policy rate can go. We defer a discussion of the ELB until [Subsection 6.2](#). We only assume that the deposit rate cannot go negative via a modification of (3.3):

$$R_t^d = \max \{1, R_t^{re}\}, \quad R_t^{re} = R_t^{TR} \quad (3.6)$$

We model NIRP as a negative shock to the underlying desired policy rate, met by an equal reduction in the interest rate on reserves, at a time where the deposit rate is constrained by zero. Implementation of such a policy requires that the reserve requirement, (2.10), which is non-binding in all of the policy experiments discussed above, binds. Intuitively, absent a reserve requirement, intermediaries are willing to hold an indeterminate amount of reserves so long as $R_t^{re} = R_t^d$. They would not wish to hold any reserves if $R_t^{re} < R_t^d$. To get intermediaries to hold reserves with a negative gap between the rate on reserves and the rate on deposits, we simply assume that the central bank requires them to do so, which makes (2.10) binding.

NIRP affects the economy through two distinct channels in our model. First, NIRP has an effect similar to forward guidance – to the extent to which ρ_r (the interest rate smoothing parameter) is large relative to H (the duration of the ZLB), lowering the interest rate on reserves in the present signals lower deposit rates in the future after the ZLB has ended. Since implementation of NIRP involves an observable action (lowering the interest rate on reserves), it may well be more credible than pure forward guidance. The second channel through which NIRP operates works in the opposite direction. $R_t^{re} < R_t^d$ works like a tax on intermediaries in the evolution of net worth, (2.5). This lowers net worth and exacerbates

that households receive utility from holding money balances. As long as the utility from money balances is additively separable from utility from consumption and leisure, this would not affect the equilibrium dynamics of the model.

the costly enforcement constraint facing intermediaries; other things being equal, this results in higher excess returns. This channel is similar to what is emphasized both theoretically and empirically in Ulate (2018) and Eggertsson, Juelsrud, Summers and Wold (2019). If forward guidance is completely credible (i.e. $\gamma = 1$), NIRP will therefore be less stimulative than pure forward guidance. If the central bank is not perfectly credible ($\gamma < 1$), then this may not be the case. How important this offsetting effect of NIRP coming through the evolution of net worth is depends on how many reserves intermediaries hold. This in turn has implications for the size of the central bank's balance sheet.

4 Comparing Alternative Monetary Policy

In this section, we compare and contrast *exogenous* changes in alternative unconventional policy tools to an exogenous shock to conventional policy. We wish to assess the efficacy of unconventional policy interventions in affecting output and other macroeconomic aggregates.

4.1 Calibration

The values assigned to relevant parameters of the model are given in Table 1. The model is solved via a linear approximation about the non-stochastic steady state.¹² When considering the effects of the ZLB, we solve a piecewise linear version of the model using the approach suggested by Guerrieri and Iacoviello (2015).

Most of the parameters take on standard values drawn from the literature (shown in the top panel). The frequency of time is quarterly, and accordingly the discount factor and steady state depreciation rate on physical capital are set to conventional values, $\beta = 0.995$ and $\delta_0 = 0.025$. The parameter on the linear term in the utilization cost function, δ_1 , is chosen to be consistent with a steady state normalization of utilization to unity.¹³ The

¹²In this steady state neither the ZLB nor the reserve requirements bind, so that the steady state reserve and deposit rates are equal.

¹³The utilization adjustment cost function takes the form: $\delta(u_t) = \delta_0 + \delta_1(u_t - 1) + \frac{\delta_2}{2}(u_t - 1)^2$.

Table 1: Calibrated Parameters

Parameter	Value or Target	Description
<i>Standard parameters</i>		
β	0.995	Discount factor
b	0.7	Habit formation
η	1	Inverse Frisch elasticity
χ	$L = 1$	Labor disutility scaling parameter / steady state labor
α	0.33	Production function exponent on capital
δ_0	0.025	Steady state depreciation
δ_1	$u = 1$	Utilization linear term / steady state utilization
δ_2	0.01	Utilization squared term
κ_I	2	Investment adjustment cost
Π	1	Steady state (gross) inflation
ϵ_p	11	Elasticity of substitution goods
ϵ_w	11	Elasticity of substitution labor
ϕ_p	0.75	Price rigidity
ϕ_w	0.75	Wage rigidity
γ_p	0	Price indexation
γ_w	0	Wage indexation
\bar{b}_G	$\frac{B_G Q_B}{4Y} = 0.41$	Steady state government debt
G	$\frac{G}{Y} = 0.2$	Steady state government spending
ρ_r	0.8	Taylor rule smoothing
ϕ_π	1.5	Taylor rule inflation
ϕ_y	0.25	Taylor rule output growth
ρ_A	0.95	AR productivity
ρ_G	0.95	AR government spending
<i>Non-Standard parameters</i>		
κ	$1 - 40^{-1}$	Bond duration
ψ	0.81	Fraction of investment from debt
σ	0.95	Intermediary survival probability
θ	$400(R^F - R^d) = 3$	Recoverability parameter / steady state spread
X	Leverage = 4	Transfer to new intermediaries / steady state leverage
Δ	1/3	Government bond recoverability
b_{cb}	$\frac{b_{cb} Q_B}{4Y} = 0.06$	Steady state central bank Treasury holdings
f_{cb}	0	Steady state central private bond holdings
ρ_θ	0.98	AR liquidity
ρ_b	0.8	AR central bank Treasury
ρ_f	0.8	AR central bank private bonds
<i>Shock sizes</i>		
s_A	0.0065	SD productivity
s_G	0.01	SD government spending
s_θ	0.04	SD liquidity

Note: this table lists the values of calibrated parameters or the target used in the calibration.

habit formation parameter is set to $b = 0.7$, and the inverse Frisch labor supply elasticity is $\eta = 1$. These are both standard values. The scaling parameter on the disutility of labor, χ , is chosen to normalize steady state labor input to unity. The exponent on capital services in the production function takes the standard value of $\alpha = 0.33$. The squared term in

the utilization adjustment cost function is $\delta_2 = 0.01$ and the investment adjustment cost parameter is $\kappa_I = 2$, both of which are conventional values.¹⁴ The elasticities of substitution for goods and labor, ϵ_p and ϵ_w , are set to be consistent with steady state markups of ten percent. The price rigidity parameters for both goods and labor, ϕ_p and ϕ_w , imply average durations between price and wage changes of one year. We assume no backward indexation of either prices or wages, so $\gamma_p = \gamma_w = 0$. The model is solved about a zero inflation steady state. The parameters of the underlying Taylor rule – ρ_r , ϕ_π , and ϕ_y – are standard. The steady state values of government spending and government debt are chosen to match a government spending share of output of 20 percent and a steady state debt-to-GDP ratio of 41 percent, the latter of which is the observed ratio of federal government liabilities to annualized nominal GDP in the fourth quarter of 2007. The AR parameters governing the productivity and government spending take on standard values.

The relatively non-standard parameters in our model relate to financial intermediaries, private and government bonds, the fraction of investment the wholesale firm must finance by issuing bonds, and the size of the central bank’s balance sheet. These are shown in the middle panel of Table 1. We follow Carlstrom, Fuerst and Paustian (2017) in setting κ to imply a ten year duration for both private and government bonds. We choose σ , the survival probability for intermediaries, to be 0.95, which is in-line with Gertler and Karadi (2011, 2013). We target a steady state excess return of private bonds over the deposit rate, $R_t^F - R_t^d$, of 300 basis points and that of government bonds over the deposit rate, $R_t^B - R_t^d$, of 100 basis points (both at an annualized frequency). These are chosen to roughly match observed spreads of Baa yields over the Federal Funds rate and the ten year Treasury yield over the Federal Funds rate.¹⁵ These targets imply a steady state value of $\theta = 0.58$ and a value of the parameter $\Delta = 1/3$. The value of X (the transfer of startup funds to new financial

¹⁴The functional form for the investment adjustment cost is standard: $S(I_t/I_{t-1}) = \frac{\kappa_I}{2} (I_t/I_{t-1} - 1)^2$.

¹⁵Note that excess returns and yields to maturity (defined and discussed below) are not necessarily the same outside of steady state, but are identical in the steady state. Over the period 1960-2008, the average Baa-FFR spread is 272 basis points and the average 10 Year-FFR spread is 86 basis points. Over a shorter sample, from 1984-2008, the average spreads are 352 and 142 basis points, respectively. Targeting spreads of 300 and 100 basis points, respectively, therefore represents a middle ground.

intermediaries) is chosen to be consistent with a steady state leverage ratio (ratio of assets to aggregate net worth) of four, which is similar to Gertler and Karadi (2011, 2013). The fraction of investment the wholesale firm must finance by issuing debt is $\psi = 0.81$. This is chosen to match the observed ratio of the value of outstanding private debt to annualized nominal GDP immediately prior to the Crisis of 1.68.¹⁶ The steady state (real) value of central bank private bond holdings, f_{cb} , is set to zero because the Fed only held Treasury debt at the end of 2007. The steady state central bank holding of government bonds, b_{cb} , is chosen to match the fraction of the Fed’s assets over annualized GDP immediately prior to the Great Recession of 6 percent.

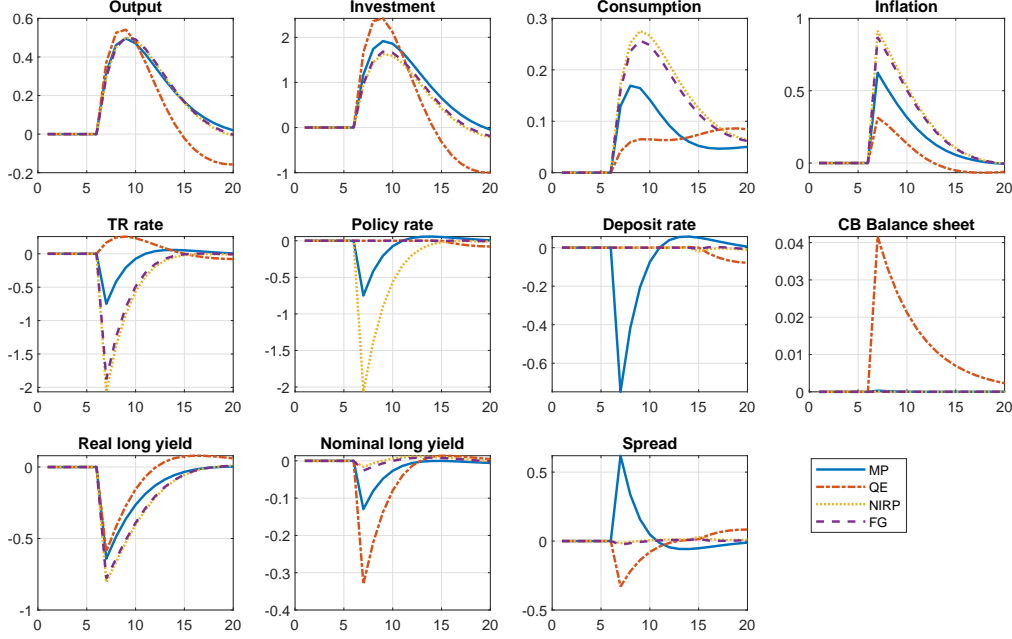
Although it can accommodate many more, our model features a parsimonious selection of exogenous shocks. The shock size are shown in the bottom panel of the table. Turning off all exogenous monetary policy shocks (both conventional and unconventional), we then pick the standard deviations of the innovations to the exogenous processes to loosely match a small set of selected moments from the data. The resulting standard deviations are $s_A = 0.0065$, $s_G = 0.01$, and $s_\theta = 0.04$. More details are provided in Appendix B.

4.2 Quantitative Analysis

In this section, we seek to quantify how much an unconventional policy tool must be moved so as to generate similar aggregate responses to a conventional policy shock. The comparison is depicted in Figure 1, which plots impulse responses of different macroeconomic aggregates to various policy interventions. In each case, the relevant exogenous shock hits in period 7. The solid blue lines depict responses to a conventional policy shock to the Taylor rule, (3.1). The size of the shock is -1 percent at an annualized percentage rate, which results in about a 0.75 percentage point decline in the policy rate on impact (the difference being accounted for by the endogenous reaction of the policy rate to inflation and output growth). The responses of aggregate variables to this shock are familiar. Output increases by about

¹⁶Outstanding private debt is measured by total credit to the private non-financial sector for the United States from FRED.

Figure 1: **Exogenous monetary policy**



Notes: Blue solid lines: a -1% shock to the annualized policy rate. Red dash-dotted lines: a QE shock, where the central bank increases its balance sheet by about 4 percent of steady state GDP by purchasing private debt, where $\rho_f = \rho_r$. Yellow dotted lines: NIRP with a shock of -2.4% to the annualized policy rate. Purple dashed lines: forward guidance with a shock of -2.2% to the desired policy interest rate. All these shocks hit the economy in period 7. For all but the blue solid lines, we generate a binding ZLB with a sequence of liquidity shocks of 1.5 standard deviations in each of the periods 1 through 6.

0.5 percent on impact and follows a hump-shaped pattern. The responses of investment and consumption are similar, though investment responds about four times as much as output and consumption about one third as much as output. Inflation rises by about 0.6 percentage points on impact and remains above trend for about three years.

Responses to unconventional policy interventions are depicted via colored dash lines (red for QE, purple for FG, and orange for NIRP). For these experiments, we assume that a sequence of liquidity shocks drives the economy to the ZLB; from the period the policy shock hits (period 7), the ZLB is expected to bind for nine quarters, or just over two years. This is in-line with the expected duration of the ZLB as estimated by Bauer and Rudebusch (2016) and Wu and Xia (2016).¹⁷ The quantitative easing experiment involves purchasing private

¹⁷In the case of the quantitative easing and forward guidance experiments, the policy rate (interest rate on

bonds (a purchase of government bonds has a qualitatively similar, albeit smaller, effect on macroeconomic aggregates, as we discuss further below). The magnitudes of the unconventional policy interventions are chosen to roughly match the impact change in output to the conventional policy shock. The autoregressive coefficient in the quantitative easing experiment, ρ_f , is set equal to the Taylor rule smoothing parameter, ρ_r , to facilitate comparison.

The unconventional shocks all have qualitatively similar effects on output compared to the conventional policy shock, although the output response to QE is less persistent than to the other interventions. The investment dynamics mirror the output response for all of the unconventional shocks. The consumption dynamics are similar for FG and NIRP compared to conventional policy, while the consumption response to a QE shock is quite muted in comparison. This is because QE has no effect on the deposit rate during the period of the ZLB, which is what is relevant in the household's Euler equation. All the shocks are inflationary, with the QE shock being the least so, while FG and NIRP are more inflationary compared to the conventional policy shock. To generate a similar output increase as a conventional policy shock, the QE experiment requires the central bank to increase the size of its balance sheet by about 4 percent relative to steady state output. Translating into the equations of our model,

$$s_f \varepsilon_{f,t} = \Psi_f s_r \varepsilon_{r,t} \quad (4.1)$$

where $\Psi_f = -7$ implies the shock size on $f_{cb,t}$ is seven times as large as the shock to the policy rate, and in the opposite direction. In other words, an easing policy could be implemented by lowering the policy rate or by increasing the size of the central bank's balance sheet. We will elaborate on this conversion rate later.

The NIRP and FG interventions are shocks to the Taylor rule of -2.4 and -2.2 percent, respectively, resulting in roughly a 2 percentage point decline in the Taylor rule rate on

reserves) and the deposit rate are both constrained by zero. For the NIRP experiment, as discussed above, only the deposit rate is constrained.

impact. This is more than twice as large as the change in the policy rate for the conventional policy shock. The main difference between FG and NIRP is that the policy rate does not react for the duration of the ZLB in the case of FG but falls immediately in the NIRP experiment. Forward guidance and NIRP shocks must be larger than conventional policy shocks to generate a similar change in output because these only affect the deposit rate in the future rather than in the present. The longer the expected duration of the ZLB, the larger must be the shock to FG or NIRP to generate a given output movement. FG is slightly more expansionary than a NIRP intervention of the same amount, which is why the NIRP shock is slightly larger than the FG shock. This is because a decline in the policy rate relative to the deposit rate reduces the profitability of intermediaries, as discussed in [Subsection 3.3](#). This in turn lowers net worth and exacerbates the constraint giving rise to equilibrium spreads. Because the model is solved about a steady state in which the size of the central bank's balance sheet is small, intermediaries do not hold large quantities of reserves and consequently the differences between FG and NIRP are small. This need not necessarily be the case, as we discuss further below.

The different policy shocks have comparable effects on output because they have similar effects on long term interest rates. To be precise, the (gross) yield to maturity, RL_t^F , on the private investment bond is implicitly defined via:¹⁸

$$Q_t = \frac{1}{RL_t^F} + \frac{\kappa}{(RL_t^F)^2} + \frac{\kappa^2}{(RL_t^F)^3} + \dots$$

Consequently,

$$RL_t^F = \frac{1}{Q_t} + \kappa \tag{4.2}$$

We further define the overall spread as the difference in the net long term private yield and

¹⁸The expression for the yield is similar for the government bond.

the net deposit rate:

$$\ln RL_t^F - \ln R_t^d \quad (4.3)$$

We define the credit spread as the difference in net yields on private and government bonds:

$$\ln RL_t^F - \ln RL_t^B \quad (4.4)$$

We define the real long term yield as:

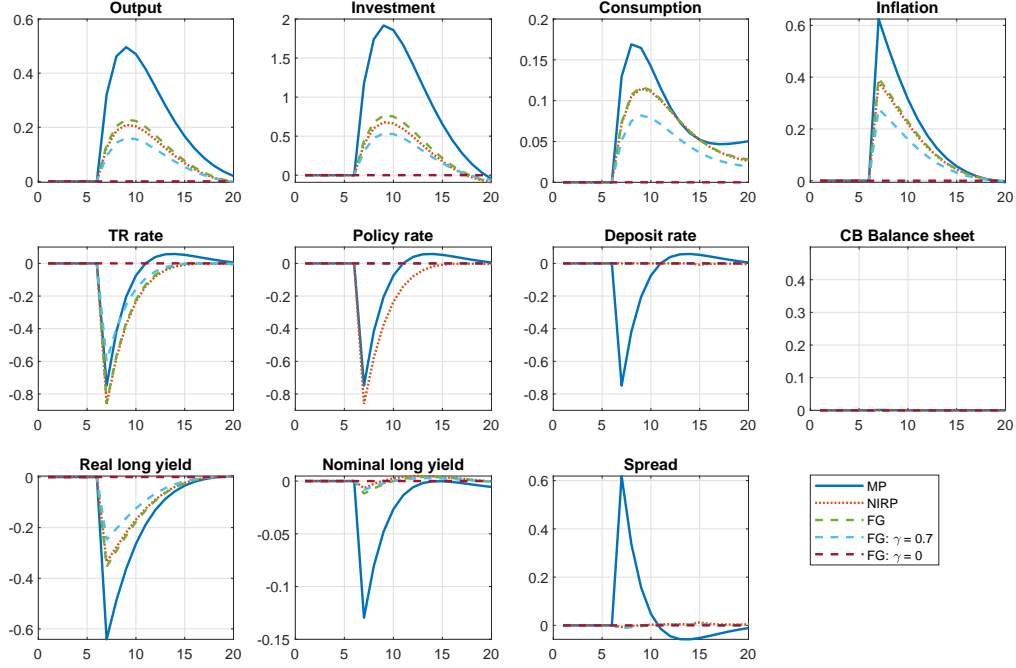
$$r_t^L = \ln RL_t^F - \mathbb{E}_t \ln \Pi_{t+1} \quad (4.5)$$

More precisely, what we call the real long term yield equals the slope of the yield curve plus the real deposit rate.¹⁹ This turns out to be the relevant metric for assessing monetary policy transmission in the model. One observes that the real long term yield declines by a similar amount both on impact and dynamically in response to both the conventional and unconventional policy interventions. This, in turn, results in very similar output dynamics to each of the policy interventions. Long term yields being the main driving force in the economy is consistent with Wu and Zhang (2017, 2019), who use the shadow rate of Wu and Xia (2016) as a summary statistic for the effects of unconventional policies.

Even though conventional and unconventional policy shocks have similar effects on the real long term yield, they have quite different effects on the overall spread between the long term private yield and the deposit rate. Conventional policy shocks indirectly affect long term rates by influencing short term rates. Since long term yields react less than short term rates, the overall spread increases. In contrast, unconventional shocks affect longer term rates without impacting the deposit rate in the short run. This results in the overall spread declining rather than rising.

¹⁹ In particular, the real net deposit rate is $\ln R_t^d - \mathbb{E}_t \ln \Pi_{t+1}$. Summing this and (4.3) gives (4.5).

Figure 2: **Exogenous NIRP vs. FG**



Notes: Blue lines: responses to the conventional monetary policy shock. Red dotted lines: response to a NIRP shock. Green dashed lines: fully credible forward guidance; light blue dashed lines: partially credible forward guidance with $\gamma = 0.7$; dark red dashed lines: non-credible forward guidance with $\gamma = 0$. All shocks are -1% annualized and hit in period 7. For all but blue solid lines, we generate a binding ZLB with a sequence of liquidity shocks of 1.5 standard deviations in each of the periods 1 - 6.

Figure 2 compares and contrasts forward guidance and NIRP across several different specifications. Blue lines depict responses to the conventional policy shock; these are identical to the solid blue lines in Figure 1 and are included to facilitate comparison. For the unconventional shocks, we generate the ZLB in the same manner as in Figure 1. Different than Figure 1, all shocks are the same magnitude of -1 percent annualized (i.e. the shock magnitudes are not adjusted so as to produce similar output responses). Dotted lines are responses to a NIRP shock. Dashed lines depict responses to forward guidance with green being fully credible ($\gamma = 1$), light blue partially credible ($\gamma = 0.7$), and dark red non-credible ($\gamma = 0$).

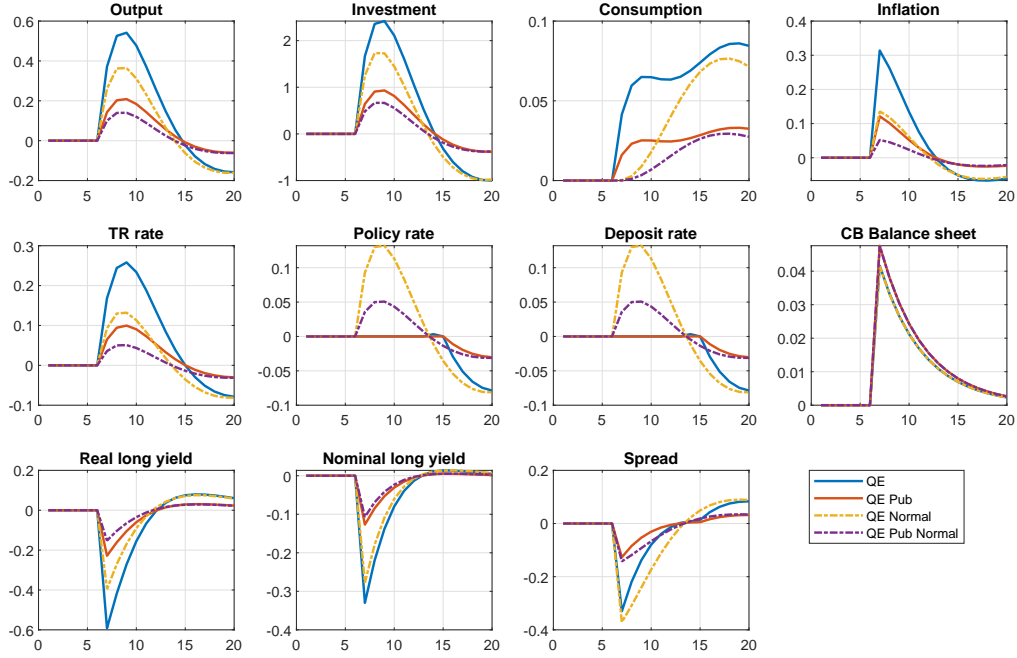
There are a couple of noteworthy results. First, although a fully credible forward guidance policy is more stimulative than NIRP, for a central bank with limited credibility, NIRP

is likely a more effective policy. In this sense, one can interpret NIRP in our model as commitment device a central bank can employ to carry out forward guidance. Pure forward guidance involves nothing more than a verbal promise of future low rates, which could be interpreted by the private sector as cheap talk, especially when the central bank has a low γ . NIRP, in contrast, involves an observable action in the present. Second, conditioning on equal sized shocks, the stimulative effect of a NIRP intervention is about half as large as a conventional policy shock. Put differently, to achieve a given stimulus on output the policy rate must be cut twice as much when the deposit rate is constrained by zero in comparison to normal times. In practice, this likely limits the usefulness of NIRP as an unconventional policy tool. A number of political and social constraints likely limit how negative interbank lending rates can go. In practice, no central bank has pushed policy rates deeply negative (e.g. the ECB’s lowest policy rate was only -40 basis points). Later, in [Section 6](#), we will use our model to shed some light on an effective lower bound for policy rates. We also return to NIRP and discuss its efficacy in relation to the total size of the central bank’s balance sheet.

[Figure 3](#) assesses different types of quantitative easing experiments. The blue solid lines are the benchmark responses, which correspond to a purchase of private debt when the policy and deposit rates are constrained by the ZLB. These responses are identical to the red dash-dotted responses in [Figure 1](#). The red solid lines are responses to a purchase of public debt (of the same magnitude as the benchmark) when short term rates are constrained. As in Gertler and Karadi (2013), private debt purchases are more stimulative than buying government debt. This result is essentially built into the model via our calibration of Δ , which governs the steady state government bond spread over the deposit rate. The dashed lines depict impulse responses to QE interventions during periods in which short term rates are unconstrained (i.e. “normal” times). QE is more stimulative when short term policy rates are constrained than when not, and private bond purchases remain more stimulative than government bond purchases.

In summary, various unconventional policy interventions can in principle have similar

Figure 3: Exogenous QE



Notes: Blue solid lines: purchasing private debt at the ZLB; red solid lines: purchasing public debt at the ZLB. Yellow dashed lines: purchasing private debt during normal times; purple dashed lines: purchasing public debt during normal times. The QE shocks hit the economy in period 7, and are the same across the four specifications. Where relevant, we generate a binding ZLB with a sequence of liquidity shocks of 1.5 standard deviations in each of the periods 1 - 6.

macroeconomic effects as a conventional cut in the short term policy rate. The requisite forward guidance and NIRP interventions are quite large, however, and the efficacy of FG depends crucially on a central bank's credibility. Further, implementing very large negative policy rates seems implausible in practice. For these reasons, QE-type interventions seem to be the best unconventional tool in practice. We therefore focus on QE as an endogenous policy tool to ameliorate the adverse consequences of a binding ZLB in the next section.

5 Endogenous Quantitative Easing and the Great Recession

The previous section established that exogenous changes in different unconventional policy tools can have similar economic effects as exogenous changes in short term policy rates. In practice, rather than exogenous shocks, effective monetary policy design entails the endogenous adjustment of policy to changing economic conditions. In this section, we study endogenous quantitative easing in response to other shocks. In particular, we wish to address the question of whether QE can serve as a close substitute for conventional policy. If so, how much QE is required? Is there an optimal reaction function for QE? How much might endogenous QE lessen the impact of a sequence of adverse shocks such as the ones that drove the US economy into the Great Recession?

So as to facilitate a comparison with conventional policy, we propose that the central bank's private bond holdings obey a Taylor type reaction function:²⁰

$$f_{cb,t} = (1 - \rho_f)f_{cb} + \rho_f f_{cb,t-1} + (1 - \rho_f)\Psi_f [\phi_\pi (\ln \Pi_t - \ln \Pi) + \phi_y (\ln Y_t - \ln Y_{t-1})] + s_f \varepsilon_{f,t} \quad (5.1)$$

The values of ϕ_π and ϕ_y are the same as in (3.1), and to further facilitate comparison we assume that $\rho_f = \rho_r$. The parameter Ψ_f measures the intensity of the reaction of bond-holdings, relative to the reaction of the desired policy rate, to deviations of inflation and output growth from target. For the exercises which follow, we assume that endogenous QE is only operative when the policy rate is constrained by the ZLB. That is, outside of the ZLB, $\Psi_f = 0$, and (5.1) collapses to (3.4). At the ZLB, we allow $\Psi_f \neq 0$. For the quantitative results which follow, we assume a value of $\Psi_f = -7$, which is the same conversion factor

²⁰We could similarly specify an endogenous reaction function for central bank holdings of government debt. All of the results which follow would be similar, though the value of Ψ_b required to mimic conventional policy would be larger.

we used above in (4.1) when studying exogenous QE. As we shall show, this conversion rate results in endogenous QE policy when the policy rate is constrained by zero closely mimicking conventional policy.

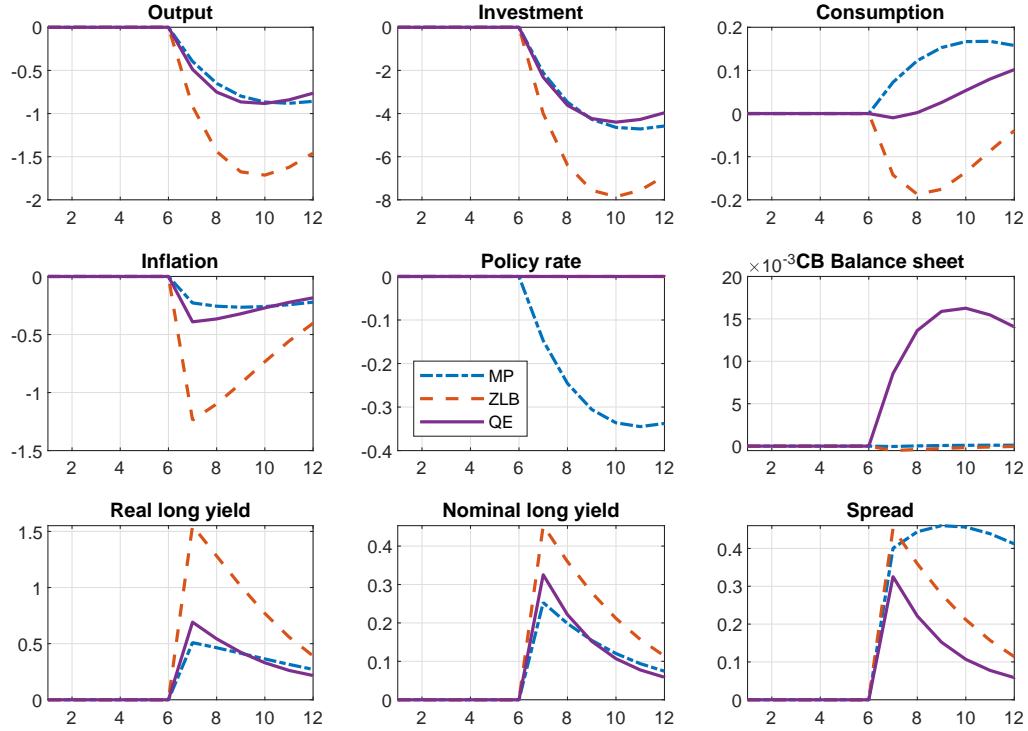
5.1 Endogenous Monetary Policy Responses to Economic Shocks

In this subsection, we study the conditional responses to different exogenous shocks when the policy rate is constrained by the ZLB. We find that endogenous QE effectively ameliorates the effects of the ZLB in response to each shock.

Liquidity shock Consider first a shock to θ_t . In the model, this measures the fraction of assets with which an intermediary may abscond in the event of bankruptcy; equivalently, $1 - \theta_t$ is the fraction of assets recoverable by creditors. An increase in θ_t therefore makes it more difficult for households to recover funds in the event of failure, which, other things being equal, makes them less willing to lend to intermediaries. We therefore refer to this shock as a liquidity shock – an increase in θ_t results in creditors pulling funds out of intermediaries, which in equilibrium results in a decline in the supply of credit and an increase in interest rate spreads. This represents a simple way for the model to capture the “run” dynamics on the shadow banking system during the recent crisis as described, for example, in Gorton and Metrick (2012).

Figure 4 plots impulse responses to a one standard deviation liquidity shock. The shock occurs in period seven. The procedure to generate the ZLB is exactly the same as in the exogenous policy experiments in Section 4. With conventional monetary policy (blue dash-dotted lines) unconstrained by the ZLB, a one standard deviation liquidity shock lowers output by about 1 percent at the peak. The shock results in a sharp increase in both the real and nominal long term yields. There is a consequent sharp decline in investment and inflation. The conventional policy response involves lowering short term rates, which partially offsets the decline in aggregate demand. The dashed red lines depict how the

Figure 4: **Liquidity shock and endogenous QE**



Notes: This figure plots impulse responses to a one standard deviation liquidity shock. The shock hits in period 7. Blue dash-dotted lines: conventional monetary policy without the ZLB. Red dashed lines: the ZLB with no unconventional monetary policy. Purple solid lines: the ZLB with endogenous QE. We create the ZLB with a sequence of liquidity shocks of 1.5 standard deviations from periods 1 - 6.

economy would react to the same shock if monetary policy were constrained by the ZLB and there is no unconventional policy response. With an inability to lower the policy rate, output falls substantially more in response to the shock – output declines by about 1.7 percent at peak. The shock is substantially more deflationary at the ZLB and the real and nominal long yields rise by significantly more in comparison to when policy is unconstrained.

The solid purple lines in Figure 4 plot responses when the policy rate is constrained by the ZLB but the central bank implements endogenous QE via (5.1). Endogenous QE is able to closely replicate the responses of output and investment to the shock under conventional policy. To replicate the responses under conventional policy (which entails an eventual 35 basis point cut in the policy rate), it is necessary for the central bank to increase the size of

its balance sheet by about 1.5 percent of steady state output at the maximum.

The underlying reason why endogenous QE is effective in response to the liquidity shock is similar to what is discussed above for exogenous policy shocks. Endogenous QE prevents long term interest rates from rising as much as they would when policy is constrained by the ZLB, thereby lowering the slope of the yield curve. This encourages investment and results in similar output dynamics in comparison to conventional policy. Consumption dynamics are different because it depends on the short term interest rate (evident in the upper right panel of the figure), but consumption contributes little to output fluctuations.

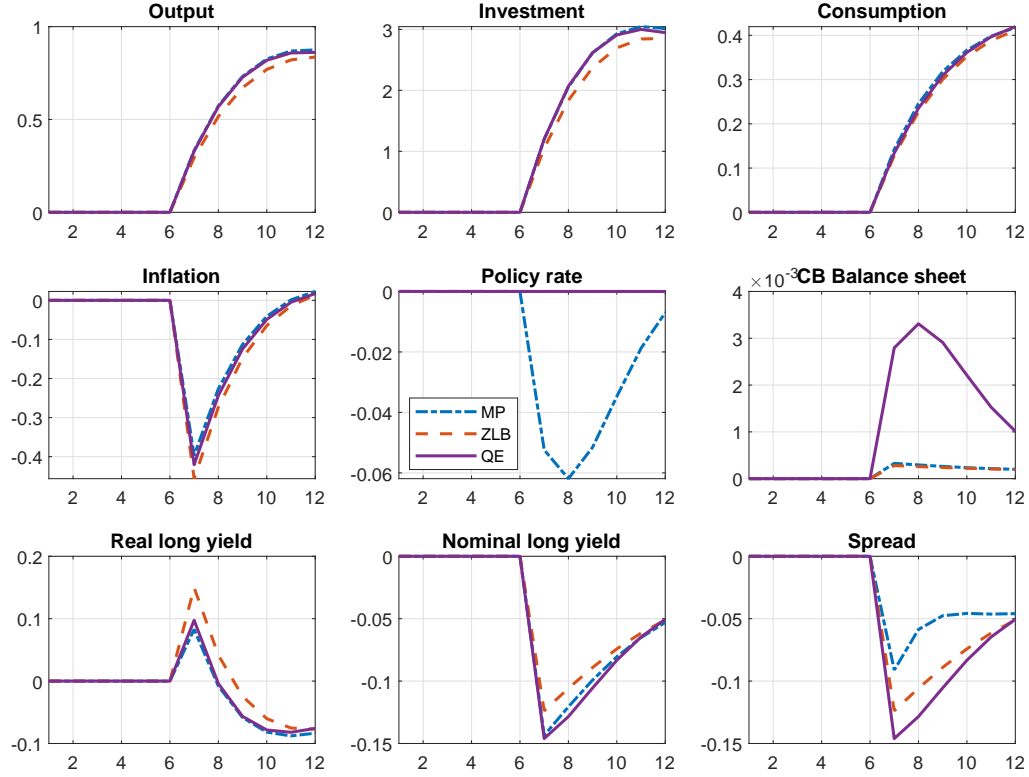
Technology shock Consider next the impulse responses to a one standard deviation technology shock. Responses under different policies scenarios are depicted in [Figure 5](#).

Under conventional policy, output, consumption, and investment all jump up on impact and continue to grow for a number of periods. Inflation and the policy rate both fall. When the policy rate is constrained and there is no unconventional policy response, output and other aggregates still rise on impact, but by less in comparison to normal times. The central bank would like to lower the policy rate to accommodate the increase in aggregate supply, but the inability to do so results in long term interest rates being higher than they otherwise would be, which dampens demand. When the central bank can endogenously adjust its balance sheet by purchasing bonds, in contrast, the responses of output, investment, and consumption virtually lie atop the responses under conventional policy. Note that this is with exactly the same value of Ψ_f as in [Figure 4](#). The requisite bond purchase to closely mimic the responses under conventional policy is not large – at the peak, the central bank needs to increase the size of its balance sheet by 0.3% of steady state output.

Government Spending Shock The other exogenous shock in our model is a shock to government spending. The impulse responses under different monetary policy scenarios are shown in [Figure 6](#).

In normal times, the government spending shock is expansionary for output, but con-

Figure 5: **Technology shock**

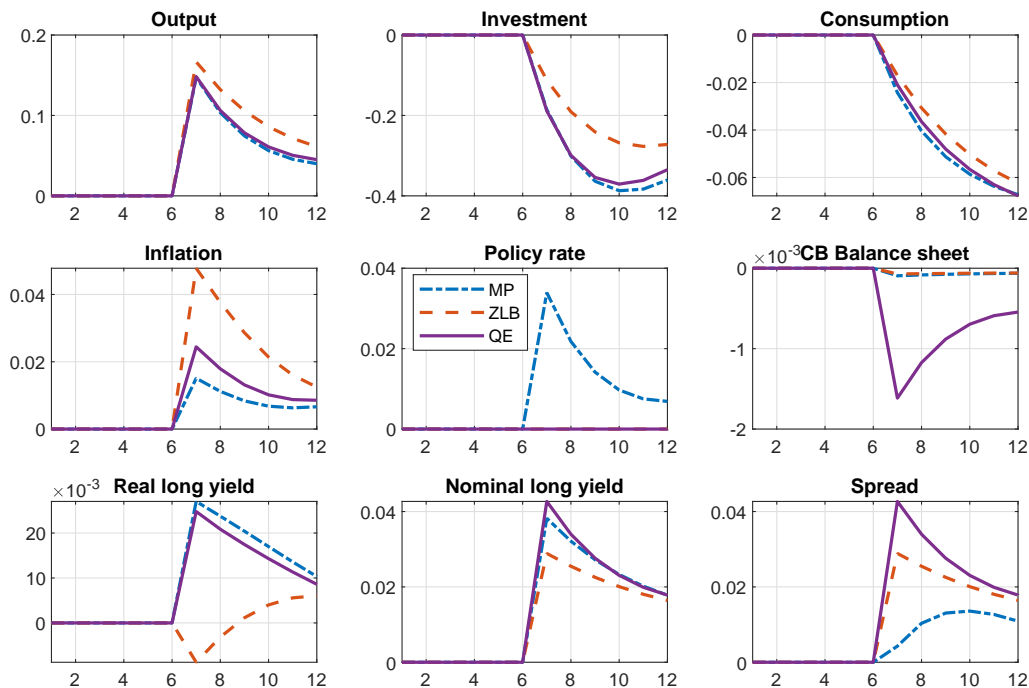


Notes: This figure plots impulse responses to a one standard deviation technology shock. The shock hits in period 7. Blue dash-dotted lines: conventional monetary policy without the ZLB. Red dashed lines: the ZLB with no unconventional monetary policy. Purple solid lines: the ZLB with endogenous QE. We create the ZLB with a sequence of liquidity shocks of 1.5 standard deviations from periods 1 - 6.

sumption and investment both decline (i.e. the multiplier is less than unity). Inflation and the policy rate both rise, with long term interest rates rising as well. When conventional policy is constrained by the ZLB and unconventional policies are unavailable, output expands by more compared to normal times. Consumption and investment both fall by less in comparison to normal times, although for this particular experiment both still decline (i.e. the multiplier remains below unity).²¹ This basic pattern is consistent with other papers that argue that the government spending multiplier can be substantially larger at the ZLB – see, e.g., Christiano, Eichenbaum and Rebelo (2011). With endogenous QE, in contrast, output

²¹For longer expected durations of the ZLB, or different parameterizations of the model, the multiplier can exceed unity (i.e. consumption and/or investment rise on impact). However, it is generally the case that our model is less sensitive to the ZLB than an otherwise similar model with frictionless financial markets.

Figure 6: **Government spending shock**



Notes: This figure plots impulse responses to a one standard deviation government spending shock. The shock hits in period 7. Blue dash-dotted lines: conventional monetary policy without the ZLB. Red dashed lines: the ZLB with no unconventional monetary policy. Purple solid lines: the ZLB with endogenous QE. We create the ZLB with a sequence of liquidity shocks of 1.5 standard deviations from periods 1 - 6.

and other variables react similarly at the ZLB compared to normal times. The underlying mechanism is the same as above conditional on the other shocks – by buying or selling bonds, the central bank can implement similar paths of long term interest rates as when it engages in conventional policy by adjusting short term rates.

5.2 Unconditional Analysis and the Great Recession

In this subsection we conduct a simulation analysis of the model. The economy starts in the non-stochastic steady state. Random technology, government spending, and liquidity shocks are then drawn for the subsequent 40 periods. To roughly mimic the experience of the Great Recession in the United States, we fix the liquidity shocks in periods 2 through 10 so as to drive the economy to the ZLB. This sequence of shocks causes the ZLB to bind for roughly

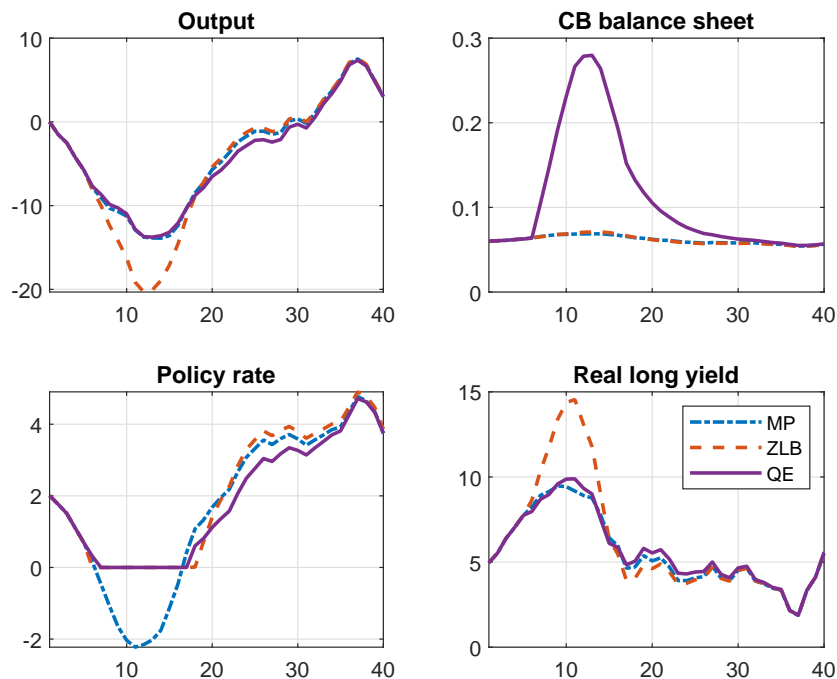
four years in expectation. Absent a ZLB constraint, it would result in the desired policy rate falling to two percentage points below zero.

Figure 7 plots simulated values of output (expressed in percentage deviations relative to steady state), the size of the central bank's balance sheet (relative to output), the policy rate (expressed at an annualized percentage rate), and the real long yield for one particular draw of random shocks. The dashed blue lines correspond to simulated values under conventional policy absent a ZLB constraint. The dashed red lines impose the ZLB and feature no unconventional policy reaction. The solid purple lines correspond to simulated paths assuming that endogenous QE becomes operative while the policy rate is constrained by zero. After the ZLB period is over, central bank bond holdings revert to obeying an exogenous AR(1) process.

Absent a ZLB constraint, for this particular draw of shocks output declines by more than 10 percent relative to steady state. It remains persistently low and does not return to steady state until about period 30, which is when the adverse sequence of initial large liquidity shocks that are employed to make the ZLB bind have mostly abated. Taking the ZLB constraint into account and assuming no unconventional policy response, in contrast, results in a much sharper output contraction. In particular, output falls a little more than 20 percent below steady state at peak. Put slightly differently, the ZLB causes output to decline by roughly twice as much in comparison to a version of the model ignoring the ZLB. The mechanism behind this amplification is the behavior of the real long yield (bottom right panel of the figure). The combination of deflation plus a fixed short term rate results in a very large spike in the real long yield.

It is apparent from Figure 7 that endogenous QE largely mitigates the effects of the binding ZLB. In fact, for the first 18 or so periods of the simulation, it is virtually impossible to distinguish between the simulated paths of output and the real long yield under endogenous QE relative to their simulated paths ignoring the ZLB. There is some difference evident after about period 20 of the simulation, where output under the ZLB with endogenous QE lies

Figure 7: **Simulation**



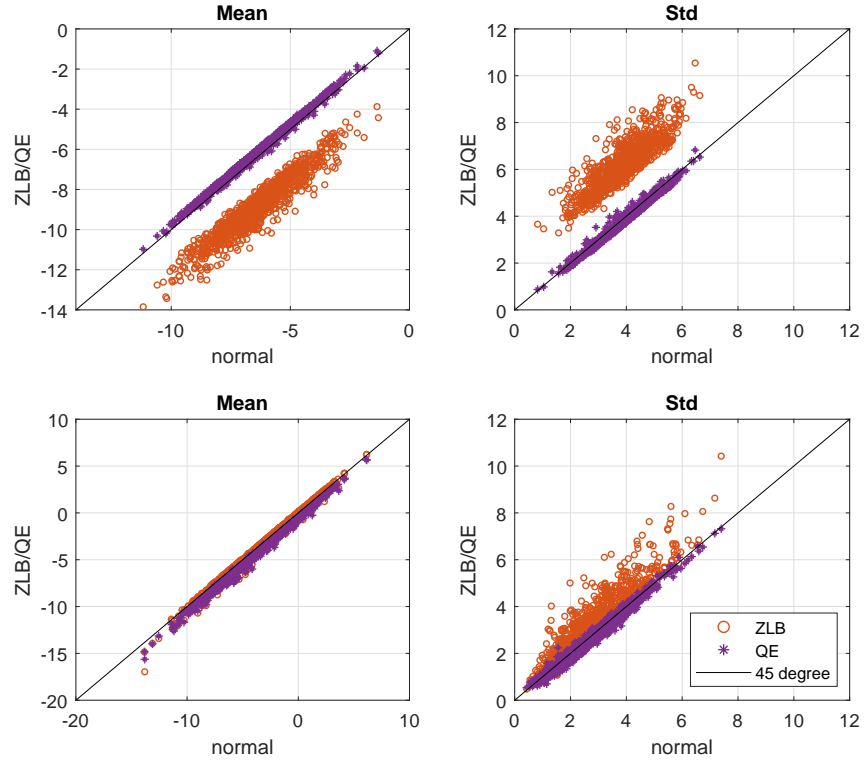
Notes: This figure plots simulated paths of different variables from one draw from our simulation. The economy begins in period 1 in steady state. From period 2 on, we draw random shocks to technology, government spending, and liquidity. We fix the size of the liquidity shock at 1.5 standard deviations from periods 2 - 10 to create the ZLB environment. Blue dash-dotted lines: normal times with conventional policy; red dashed lines: ZLB; solid purple lines: ZLB with endogenous QE. The size of the central bank's balance sheet is expressed relative to current period output.

slightly below the simulated path of output under conventional policy. This is driven by the unwinding of the balance sheet subsequent to the ZLB period and is discussed further in [Section 6](#).

The sequence of shocks used to generate [Figure 7](#) pushes the desired short term policy rate more than 2 percentage points below zero. This is roughly in-line with the estimated decline in the shadow rate series of Wu and Xia (2016). To generate the same response of output, a central bank following our endogenous QE rule purchases bonds in an amount that pushes the size of its balance sheet to about 25 percent of GDP. This is roughly the experience of the US during the Great Recession, where between QE1 and QE3 the Fed increased the size of its balance sheet to roughly 25 percent of GDP.

[Figure 7](#) focuses on one particular draw of shocks over a 40 quarter time horizon. In

Figure 8: **Simulation**



Notes: We generate 1000 different simulations of 40 periods each. For each simulation, from period 2 on we draw random shocks to technology, government spending and liquidity. We fix the size of the liquidity shock at 1.5 standard deviations from periods 2 - 10 to create the ZLB environment in each simulation. We compute the mean and standard deviation over periods 1 - 15 (top row) or over periods 16 - 40 (bottom row) for output for each of the 1000 different draws. We compare between ZLB (red circles) and endogenous QE (purple stars) on the vertical axes against normal times on the horizontal axes. Each of the 1000 circles or stars represent moments from each of the 1000 draws. Solid black line: 45 degree line. Units: percentage deviations from the steady state.

Figure 8 we graphically depict results from repeating this exercise 1000 different times. The figure shows scatter plots of moments of output when the short rate is subject to the ZLB (with and without endogenous QE) on the vertical axes against moments of output under conventional policy on the horizontal axes. The moments on which we focus are the mean and standard deviation of output, expressed in percentage deviations from steady state. Red circles correspond to moments with a binding ZLB but no endogenous QE, while purple stars are moments at the ZLB with endogenous QE. For ease of interpretation, we include 45 degree lines in each plot.

Focus first on the upper panel of [Figure 8](#), which is based on the first 15 periods of each of the 1000 simulations. The ZLB binds during these periods for all of the 1000 simulations. The red circles lie far below the 45 degree line in the case of the mean, and far above the 45 degree line in the case of the standard deviation. In other words, when the economy is at the ZLB and unconventional policies are unavailable, output is on average significantly lower and more volatile relative to a situation in which there is no ZLB. The purple stars, in contrast, align tightly with the 45 degree line for both the mean and standard deviation scatter plots. This suggests that endogenous QE can serve as a highly effective substitute for conventional policy during periods in which the ZLB binds.

Next, focus on the bottom panel of [Figure 8](#), which computes moments based on periods 16 to 40 of each simulation. The ZLB begins to lift in period 16. As a result, there is little observable difference between mean output with and without endogenous QE. Relative to the upper panel of the figure, the purple stars on average lie slightly below the 45 degree line in the case of mean output. This suggests that in the simulations where endogenous QE is undertaken during a ZLB period, the economy produces slightly less on average after the ZLB has ended. This is also evident in [Figure 7](#) for one draw. In terms of the volatility of output, the purple stars still align closely with the 45 degree line in the lower panel. There is, on average, slightly more volatility in periods 16-40 in the economy without endogenous QE, though it is markedly less than in periods 1-15 as shown in the upper panel. What is driving this is that, as noted above, while the ZLB begins to lift in period 16, in many of the simulations it lasts slightly longer.

[Table 2](#) summarizes these same moments by reporting averages across the 1000 simulations. The table corroborates the visual impressions from [Figure 8](#). Even without taking the ZLB into account, output would be substantially below the steady state given the large sequence of liquidity shocks fed into the model. The ZLB with no unconventional policy exacerbates this. Focusing on the first fifteen periods of the simulation, which correspond to the first rows of each block of the table, the level of output would be on average about

Table 2: Simulated Means and Volatilities

	Mean			Std		
	MP	ZLB	QE	MP	ZLB	QE
Output	-6.36	-8.84	-6.11	3.84	6.14	3.72
	-3.48	-3.69	-4.11	2.70	3.21	2.57
Real long yield	7.65	9.41	7.65	1.33	2.38	1.12
	5.81	5.75	6.04	0.86	0.97	0.85

Notes: We generate 1000 different simulations of 40 periods each. For each simulation, from period 2 on we draw random shocks to technology, government spending, and liquidity. We fix the size of the liquidity shock at 1.5 standard deviations from periods 2 - 10 to create the ZLB environment in each simulation. We compute the mean and standard deviation over periods 1 - 15 (first row) or over periods 16 - 40 (second row) for each draw. The table reports the average across 1000 draws. MP stands for conventional monetary policy with no ZLB constraint; ZLB refers to the zero lower bound; and QE denotes the ZLB on short term rates but with endogenous QE. Units are percentage deviations from the steady state for output and annualized percentage points for the real long yield.

2.5 percentage points lower and about 60 percent more volatile taking the ZLB into account versus not. The reason for the lower mean and enhanced volatility is the behavior of the real long yield, which is on average both higher and more volatile when taking the ZLB into account. In contrast, with endogenous QE, during the ZLB period the mean and volatility of output are roughly equal to what they would be absent a ZLB constraint.

The second rows in each panel of [Table 2](#) show moments calculated from periods 16 to 40 of each simulation. These moments are calculated mainly during periods after the ZLB has ended. Output would still be well below steady state in these periods even if there were no ZLB constraint. Interestingly, the version of the model taking the ZLB into account with no endogenous policy comes closer to matching the mean of output than does the version of the model with endogenous QE. This suggests a tradeoff – endogenous QE results in significantly better performance during a ZLB period at the expense of slightly worse outcomes once the ZLB has ended (at least in terms of mean output relative to steady state).²² As alluded to above, this is driven by unwinding of the balance sheet once the ZLB has ended. This unwinding exerts a mild contractionary effect on the path of output, and raises questions about the optimal speed of unwinding. We turn to such questions next.

²²In contrast, as is also evident in the bottom panel of [Figure 8](#), output remains slightly more volatile in periods 16-40 in the simulation where no endogenous QE is undertaken relative to the simulation with endogenous QE.

6 Issues Going Forward: QT, NIRP, and ELB

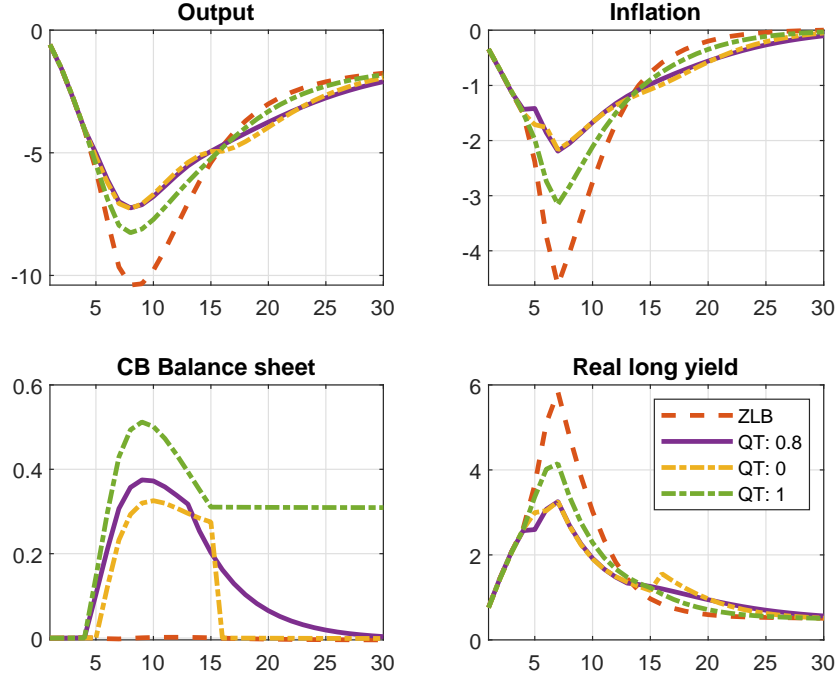
Combatting a sequence of negative demand shocks and minimizing the adverse consequences of the ZLB potentially requires a large expansion in the size of a central bank’s balance sheet. This is true both in our model as well as in practice for the myriad countries that engaged in aggressive bond purchasing programs in the wake of the Great Recession. In this section, we discuss three broad issues related to the post-QE size of central bank balance sheets going forward. First, an imminent issue facing policymakers around the globe is how to unwind their balance sheets while minimizing potential disruptions to the economy. We discuss the unwinding process, also known as quantitative tightening (QT), in [Subsection 6.1](#). Second, there exists an interaction between the size of a central bank’s balance sheet and the efficacy of NIRP. This is discussed in [Subsection 6.2](#). Third, the size of the balance sheet constrains how negative policy rates can go. We make a first attempt at endogenizing the effective lower bound (ELB) in our model in [Subsection 3.3](#).

6.1 Quantitative Tightening

In the previous section, we showed that endogenous QE can serve as an effective substitute for conventional monetary policy during ZLB periods. At the same time, we showed that the economy tends to perform somewhat worse *after* the ZLB period has ended when endogenous QE is undertaken as an antidote to the ZLB. What drives this is the unwinding of the central bank’s balance sheet as it transitions away from bond purchases.

We wish to study the consequences of different paces of balance sheet normalization. We design our QT experiment as an extension on the impulse responses depicted in [Figure 4](#). Instead of showing the responses of variables to the incremental one standard deviation θ shock, we plot the responses based on all the shocks, including those that we use to create the ZLB environment. In this way, we are able to generate a realistic depiction of the post-Great Recession experience. This is shown in [Figure 9](#).

Figure 9: **Quantitative tightening**



Notes: Red dashed lines: ZLB without unconventional monetary policy. Purple solid lines: QT according to (3.4) with $\rho_f = 0.8$; orange dash-dotted lines: QT according to (3.4) with $\rho_f = 0$; green dash-dotted lines: QT according to (3.4) with $\rho_f = 0.1$.

Red dashed and purple solid lines are the same as those in Figure 4, except that the sequence of liquidity shocks used to generate the ZLB are no longer differenced out.²³ The red dashed lines plot simulated paths of variables under the ZLB constraint with no endogenous QE. The purple solid lines show paths of variables under the ZLB constraint with endogenous QE, followed by QT, with bond-holdings reverting to the AR(1) process as specified in (3.4) with the same autoregressive parameter of $\rho_f = 0.8$. The dash-dotted lines also implement QT after QE, but with different values of ρ_f (i.e. different speeds of unwinding). Paths under immediate unwinding, with $\rho_f = 0$, are shown in orange, whereas the green dash-dotted lines represent simulated paths with $\rho_f = 1$. This case corresponds to the central bank carrying

²³The impulse responses shown in Figure 4 are constructed by considering two simulations. In both simulations there are a sequence of liquidity shocks in periods 1-6. These are used to make the ZLB bind. In the second simulation, there is an additional, one standard deviation liquidity shock that hits the economy in period 7. The plotted impulse responses are the differences between these two simulations. In Figure 9, we instead plot the second simulation without differencing out the first simulation.

forward the large balance sheet accumulated during QE with no unwinding.

Interestingly, the anticipated speed of unwinding after the ZLB ends has an important impact for how the economy fares during the ZLB. Relative to our baseline case of $\rho_f = 0.8$, the orange lines with $\rho_f = 0$ show a somewhat smaller decrease in output during the ZLB period, followed by a lower trajectory of output after the ZLB has ended. In contrast, no normalization with $\rho_f = 1$ results in output declining more during the ZLB period, though the recovery after the ZLB has ended is somewhat stronger. The tradeoff seems clear – rapid QT results in better outcomes during the early stages of the recession at the expense of worse performance after the ZLB has ended, whereas no unwinding exacerbates the recession. $\rho_f = 0.8$, in contrast, represents a desirable middle ground with a mild recession and a smooth transition away from bond purchases.

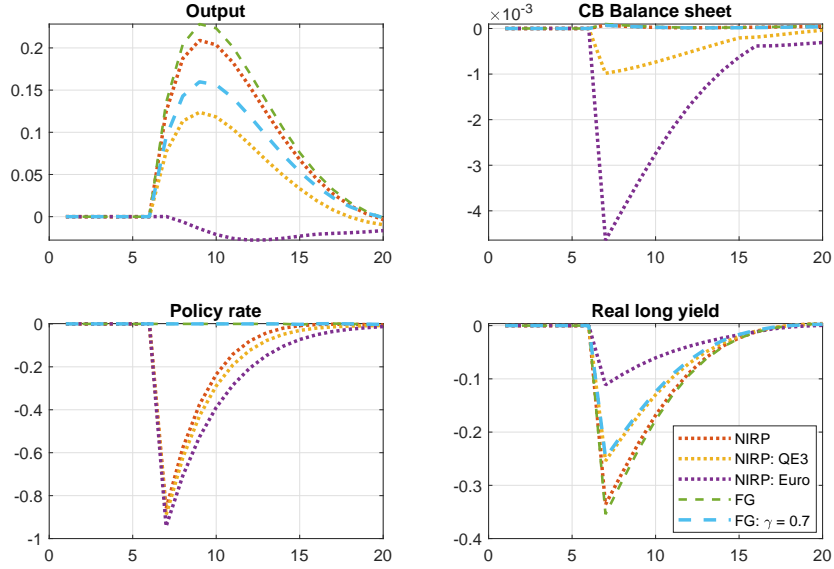
6.2 Interaction Between QE and NIRP

Moving forward, it is also important to understand interactions between different types of unconventional policy interventions. The size of a central bank’s balance is relevant for the efficacy of NIRP. Negative policy rates function as a type of tax on intermediaries; the more reserves intermediaries hold, the more burdensome is the tax. This tends to undo the stimulative signaling feature of NIRP.

Dotted lines in [Figure 10](#) depict impulse responses to 100 basis point NIRP shocks conditioning on different sizes of the central bank’s balance sheet. Red dotted lines are responses when the central bank’s balance sheet is about 6 percent of steady state GDP; this is identical to [Figure 2](#). The orange and purple dotted lines condition on balance sheet sizes calibrated to the size of the Fed’s post-QE3 balance sheet and the ECB’s balance sheet as of the end of 2018. For comparison, dashed lines plot responses to forward guidance shocks assuming varying degrees of credibility. The green and light blue responses are identical to [Figure 2](#).

As the central bank’s balance sheet becomes large, NIRP becomes less stimulative. For example, conditioning on a balance sheet calibrated to post-QE3 in the US, a NIRP shock

Figure 10: **Exogenous NIRP and balance sheet**

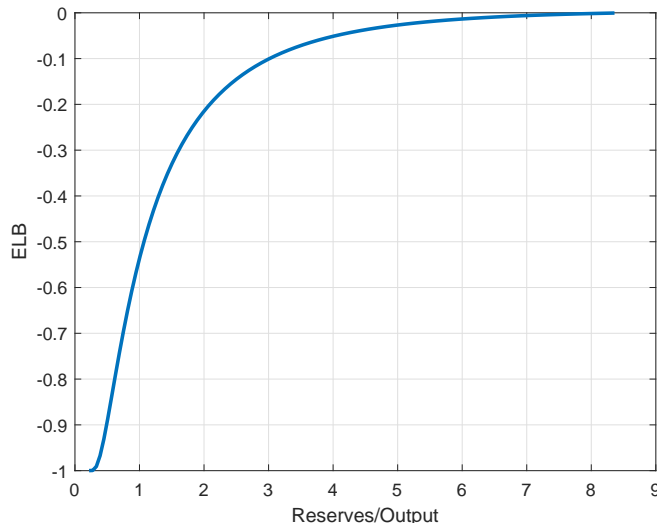


Notes: Dotted lines are plots of responses to exogenous NIRP shocks conditioning on different sizes of the central bank's balance sheet. Red dotted lines: NIRP with the steady state central bank's balance sheet of 6% of the GDP (calibrated to pre-QE); orange dotted lines: NIRP with the steady state balance sheet of 25% GDP (calibrated to post QE3); purple dotted lines: NIRP with the steady state balance sheet of 38% of GDP (calibrated to the Euro area to 2018Q4). Green dashed lines: fully credible forward guidance; light blue dashed lines: partially credible forward guidance with $\gamma = 0.7$. We generate a binding ZLB with a sequence of liquidity shocks of 1.5 standard deviations in each of the periods 1 - 6.

becomes half as stimulative as the benchmark in red, and less stimulative than non-credible forward guidance ($\gamma = 0.7$). When the balance sheet is as large as in the Euro area, a cut in the policy rate into negative territory is actually mildly contractionary. This is because the contractionary banking channel outweighs the expansionary forward guidance channel of NIRP.

For countries such as the US that have neither experimented with NIRP nor amassed a particularly large balance sheet, this interaction between the size of a central bank's balance sheet and the efficacy of NIRP may seem academic. But this is not so for other countries, such as Japan and the Euro area, that have accumulated much larger balance sheets and have already pushed policy rates into negative territory. Another interesting implication from our model relates to the timing of unconventional policy interventions. In practice, economies that eventually turned to negative rates all engaged in heavy QE programs prior

Figure 11: **ELB**



Notes: This figure plots the minimum steady state value of $R^{re} - 1$ (expressed at an annualized rate) as a function of the steady state size of the central bank's balance sheet (relative to output). For this figure, we set $\beta = 1$ so that the deposit rate, R^d , is fixed at its lower bound of one (or zero in net terms).

to doing so. From the perspective of our model, it would have instead been more effective to first push policy rates into negative territory and then to engage in QE programs.

6.3 Implications of a Large Balance Sheet for the ELB

In practice, no central bank has experimented with moving policy rates into deep negative territory. An important question is how negative rates could conceivably go without adverse consequences.

Though we have not imposed it thus far, in principle there exists a lower bound on the interest rate on reserves in our model. In particular, if $\theta_t \phi_t$ were to fall below one (i.e. net worth is more valuable in a household than in an intermediary), intermediaries might choose to voluntarily exit rather than continue. How low R_t^{re} can go without violating $\theta_t \phi_t \geq 1$ depends on (i) the magnitude of excess returns on private and government bonds and (ii) the total value of reserves issued by the central bank. The lower are excess returns, or the larger is the volume of reserves, the higher will be the ELB.

Figure 11 plots the minimum value of steady state R^{re} consistent with $\theta \phi \geq 1$ as a

function of the quantity of reserves (as a fraction of output). When there are no (or few) reserves, there is essentially no limit on how negative R^{re} can go. As there are more reserves in the system, and reserves constitute a higher fraction of intermediary assets, the effective lower bound becomes tighter. The intuition is straightforward. Negative policy rates are essentially a tax on the holders of reserves. The more reserves there are, the more punitive is this tax, and at some point intermediaries might find it undesirable to continue.

Caution is in order when interpreting the quantitative magnitudes in [Figure 11](#) seriously. Our model abstracts from many real-world features that might further inhibit a central bank’s ability to push policy rates into deep negative territory. Nevertheless, we think that the qualitative relationship between the size of the central bank’s balance sheet and the ELB is both instructive and intuitive. It is also potentially increasingly relevant for countries such as Japan and the Euro area that have now amassed very large balance sheets. While more work remains to be done, we provide a useful first benchmark to think more seriously about this issue.

7 Conclusion

In this paper, we developed a quantitative DSGE framework to systematically evaluate three different types of unconventional policy tools – quantitative easing (QE), forward guidance (FG), and negative interest rate policy (NIRP) – as well as their interactions. This represents an important step forward in the literature, which has to date tended to focus on each policy tool in isolation. We emphasize two distinct channels by which NIRP can transmit to the economy: a forward guidance channel and a banking channel. We introduce a novel way to model forward guidance, where our approach ameliorates the forward guidance “puzzle.”

We find that unconventional policies can all stimulate output as much as a conventional policy shock, though we argue that quantitative easing is likely to be the most desirable among unconventional options. We specify an endogenous feedback rule for QE and show

that it can effectively neutralize the adverse consequences of a binding zero lower bound on the short term policy rate. In an application meant to mimic the experience of the US during the Great Recession, we show that our endogenous QE rule, which expands the central bank's balance sheet to 25% of GDP, provides stimulus roughly equivalent to moving the policy rate 2 percentage points below zero. This is in-line with empirical estimates of the shadow rate from Wu and Xia (2016).

Though endogenous QE is an effective substitute for conventional policy when faced with a binding ZLB, its use is not without cost or potential complication. When unwinding the balance sheet, a smooth balance sheet normalization is preferred. On the contrary, carrying a large balance sheet forward would cause a deeper recession and force the central bank to expand their balance sheet even more at the peak. Moreover, a larger balance sheet undermines the efficacy of NIRP or even makes it contractionary. We provide a first attempt at endogenizing an effective lower bound (ELB) for policy rates, which also depends on the overall size of a central bank's balance sheet.

References

- Abo-Zaid, Salem and Julio Garín**, “Optimal Monetary Policy and Imperfect Financial Markets: A Case for Negative Nominal Interest Rates?,” *Economic Inquiry*, 2016, 54 (1), 215–228.
- Bauer, Michael D. and Glenn D. Rudebusch**, “The Signaling Channel for Federal Reserve Bond Purchases,” *International Journal of Central Banking*, 2014, 10 (3), s114–s133.
- and –, “Monetary Policy Expectations at the Zero Lower Bound,” *Journal of Money, Credit and Banking*, 2016, 48 (7), 1439–1465.
- Campbell, Jeffrey R., Charles L. Evans, Jonas D.M. Fisher, and Alejandro Justiniano**, “Macroeconomic Effects of Federal Reserve Forward Guidance,” *Brookings Papers on Economic Activity*, 2012, 43 (1 (Spring), 1–80.
- Carlstrom, Charles T., Timothy S. Fuerst, and Matthias Paustian**, “Targeting Long Rates in a Model with Segmented Markets,” *American Economic Journal: Macroeconomics*, January 2017, 9 (1), 205–42.
- Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans**, “Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy,” *Journal of Political Economy*, 2005, 113, 1–45.
- , – , and **Sergio Rebelo**, “When is the Government Spending Multiplier Large?,” *Journal of Political Economy*, 2011, 119 (1), 78–121.
- Clarida, Richard H.**, “What Has – and Has Not – Been Learned about Monetary Policy in a Low-Inflation Environment? A Review of the 2000s,” *Journal of Money, Credit and Banking*, 2012, 44 (s1), 123–140.

- Coibion, Olivier and Yuriy Gorodnichenko**, “Monetary Policy, Trend Inflation, and the Great Moderation: An Alternative Interpretation,” *American Economic Review*, February 2011, *101* (1), 341–70.
- Debortoli, Davide, Jordi Galí, and Luca Gambetti**, “On the Empirical (Ir)Relevance of the Zero Lower Bound Constraint,” 2016. Universitat Pompeu Fabra working paper.
- Del Negro, Marco, Domenico Giannone, Marc P. Giannoni, and Andrea Tambalotti**, “Safety, Liquidity, and the Natural Rate of Interest,” *Brookings Papers on Economic Activity*, 2017, *Spring*, 235 – 294.
- , **Marc P. Giannoni, and Chiristina Patterson**, “The Forward Guidance Puzzle,” 2015. Federal Reserve Bank of New York Staff Reports, No. 574.
- Eggertsson, Gauti, Ragnar Juelsrud, Lawrence H. Summers, and Ella Getz Wold**, “Negative Nominal Interest Rates and the Bank Lending Channel,” 2019. NBER Working Paper No. 25416.
- Gagnon, Joseph, Matthew Raskin, Julie Remache, and Brian Sack**, “The Financial Market Effects of the Federal Reserve’s Large-Scale Asset Purchases,” *International Journal of Central Banking*, 2011, *7*, 3–43.
- Garín, Julio, Robert Lester, and Eric Sims**, “Are Supply Shocks Contractionary at the ZLB? Evidence from Utilization-Adjusted TFP Data,” *The Review of Economics and Statistics*, 2019, *101* (1), 160–175.
- Gertler, Mark and Peter Karadi**, “A Model of Unconventional Monetary Policy,” *Journal of Monetary Economics*, 2011, pp. 17–34.
- **and** – , “QE 1 vs. 2 vs. 3 . . . : A Framework for Analyzing Large-Scale Asset Purchases as a Monetary Policy Tool,” *International Journal of Central Banking*, 2013, *9*(S1), 5–53.

- Gorton, Gary and Andrew Metrick**, “Securitized Banking and the Run on Repo,” *Journal of Financial Economics*, 2012, *104* (3), 425–451.
- Greenwood, Robin and Dimitri Vayanos**, “Bond Supply and Excess Bond Returns,” *The Review of Financial Studies*, 01 2014, *27* (3), 663–713.
- Guerrieri, Luca and Matteo Iacoviello**, “OccBin: A Toolkit for Solving Dynamic Models with Occasionally Binding Constraints Easily,” *Journal of Monetary Economics*, 2015, *70* (C), 22–38.
- Hamilton, James D. and Jing Cynthia Wu**, “The Effectiveness of Alternative Monetary Policy Tools in a Zero Lower Bound Environment,” *Journal of Money, Credit, and Banking*, 2012, *44* (s1), 3–46.
- Kortela, Tomi**, “A Shadow Rate Model with Time-Varying Lower Bound of Interest Rates,” 2016. Bank of Finland Research Discussion Paper.
- Krishnamurthy, Arvind and Annette Vissing-Jorgensen**, “The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy,” *Brookings Papers on Economic Activity*, 2011, *2*, 215–265.
- Kuttner, Kenneth N.**, “Outside the Box: Unconventional Monetary Policy in the Great Recession and Beyond,” *Journal of Economic Perspectives*, November 2018, *32* (4), 121–46.
- Laubach, Thomas and John C. Williams**, “Measuring the Natural Rate of Interest,” *The Review of Economics and Statistics*, 2003, *85* (4), 1063–1070.
- Lemke, Wolfgang and Andreea L. Vladu**, “Below the Zero Lower Bound: A Shadow-Rate Term Structure Model for the Euro Area,” 2016. Deutsche Bundesbank Discussion Paper.

- Mau, Ronald**, “Purchases and Sales of Long Bonds as a Monetary Policy Instrument,” 2019. Working Paper.
- McKay, Alisdair, Emi Nakamura, and Jon Steinsson**, “The Power of Forward Guidance Revisited,” *American Economic Review*, October 2016, *106* (10), 3133–58.
- Mouabbi, Sarah and Jean-Guillaume Sahuc**, “Evaluating the Macroeconomic Effects of the ECB’s Unconventional Monetary Policies,” 2017. Working paper.
- Orphanides, Athanasios and John C. Williams**, “Monetary Policy with Imperfect Knowledge,” *Journal of the European Economic Association*, 2006, *4* (23), 366–375.
- Smets, Frank and Rafael Wouters**, “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach,” *American Economic Review*, 2007, *97* (3), 586–606.
- Swanson, Eric T.**, “The Federal Reserve is Not Very Constrained by the Lower Bound on Nominal Interest Rates,” 2018. Working Paper.
- , “Measuring the Effects of Federal Reserve Forward Guidance and Asset Purchases on Financial Markets,” 2018. working paper, University of California, Irvine.
- **and John C. Williams**, “Measuring the Effect of the Zero Lower Bound on Medium- and Longer-Term Interest Rates,” *American Economic Review*, 2014, *104* (10), owner = Tony,), 3154–3185.
- Taylor, John B.**, “Discretion versus Policy Rules in Practice,” *Carnegie-Rochester Conference Series on Public Policy*, 1993, *39*, 195–214.
- Ulate, Mauricio**, “Going Negative at the Zero Lower Bound: The Effects of Negative Nominal Interest Rates,” 2018. Working Paper.
- Vayanos, Dimitri and Jean-Luc Vila**, “A Preferred-Habitat Model of the Term Structure of Interest Rates,” 2009. Working paper, London School of Economics.

- Walker, Ray**, “Monetary Policy and the Limits to Arbitrage: Insights from a New Keynesian Preferred Habitat Model,” 2019. Working Paper.
- Walsh, Carl**, “Speed Limit Policies: The Output Gap and Optimal Monetary Policy,” *American Economic Review*, March 2003, *93* (1), 265–278.
- Woodford, Michael**, “Fiscal Requirements for Price Stability,” *Journal of Money, Credit and Banking*, August 2001, *33* (3), 669–728.
- Wu, Jing Cynthia and Fan Dora Xia**, “Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound,” *Journal of Money, Credit, and Banking*, 2016, *48* (2-3), 253–291.
- **and Ji Zhang**, “A Shadow Rate New Keynesian Model,” 2017. NBER Working Paper No. 22856.
- **and —**, “Global Effective Lower Bound and Unconventional Monetary Policy,” *Journal of International Economics*, 2019, *118*, 200–216.

A Model Derivations

A.1 The Financial Intermediary's Problem

Write the net worth accumulation equation, (2.5), in real terms by dividing both sides by the aggregate price level, noting that $\Pi_t = P_t/P_{t-1}$:

$$\begin{aligned} \Pi_t n_{i,t} = & (R_t^F - R_{t-1}^d) Q_{t-1} f_{i,t-1} + (R_t^B - R_{t-1}^d) Q_{B,t-1} b_{i,t-1} \\ & + (R_{t-1}^{re} - R_{t-1}^d) re_{i,t-1} + R_{t-1}^d n_{i,t-1} \end{aligned} \quad (\text{A.1})$$

which implies

$$\begin{aligned} \Lambda_{t,t+1} \Omega_{t+1} n_{i,t+1} = \\ \Lambda_{t,t+1} \Omega_{t+1} \Pi_{t+1}^{-1} [(R_{t+1}^F - R_t^d) Q_t f_{i,t} + (R_{t+1}^B - R_t^d) Q_{B,t} b_{i,t} + (R_t^{re} - R_t^d) re_{i,t} + R_t^d n_{i,t}] \end{aligned} \quad (\text{A.2})$$

An intermediary's value function, (2.8), can be written as

$$V_{i,t} = \max (1 - \sigma) \mathbb{E}_t \Lambda_{t,t+1} n_{i,t+1} + \sigma \mathbb{E}_t \Lambda_{t,t+1} V_{i,t+1} \quad (\text{A.3})$$

$$\begin{aligned} = \max (1 - \sigma) \mathbb{E}_t \Lambda_{t,t+1} \left[(R_{t+1}^F - R_t^d) \Pi_{t+1}^{-1} Q_t f_{i,t} + (R_{t+1}^B - R_t^d) \Pi_{t+1}^{-1} Q_{B,t} b_{i,t} \right. \\ \left. + (R_t^{re} - R_t^d) \Pi_{t+1}^{-1} re_{i,t} + R_t^d \Pi_{t+1}^{-1} n_{i,t} \right] + \sigma \mathbb{E}_t \Lambda_{t,t+1} V_{i,t+1} \end{aligned} \quad (\text{A.4})$$

A Lagrangian with the constraints (2.9) - (2.10) is

$$\begin{aligned} \mathbb{L}_{i,t} = \max(1 + \lambda_t) \mathbb{E}_t \left\{ (1 - \sigma) \Lambda_{t,t+1} \left[(R_{t+1}^F - R_t^d) \Pi_{t+1}^{-1} Q_t f_{i,t} + (R_{t+1}^B - R_t^d) \Pi_{t+1}^{-1} Q_{B,t} b_{i,t} + \right. \right. \\ \left. \left. (R_t^{re} - R_t^d) \Pi_{t+1}^{-1} re_{i,t} + R_t^d \Pi_{t+1}^{-1} n_{i,t} \right] + \sigma \Lambda_{t,t+1} V_{i,t+1} \right\} - \lambda_t \theta_t (Q_t f_{i,t} + \Delta Q_{B,t} b_{i,t}) + \omega_t (re_{i,t} - \bar{r} e_{i,t}) \end{aligned}$$

where λ_t and ω_t are Lagrangian multipliers. The first order necessary conditions are given in (2.11) - (2.13).

To derive expressions for the auxiliary variables Ω_t and ϕ_t in (2.14) - (2.15), we guess that the value function is linear in net worth as in (2.16). Combining this guess with a binding constraint on (2.9) yields (2.17). (2.17), (A.2), and the first order conditions yield:

$$\begin{aligned} & \mathbb{E}_t [\Lambda_{t,t+1} \Omega_{t+1} n_{i,t+1}] \\ = & \mathbb{E}_t [\Lambda_{t,t+1} \Omega_{t+1} \Pi_{t+1}^{-1} (R_{t+1}^F - R_t^d) (Q_t f_{i,t} + \Delta Q_{B,t} b_{i,t}) \\ & + \mathbb{E}_t [\Lambda_{t,t+1} \Omega_{t+1} \Pi_{t+1}^{-1} (R_t^{re} - R_t^d) re_{i,t} + \mathbb{E}_t [\Lambda_{t,t+1} \Omega_{t+1} \Pi_{t+1}^{-1} R_t^d n_{i,t}]] \\ = & \frac{\lambda_t}{1 + \lambda_t} \theta_t \phi_t n_{i,t} - \frac{\omega_t}{1 + \lambda_t} re_{i,t} + \mathbb{E}_t [\Lambda_{t,t+1} \Omega_{t+1} \Pi_{t+1}^{-1} R_t^d n_{i,t}] \end{aligned}$$

Rewrite the value function in (A.3) using (2.14) and (2.16):

$$\begin{aligned} \theta_t \phi_t n_{i,t} &= \max \mathbb{E}_t [\Lambda_{t,t+1} n_{i,t+1} \Omega_{t+1}] \\ &= \frac{\lambda_t}{1 + \lambda_t} \theta_t \phi_t n_{i,t} - \frac{\omega_t}{1 + \lambda_t} re_{i,t} + \mathbb{E}_t [\Lambda_{t,t+1} \Omega_{t+1} \Pi_{t+1}^{-1} R_t^d n_{i,t}] \end{aligned}$$

Assume

$$\frac{re_{i,t}}{d_{i,t}} = \frac{re_t}{d_t}$$

(2.10) implies $\frac{re_t}{d_t} = \varsigma_t$ when the constraint binds, and aggregation $\sum_i re_{i,t} = \sum_i \frac{re_t}{d_t} d_{i,t}$ holds automatically. This leads to

$$\theta_t \phi_t = \frac{\lambda_t}{1 + \lambda_t} \theta_t \phi_t - \frac{\omega_t}{1 + \lambda_t} \frac{re_t}{n_t} + \mathbb{E}_t[\Lambda_{t,t+1} \Omega_{t+1} \Pi_{t+1}^{-1}] R_t^d$$

Simplify,

$$\theta_t \phi_t = (1 + \lambda_t) \mathbb{E}_t[\Lambda_{t,t+1} \Omega_{t+1} \Pi_{t+1}^{-1}] R_t^d - \frac{\omega_t re_t}{n_t} \quad (\text{A.5})$$

(A.5) is the same as (2.15). Use (2.14) and (2.16), (A.5) can be rewritten as

$$\begin{aligned} \frac{\partial V_{i,t}}{\partial n_{i,t}} &= (1 + \lambda_t) \mathbb{E}_t \Lambda_{t,t+1} \Pi_{t+1}^{-1} R_t^d \left(1 - \sigma + \sigma \frac{\partial V_{i,t+1}}{\partial n_{i,t+1}} \right) - \frac{\omega_t re_t}{n_t} \\ &= (1 + \lambda_t) \left((1 - \sigma) \mathbb{E}_t \Lambda_{t,t+1} \Pi_{t+1}^{-1} R_t^d + \sigma (1 - \sigma) \mathbb{E}_t \left[\Lambda_{t,t+1} \Pi_{t+1}^{-1} R_t^d (1 + \lambda_{t+1}) \mathbb{E}_{t+1} \Lambda_{t+1,t+2} \Pi_{t+2}^{-1} R_{t+1}^d \right] + \dots \right) \\ &\quad - \frac{\omega_t re_t}{n_t} \end{aligned}$$

Since $\lambda_{t+j} \geq 0 \forall j$, the derivative of the value function with respect to net worth satisfies:

$$\begin{aligned} \frac{\partial V_{i,t}}{\partial n_{i,t}} &\geq (1 + \lambda_t) \left((1 - \sigma) \mathbb{E}_t \Lambda_{t,t+1} \Pi_{t+1}^{-1} R_t^d + \sigma (1 - \sigma) \mathbb{E}_t \left[\Lambda_{t,t+1} \Pi_{t+1}^{-1} R_t^d \mathbb{E}_{t+1} \Lambda_{t+1,t+2} \Pi_{t+2}^{-1} R_{t+1}^d \right] + \dots \right) - \frac{\omega_t re_t}{n_t} \\ &= (1 + \lambda_t) (1 - \sigma) (1 + \sigma + \sigma^2 + \dots) - \frac{\omega_t re_t}{n_t} \\ &= 1 + \lambda_t - \frac{\omega_t re_t}{n_t} \end{aligned}$$

where we use the household's first order condition in (2.24).

When neither constraint binds, then $\lambda_{t+j} = \omega_t = 0$, and $\frac{\partial V_{i,t}}{\partial n_{i,t}} = 1$, so that $\phi_t = \frac{1}{\theta_t}$. When $\lambda_{t+j} > 0$ for any j and the reserve requirement is non-binding, $\frac{\partial V_{i,t}}{\partial n_{i,t}} > 1$. When both constraints bind the size of the derivative of the value function with respect to net worth relative to unity is indeterminate.

A.2 Nominal Rigidities

A.2.1 Labor Union

Profit of a typical labor union in nominal terms is

$$DIV_{L,t}(h) = W_t(h) L_{d,t}(h) - MRS_t L_t(h)$$

Imposing that $L_t(h) = L_{d,t}(h)$ and using the demand curve, (2.25), this can be written:

$$DIV_{L,t}(h) = W_t(h)^{1-\epsilon_w} W_t^{\epsilon_w} L_{d,t} - MRS_t W_t(h)^{-\epsilon_w} W_t^{\epsilon_w} L_{d,t}$$

A labor union given the opportunity to set the wage $W_t(h)$ in period t must take into account the possibility that it will be stuck with this wage for some time. In particular, the probability that a wage chosen in t is still relevant in $t + j$ is ϕ_w^j . The possibility of indexation results in the nominal wage in $t + j$ for a union who has not updated since period t being: $W_t(h) \left(\frac{P_{t+j-1}}{P_{t-1}} \right)^{\gamma_w}$. The problem of a labor union updating in period t is to choose a wage to maximize the present discounted value of real profits, where

discounting is by the household's stochastic discount factor and the probability of non-wage adjustment:

$$\max_{W_t(h)} \mathbb{E}_t \sum_{j=0}^{\infty} \phi_w^j \Lambda_{t,t+j} \left[\left(\frac{P_{t+j-1}}{P_{t-1}} \right)^{(1-\epsilon_w)\gamma_w} W_t(h)^{1-\epsilon_w} P_{t+j}^{\epsilon_w-1} w_{t+j}^{\epsilon_w} L_{d,t+j} - mrs_{t+j} \left(\frac{P_{t+j-1}}{P_{t-1}} \right)^{-\epsilon_w\gamma_w} W_t(h)^{-\epsilon_w} P_{t+j}^{\epsilon_w} w_{t+j}^{\epsilon_w} L_{d,t+j} \right]$$

where $\Lambda_{t,t+j} = \Lambda_{t,t+1} \dots \Lambda_{t+j-1,t+j}$.

The first order condition is

$$(\epsilon_w - 1) W_t(h)^{-\epsilon_w} \mathbb{E}_t \sum_{j=0}^{\infty} \phi_w^j \Lambda_{t,t+j} \left(\frac{P_{t+j-1}}{P_{t-1}} \right)^{(1-\epsilon_w)\gamma_w} P_{t+j}^{\epsilon_w-1} w_{t+j}^{\epsilon_w} L_{d,t+j} = \epsilon_w W_t(h)^{-\epsilon_w-1} \mathbb{E}_t \sum_{j=0}^{\infty} \phi_w^j \Lambda_{t,t+j} mrs_{t+j} \left(\frac{P_{t+j-1}}{P_{t-1}} \right)^{-\epsilon_w\gamma_w} P_{t+j}^{\epsilon_w} w_{t+j}^{\epsilon_w} L_{d,t+j}$$

The reset wage is the same across all labor unions. Hence, drop the h index, and the optimal price W_t^* can be written as:

$$W_t^* = \frac{\epsilon_w}{\epsilon_w - 1} \frac{F_{1,t}}{F_{2,t}}$$

where $F_{1,t}$ and $F_{2,t}$ are recursive representations of the infinite sums above:

$$\begin{aligned} F_{1,t} &= mrs_t P_t^{\epsilon_w} w_t^{\epsilon_w} L_{d,t} + \phi_w \Lambda_{t,t+1} \Pi_t^{-\epsilon_w\gamma_w} F_{1,t+1} \\ F_{2,t} &= P_t^{\epsilon_w-1} w_t^{\epsilon_w} L_{d,t} + \phi_w \Lambda_{t,t+1} \Pi_t^{(1-\epsilon_w)\gamma_w} F_{2,t+1} \end{aligned}$$

Hence (2.27) - (2.29) for $f_{1,t} = F_{1,t}/P_t^{\epsilon_w}$ and $f_{2,t} = F_{2,t}/P_t^{\epsilon_w-1}$.

Aggregation Integrate (2.25) across h , noting that $\int_0^1 L_{d,t}(h)dh = L_t$. Using the demand function for a union's labor, (2.25), yields

$$L_t = L_{d,t} v_t^w \quad (\text{A.6})$$

where v_t^w is a measure of wage dispersion:

$$v_t^w = \int_0^1 \left(\frac{w_t(h)}{w_t} \right)^{-\epsilon_w} dh$$

Note that this can be written in terms of real wages since it is a ratio. Because of properties of Calvo wage-setting, we can write this as

$$\begin{aligned} v_t^w &= (1 - \phi_w) \left(\frac{w_t^*}{w_t} \right)^{-\epsilon_w} + \int_{1-\phi_w}^1 \left(\frac{\Pi_{t-1}^{\gamma_w} W_{t-1}(h)}{W_t} \right)^{-\epsilon_w} dh \\ &= (1 - \phi_w) \left(\frac{w_t^*}{w_t} \right)^{-\epsilon_w} + \Pi_{t-1}^{-\gamma_w \epsilon_w} W_t^{\epsilon_w} W_{t-1}^{-\epsilon_w} \int_{1-\phi_w}^1 \left(\frac{W_{t-1}(h)}{W_{t-1}} \right)^{-\epsilon_w} dh \end{aligned}$$

which may be written as

$$v_t^w = (1 - \phi_w) \left(\frac{w_t^*}{w_t} \right)^{-\epsilon_w} + \phi_w \Pi_{t-1}^{-\gamma_w \epsilon_w} W_t^{\epsilon_w} W_{t-1}^{-\epsilon_w} v_{t-1}^w$$

Expressing this in real terms gives

$$v_t^w = (1 - \phi_w) \left(\frac{w_t^*}{w_t} \right)^{-\epsilon_w} + \phi_w \left(\frac{\Pi_t}{\Pi_{t-1}^{\gamma_w}} \right)^{\epsilon_w} \left(\frac{w_t}{w_{t-1}} \right)^{\epsilon_w} v_{t-1}^w \quad (\text{A.7})$$

From (2.26), we have

$$W_t^{1-\epsilon_w} = (1 - \phi_w) (W_t^*)^{1-\epsilon_w} + \int_{1-\phi_w}^1 (\Pi_{t-1}^{\gamma_w} W_{t-1}(h))^{1-\epsilon_w} dh$$

Via a law of large numbers, this is

$$W_t^{1-\epsilon_w} = (1 - \phi_w) (W_t^*)^{1-\epsilon_w} + \Pi_{t-1}^{\gamma_w(1-\epsilon_w)} \phi_w W_{t-1}^{1-\epsilon_w}$$

Dividing both sides by $P_t^{1-\epsilon_w}$ gives

$$w_t^{1-\epsilon_w} = (1 - \phi_w) (w_t^*)^{1-\epsilon_w} + \phi_w \Pi_{t-1}^{\gamma_w(1-\epsilon_w)} \Pi_t^{\epsilon_w-1} w_{t-1}^{1-\epsilon_w} \quad (\text{A.8})$$

A.2.2 Retailers

The nominal profit of a retail firm is

$$DIV_{R,t}(f) = P_t(f)Y_t(f) - P_{m,t}Y_{m,t}(f)$$

Plugging in production function equating $Y_{m,t}(f) = Y_t(f)$ and the demand function, we get:

$$DIV_{R,t}(f) = P_t(f)^{1-\epsilon_p} P_t^{\epsilon_p} Y_t - P_{m,t} P_t(f)^{-\epsilon_p} P_t^{\epsilon_p} Y_t$$

Each period, a retailer faces a constant hazard, $1 - \phi_p$, of being able to adjust its price. Consider the problem of a retailer given the opportunity to adjust in period t . When it sets a price, it must take into account that the price chosen in t will still be in effect in period $t + j$ with probability ϕ_p^j . Indexation to lagged inflation means that an unupdated price in period $t + j$ will be: $P_t(f) \left(\frac{P_{t+j-1}}{P_{t-1}} \right)^{\gamma_p}$. When choosing a price, the retailer maximizes the present discounted value of real profits returned to the household, where discounting is by the household's stochastic discount factor augmented by the probability of non-adjustment:

$$\max_{P_t(f)} \mathbb{E}_t \sum_{j=0}^{\infty} \phi_p^j \Lambda_{t,t+j} \left[P_t(f)^{1-\epsilon_p} \left(\frac{P_{t+j-1}}{P_{t-1}} \right)^{(1-\epsilon_p)\gamma_p} P_{t+j}^{\epsilon_p-1} Y_{t+j} - P_{m,t+j} P_t(f)^{-\epsilon_p} \left(\frac{P_{t+j-1}}{P_{t-1}} \right)^{-\epsilon_p \gamma_p} P_{t+j}^{\epsilon_p-1} Y_{t+j} \right]$$

The first order condition is

$$\begin{aligned} (\epsilon_p - 1) P_t(f)^{-\epsilon_p} \mathbb{E}_t \sum_{j=0}^{\infty} \phi_p^j \Lambda_{t,t+j} \left(\frac{P_{t+j-1}}{P_{t-1}} \right)^{(1-\epsilon_p)\gamma_p} P_{t+j}^{\epsilon_p-1} Y_{t+j} = \\ \epsilon_p P_t(f)^{-\epsilon_p-1} \mathbb{E}_t \sum_{j=0}^{\infty} \phi_p^j \Lambda_{t,t+j} P_{m,t+j} \left(\frac{P_{t+j-1}}{P_{t-1}} \right)^{-\epsilon_p \gamma_p} P_{t+j}^{\epsilon_p} Y_{t+j} \end{aligned}$$

Define variables:

$$\begin{aligned} X_{1,t} &= \sum_{j=0}^{\infty} (\phi_p \beta)^j \frac{\mu_{t+j}}{\mu_t} P_{m,t+j} \left(\frac{P_{t+j-1}}{P_{t-1}} \right)^{-\epsilon_p \gamma_p} P_{t+j}^{\epsilon_p} Y_{t+j} \\ X_{2,t} &= \sum_{j=0}^{\infty} (\phi_p \beta)^j \frac{\mu_{t+j}}{\mu_t} \left(\frac{P_{t+j-1}}{P_{t-1}} \right)^{(1-\epsilon_p)\gamma_p} P_{t+j}^{\epsilon_p-1} Y_{t+j} \end{aligned}$$

These can be written recursively as:

$$\begin{aligned} X_{1,t} &= p_{m,t} P_t^{\epsilon_p} Y_t + \phi_p \Lambda_{t,t+1} \Pi_t^{-\epsilon_p \gamma_p} X_{1,t+1} \\ X_{2,t} &= P_t^{\epsilon_p - 1} Y_t + \phi_p \Lambda_{t,t+1} \Pi_t^{(1-\epsilon_p) \gamma_p} X_{2,t+1} \end{aligned}$$

Note, all retailers set the same price. We call this reset price P_t^* . Hence, the first order condition can be written as

$$P_t^* = \frac{\epsilon_p}{\epsilon_p - 1} \frac{X_{1,t}}{X_{2,t}}$$

Defining $x_{1,t} = X_{1,t}/P_t^{\epsilon_p}$, $x_{2,t} = X_{2,t}/P_t^{\epsilon_p - 1}$, and $p_t^* = P_t^*/P_t$ gives (2.32) - (2.34).

Aggregation Integrate the demand for retail output, (2.30), across retailers, noting that $Y_t(f) = Y_{m,t}(f)$, $\int_0^1 Y_{m,t}(f) df = Y_{m,t}$. This yields

$$Y_t v_t^p = Y_{m,t} \quad (\text{A.9})$$

where

$$v_t^p = \int_0^1 \left(\frac{P_t(f)}{P_t} \right)^{-\epsilon_p} df$$

This is a measure of price dispersion. Using properties of Calvo pricing with indexation to lagged inflation yields:

$$\begin{aligned} v_t^p &= (1 - \phi_p) (p_t^*)^{-\epsilon_p} + \int_{1-\phi_p}^1 \left(\frac{\Pi_{t-1}^{\gamma_p} P_{t-1}(f)}{P_t} \right)^{-\epsilon_p} df \\ &= (1 - \phi_p) (p_t^*)^{-\epsilon_p} + \Pi_{t-1}^{-\gamma_p \epsilon_p} P_t^{\epsilon_p} P_{t-1}^{-\epsilon_p} \int_{1-\phi_p}^1 \left(\frac{P_{t-1}(f)}{P_{t-1}} \right)^{-\epsilon_p} df \end{aligned}$$

Via a law of large numbers, this reduces to

$$v_t^p = (1 - \phi_p) (p_t^*)^{-\epsilon_p} + \phi_p \left(\frac{\Pi_t}{\Pi_{t-1}^{\gamma_p}} \right)^{\epsilon_p} v_{t-1}^p \quad (\text{A.10})$$

Similarly, the aggregate price index, (2.31), may be written as:

$$\begin{aligned} P_t^{1-\epsilon_p} &= (1 - \phi_p) (P_t^*)^{1-\epsilon_p} + \int_{1-\phi_p}^1 \Pi_{t-1}^{\gamma_p (1-\epsilon_p)} P_{t-1}(f)^{1-\epsilon_p} df \\ &= (1 - \phi_p) (P_t^*)^{1-\epsilon_p} + \phi_p \Pi_{t-1}^{\gamma_p (1-\epsilon_p)} P_{t-1}^{1-\epsilon_p} \end{aligned}$$

Divide both sides by $P_t^{1-\epsilon_p}$ to obtain

$$1 = (1 - \phi_p) (p_t^*)^{1-\epsilon_p} + \phi_p \Pi_{t-1}^{\gamma_p (1-\epsilon_p)} \Pi_t^{\epsilon_p - 1} \quad (\text{A.11})$$

A.3 Wholesale Firms

Write wholesaler profits, (2.38), in real terms by dividing by the aggregate price level:

$$div_{m,t} = p_{m,t} A_t (u_t K_t)^\alpha L_{d,t}^{1-\alpha} - w_t L_{d,t} - p_t^k \hat{I}_t + Q_t \left(\frac{F_{m,t}}{P_t} - \kappa \frac{F_{m,t-1}}{P_{t-1}} \Pi_t^{-1} \right) - \frac{F_{m,t-1}}{P_{t-1}} \Pi_t^{-1}$$

Note that $F_{m,t}$ is the amount of nominal bonds.

The firm discounts real dividends using the stochastic discount factor of the household, $\Lambda_{t,t+j}$. A

Lagrangian with the constraints (2.36) and (2.37) is

$$\begin{aligned} \mathbb{L}_{m,t} = \mathbb{E}_t \sum_{j=0}^{\infty} \Lambda_{t,t+j} \Bigg\{ & p_{m,t+j} A_{t+j} (u_{t+j} K_{t+j})^\alpha L_{d,t+j}^{1-\alpha} - w_{t+j} L_{d,t+j} - p_{t+j}^k \widehat{I}_{t+j} + Q_{t+j} \left(\frac{F_{m,t+j}}{P_{t+j}} - \kappa \frac{F_{m,t+j-1}}{P_{t+j-1}} \Pi_{t+j}^{-1} \right) \\ & - \frac{F_{m,t+j-1}}{P_{t+j-1}} \Pi_{t+j}^{-1} + \nu_{1,t+j} \left(\widehat{I}_{t+j} + (1 - \delta(u_{t+j})) K_{t+j} - K_{t+j+1} \right) \\ & + \nu_{2,t+j} \left(Q_{t+j} \left(\frac{F_{m,t+j}}{P_{t+j}} - \kappa \frac{F_{m,t+j-1}}{P_{t+j-1}} \Pi_{t+j}^{-1} \right) - \psi p_{t+j}^k \widehat{I}_{t+j} \right) \Bigg\} \end{aligned}$$

The first order conditions are

$$\frac{\partial \mathbb{L}_{m,t}}{\partial L_{d,t}} = p_{m,t} (1 - \alpha) A_t (u_t K_t)^\alpha L_{d,t}^{-\alpha} - w_t = 0 \quad (\text{A.12})$$

$$\frac{\partial \mathbb{L}_{m,t}}{\partial \widehat{I}_t} = -p_t^k + \nu_{1,t} - p_t^k \psi \nu_{2,t} = 0 \quad (\text{A.13})$$

$$\frac{\partial \mathbb{L}_{m,t}}{\partial u_t} = p_{m,t} \alpha A_t (u_t K_t)^{\alpha-1} K_t L_{d,t}^{1-\alpha} - \nu_{1,t} \delta'(u_t) K_t = 0 \quad (\text{A.14})$$

$$\frac{\partial \mathbb{L}_{m,t}}{\partial K_{t+1}} = \mathbb{E}_t \Lambda_{t,t+1} \left[\alpha p_{m,t+1} A_{t+1} (u_{t+1} K_{t+1})^{\alpha-1} u_{t+1} L_{d,t+1}^{1-\alpha} + \nu_{1,t+1} (1 - \delta(u_{t+1})) \right] \quad (\text{A.15})$$

$$-\nu_{1,t} = 0 \quad (\text{A.16})$$

$$\frac{\partial \mathbb{L}_{m,t}}{\partial F_{m,t}} = \frac{Q_t}{P_t} + \nu_{2,t} \frac{Q_t}{P_t} - \mathbb{E}_t \Lambda_{t,t+1} \left[\frac{1}{P_t} \Pi_{t+1}^{-1} + \kappa \frac{Q_{t+1}}{P_t} \Pi_{t+1}^{-1} + \nu_{2,t+1} \kappa \frac{Q_{t+1}}{P_t} \Pi_{t+1}^{-1} \right] = 0 \quad (\text{A.17})$$

Define $M_{1,t} = 1 + \psi \nu_{2,t}$, $M_{2,t} = 1 + \nu_{2,t}$. Then we can eliminate $\nu_{1,t}$ using $\nu_{1,t} = p_t^k M_{1,t}$ from (A.13). Simplifying the rest of these expressions yields (2.39) - (2.43).

A.4 Capital producer

Nominal dividend earned by the capital producer is

$$DIV_{k,t} = P_t^k \left[1 - S \left(\frac{I_t}{I_{t-1}} \right) \right] I_t - P_t I_t$$

In real terms,

$$div_{k,t} = p_t^k \left[1 - S \left(\frac{I_t}{I_{t-1}} \right) \right] I_t - I_t$$

The objective is to maximize the present discounted value of profit using the stochastic discount factor of the household:

$$\max_{I_t} \mathbb{E}_t \sum_{j=0}^{\infty} \Lambda_{t,t+j} \left\{ p_{t+j}^k \left[1 - S \left(\frac{I_{t+j}}{I_{t+j-1}} \right) \right] I_{t+j} - I_{t+j} \right\}$$

The first order condition is (2.45).

A.5 Aggregation and Exogenous Processes

The model features three exogenous variables – neutral productivity, A_t ; government spending, G_t ; and the liquidity process, θ_t . These follow AR(1) processes in the log:

$$\ln A_t = \rho_A \ln A_{t-1} + s_A \varepsilon_{A,t} \quad (\text{A.18})$$

$$\ln G_t = (1 - \rho_G) \ln G + \rho_G \ln G_{t-1} + s_G \varepsilon_{G,t} \quad (\text{A.19})$$

$$\ln \theta_t = (1 - \rho_\theta) \ln \theta + \rho_\theta \ln \theta_{t-1} + s_\theta \varepsilon_{\theta,t} \quad (\text{A.20})$$

Autoregressive parameters are restricted to lie between zero and one and shocks are drawn from standard normal distributions, with s_A , s_G , and s_θ denoting the standard deviations of the shocks. G and θ are non-stochastic steady state values; the non-stochastic steady state value of productivity is normalized to unity.

Aggregate balance sheet condition of financial intermediaries in (2.4), and write in real terms (variables without i indexes are integrated across intermediaries, and $d_t = D_t/P_t$ is real deposits):

$$Q_t f_t + Q_{B,t} b_t + re_t = d_t + n_t \quad (\text{A.21})$$

Aggregate net worth dynamics in (A.1):

$$n_t = \sigma \Pi_t^{-1} \left[(R_t^F - R_{t-1}^d) Q_{t-1} f_{t-1} + (R_t^B - R_{t-1}^d) Q_{B,t-1} b_{t-1} + (R_{t-1}^{re} - R_{t-1}^d) re_{t-1} + R_{t-1}^d n_{t-1} \right] + X \quad (\text{A.22})$$

where σ is the fraction of continuing financial intermediaries, and X is the start up funds given to the new intermediaries.

Aggregating (2.17) across intermediaries yields:

$$Q_t f_t + \Delta Q_{B,t} b_t \leq \phi_t n_t \quad (\text{A.23})$$

where (A.23) holds with equality when (2.9) binds.

The aggregate resource constraint is standard:

$$Y_t = C_t + I_t + G_t \quad (\text{A.24})$$

A.6 Equilibrium

The full set of equilibrium conditions include:

- financial intermediaries (7 equations): (2.6), (2.7), (2.11) - (2.15)
- households (4 equations): (2.21) - (2.24)
- labor market (3 equations): (2.27) - (2.29)
- production (13 equations): for retail firms, (2.32) - (2.34), for wholesale firms, (2.35) - (2.37), (2.39) - (2.43), for capital producing firms, (2.44) - (2.45)
- fiscal authority (1 equation): (2.46)
- central bank (2 equations): (2.47), plus its remittance to the fiscal authority, which satisfies:

$$T_{cb,t} = (1 + \kappa Q_t) \Pi_t^{-1} f_{cb,t-1} + (1 + \kappa Q_{B,t}) \Pi_t^{-1} b_{cb,t-1} - R_{t-1}^{re} \Pi_t^{-1} re_{t-1}$$

- monetary policy (5 equations): (3.1), plus two equations for R_t^{re} , R_t^d ((3.2) for conventional policy, (3.3) for the conventional ZLB, or (3.6) for NIRP), as well as (5.1) (which nests (3.4) when $\Psi_f = 0$) and (3.5).
- aggregation (15 equations): exogenous processes include (A.18) - (A.20). Bond market clearing conditions are (2.48) - (2.49), where $f_t = \sum_i f_{i,t}$ and $b_t = \sum_i b_{i,t}$, and other aggregate conditions are (A.6) - (A.11) and (A.21) - (A.24).

These are 50 equations for 50 variables $\left\{ R_t^F, R_t^B, R_t^{re}, R_t^d, R_t^{TR}, Q_t, Q_{B,t}, \Lambda_{t,t+1}, \Omega_t, \Pi_t, \lambda_t, \omega_t, \phi_t, re_t, n_t, \mu_t, C_t, L_t, mrs_t, w_t^*, f_{1,t}, f_{2,t}, w_t, L_{d,t}, p_t^*, x_{1,t}, x_{2,t}, p_{m,t}, Y_t, Y_{m,t}, u_t, K_t, \hat{I}_t, p_t^k, f_{m,t}, M_{1,t}, M_{2,t}, I_t, T_t, T_{cb,t}, f_{cb,t}, b_{cb,t}, A_t, G_t, \theta_t, f_t, b_t, d_t, v_t^p, v_t^w \right\}$.

B Model Moments

Table B.1: Unconditional Moments: Model vs. Data

		$d \ln Y_t$	$d \ln C_t$	$d \ln I_t$	$d \ln L_t$	π_t	Credit Spread
Volatility	Model	2.93	0.33	4.63	0.92	0.30	0.17
	Data	2.10	0.73	3.46	1.15	0.43	0.23
Cyclicalities	Model	1.00	0.16	0.95	0.50	-0.06	-0.56
	Data	1.00	0.48	0.80	0.53	-0.10	-0.23

Notes: this table lists selected moments from the model as well as counterparts in the data. Data moments are calculated from 1984q1 - 2007q3. Volatilities for output growth, $d \ln Y_t$, are standard deviations of quarterly output growth expressed at annualized percentage rate. Other volatilities are expressed relative to the volatility of output growth. Cyclicalities are correlations with output growth.

Moments from the model, along with counterparts in the data, are presented in [Table B.1](#).²⁴ As noted in the text, given non-shock parameters, we choose the standard deviations of the three exogenous shocks to produce selected moments which roughly align with the data. For these moments we focus on pre-crisis data, so the model is solved ignoring the possibility of a binding ZLB. The model does a good job at replicating key moments from the data. Output is slightly too volatile and consumption is slightly too smooth in the model relative to the data, while investment is a bit too volatile in the model. Nevertheless, qualitatively, the volatilities of these variables relative to output growth are in-line with observed data. The model does a good job of matching the relative volatilities of hours, inflation, and the observed credit spread, which in the model is defined as the difference between yields to maturity on private and government bonds.²⁵ The model does a good job at matching the cyclicalities of investment, while it underpredicts the cyclicalities of consumption growth.²⁶ The model does an excellent job at matching the cyclicalities of hours and inflation. It produces slightly too much countercyclicality in the credit spread, but qualitatively it is not far off. In terms of the relative importance of the different shocks, the productivity shock accounts for roughly 40 percent of the volatility in output growth, the government spending shock 10 percent, and the liquidity shock the remaining 50 percent.

²⁴For the data, $d \ln Y_t$ corresponds to the growth rate of real GDP measured at an annualized percentage rate. We measure consumption as the sum of (nominal) consumption of non-durables and services, deflated by the GDP deflator to put it in real terms. Our measure of investment is the sum of (nominal) gross private fixed investment plus consumption of durables, also deflated by the GDP deflator. Our measure of labor hours is the index of total hours worked in the non-farm business sector. Inflation is the growth rate of the GDP price deflator. The credit spread is the Baa-10 Year Treasury spread.

²⁵To be precise, RL_t^F is the gross yield to maturity on private bonds, and $RL_t^B = \frac{1}{Q_{B,t}} + \kappa$ is the gross yield to maturity on government debt. The yield spread is then defined as $\ln RL_t^F - \ln RL_t^B$.

²⁶Discrepancies between model and data moments, and in particular that concerning the cyclicalities of consumption, can easily be made smaller by including additional shocks (e.g. preference shocks).