Unconventional Monetary Policies and Inequality

Donggyu Lee*
Federal Reserve Bank of New York

Nov 2025

Abstract

This paper examines the aggregate and distributional effects of the Federal Reserve's unconventional monetary policy, comprising quantitative easing and forward guidance, during the 2009–2015 effective lower bound episode, using an estimated medium-scale heterogeneous agent New Keynesian model. Relative to a counterfactual without unconventional policy, these interventions stimulate aggregate activity and generate broad-based welfare gains, with an average gain equivalent to a 0.27 percent increase in lifetime consumption. These welfare gains are mildly U-shaped across the wealth distribution, with households at the bottom gaining mainly from a lower unemployment rate, while higher profits and equity prices benefit the top disproportionately. Comparing these outcomes to an unconstrained conventional interest rate rule, I find that once unconventional tools are available the effective lower bound is not very costly in aggregate terms, while conventional policy alone would have implied somewhat more uneven distributional effects.

JEL classification: E12, E30, E52, E58

Keywords: unconventional monetary policy, inequality, heterogeneous agent New Keynesian model, quantitative easing, forward guidance, effective lower bound.

*Macroeconomic and Monetary Studies Department (donggyu.lee@ny.frb.org). This paper was previously circulated under the title "Quantitative Easing and Inequality" and was part of my doctoral dissertation at the University of Maryland, College Park. I am deeply grateful to my advisors Borağan Aruoba, Luminita Stevens, Thomas Drechsel and John Shea for their invaluable guidance and support. I also thank Pierre De Leo, Guido Kuersteiner, Seth Murray, Gonzalo Garcia-Trujillo, Jun Hee Kwak, and Eugene Oue for helpful comments. The views expressed in this paper are those of the author and do not necessarily reflect the position of the Federal Reserve Bank of New York, or the Federal Reserve System.

1 Introduction

In recent decades, income and wealth inequality have been increasing in the United States, which motivated the use of heterogeneous agent macroeconomic models that capture the large degree of household inequality observed in the data. A particularly lively debate has centered on the distributional consequences of monetary policy.[1] An existing empirical literature measures the distributional consequences of both conventional monetary policy and unconventional monetary policy (UMP).[2] By contrast, evidence from estimated structural models with household heterogeneity is scarce. This gap matters because the Federal Reserve's UMP in the aftermath of the Great Recession, especially quantitative easing (QE), has often been criticized for exacerbating disparities in income and wealth among U.S. households.[3] However, whether UMP raises inequality remains an open question.

This paper provides a structural evaluation of the aggregate and distributional effects of UMP, specifically QE and forward guidance (FG), using a medium-scale heterogeneous agent New Keynesian (HANK) model. The central challenge is to capture substantial heterogeneity in household income and wealth and to trace how monetary policy affects these variables across the distribution. To meet this challenge, I augment a standard incomplete markets heterogeneous agent model with a two asset portfolio structure consisting of liquid deposits and illiquid equity, a frictional labor market with wage rigidity, and a banking sector. The policy block includes both conventional interest rate policy and UMP, with QE modeled as central bank asset purchases following Gertler and Karadi (2011), and FG modeled as excess effective lower bound (ELB) duration following Jones et al. (2022).

The model is calibrated to cross sectional data on household income and wealth from the 2007 Survey of Consumer Finances (SCF) to capture heterogeneity in wealth and income composition and the extent of inequality. In the steady state, the top 10 percent of wealthy households hold about 70 percent of total wealth, primarily in the form of equity, with business and asset income accounting for about 50 percent of their total income, consistent with U.S. data. By contrast, households in the bottom 80 percent rely predominantly on labor income, and a larger share near the bottom is unemployed and thus more exposed to unemployment risk.[4]

[1]For trends in inequality, see Heathcote et al. (2010), Saez and Zucman (2016), Gould (2019). For the discussion on inequality and monetary policy, see Yellen (2014); Bernanke (2015), Draghi (2016).

[2]For conventional monetary policy, see, for instance, Coibion et al. (2017) and Mumtaz and Theophilopoulou (2017). For unconventional monetary policy, see Bivens (2015) and Montecino and Epstein (2015).

[3]See, for instance, Schwartz (2013); Cohan (2014).

[4]According to the 2007 SCF data, labor income (wages and salaries) constitutes about 80 percent of the total income for the bottom 80 percent of the wealth distribution, with the remainder largely consisting of transfer income. In stark contrast, for the top 0.1 percent of wealthy households, labor

The model is then estimated on U.S. aggregate time series using Bayesian methods to discipline the parameters that govern its dynamics, explicitly accounting for the binding effective lower bound and the Federal Reserve's UMP, that is quantitative easing and forward guidance, from 2009 to 2015. The estimated model implies substantial nominal wage rigidity, relatively flexible employment adjustment, and volatile profits. Importantly, profits are strongly procyclical in response to demand shocks, consistent with the empirical evidence and in contrast to many New Keynesian models in which profits are countercyclical.[5] This behavior of profits is driven by search and matching frictions in the labor market combined with wage rigidity and firm fixed costs, the latter inspired by Anderson et al. (2018).[6] As a result, the model can capture the positive impact of a boost in aggregate demand on both employment and profits.

With the estimated model, I examine the aggregate and distributional impact of UMP in the aftermath of the Great Recession, when the policy rate was at its lower bound. While conventional interest rate policy was inactive due to the lower bound, the central bank supported the aggregate economy by transforming demand for a nonproductive asset, that is, bonds or deposits, into demand for a productive asset (capital) and by lowering the expected path of future interest rates. At the posterior mode, the model implies that, from 2009 to 2015, UMP on average increased output by about 1 percent, raised profits by 3.3 percent, increased equity prices by about 1 percent, and lowered the unemployment rate by 1.4 percentage points relative to a counterfactual without UMP. UMP raises real wages only slightly, with an average increase of 0.1 percent, reflecting wage rigidity. These magnitudes are in line with existing empirical evidence on unconventional monetary policy.

These aggregate effects translate into positive welfare gains for all households. At the posterior mode, the average welfare gain from UMP, relative to a counterfactual without UMP and measured in consumption equivalents at the beginning of the ELB episode, is about 0.27 percent of lifetime consumption. Across the wealth distribution, welfare gains are mildly U shaped, reflecting heterogeneity in wealth and income composition and exposure to unemployment risk. Most households benefit from lower unemployment induced by the boost to aggregate demand from UMP. Because unemployment is more prevalent at the bottom of the wealth distribution,

income makes up only 17 percent, and transfer income accounts for less than 1 percent of their total income. The remaining 85 percent primarily comes from profit-related sources, such as business income and dividends. For the top 10 percent of households, labor income comprises about 50 percent.

[5] For discussion of the cyclicality of profits in New Keynesian models, see, for instance, Broer et al. (2019) and Bilbiie and Känzig (2023).

[6] These features dampen the response of real marginal costs and, following a positive demand shock, imply a fall in average costs, which contributes to the procyclicality of profits. In addition, profits from the financial sector also contribute to the strong procyclicality of aggregate profits.

their gains are larger, about 0.3 percent of lifetime consumption, than for the middle, about 0.26 percent. For the top 10 percent of the wealth distribution, welfare gains arise primarily through higher profits and equity prices and are again about 0.3 percent of lifetime consumption.

Given the broad based welfare gains, it is natural to ask how UMP affected inequality as captured by measures commonly used in the empirical literature. To facilitate comparison, I focus on the income Gini index and the top 10 percent income share. Because welfare gains display a mild U shape across wealth, distributional effects are nonlinear and can appear mixed across inequality measures. At the posterior mode, during the ELB episode the model implies that the top 10 percent income share rose by 0.17 percentage point, mainly through higher profits and equity prices, while UMP lowered the income Gini index by 0.04 percentage point on average by reducing unemployment. Taken together, these results help reconcile the mixed findings in empirical work on the distributional consequences of UMP.

Finally, I decompose the effects of UMP into the contributions from QE and FG and compare UMP with conventional monetary policy. In the decomposition, at the posterior mode, QE accounts for about 55 percent of the total stimulus to output during the ELB episode, with the remainder attributed to FG. This split carries over to welfare and inequality measures. FG amplifies the mild U-shaped profile of welfare gains, strengthens the decline in the Gini through lower unemployment, and raises the top 10 percent income share by further increasing profits and equity prices. I then consider a counterfactual in which the policy rate is unconstrained and follows the estimated rule during the ELB episode. On aggregates, UMP is about as effective as conventional policy. On distributional outcomes, conventional policy is somewhat more adverse because it channels larger gains to the financial sector through lower borrowing costs, whereas UMP, especially QE, crowds out some private investment even as higher equity prices initially expand financial sector balance sheets.

Related Literature

This paper contributes to the empirical literature on the distributional consequences of monetary policy by quantifying the effects of UMP on welfare and on inequality measures within a single structural framework. Findings for UMP are mixed across settings and measures. Reductions in inequality through employment are emphasized in Bivens (2015), Casiraghi et al. (2018), and Lenza and Slacalek (2018), while increases through asset price channels are emphasized in Montecino and Epstein (2015) and Domanski et al. (2016). The approach here helps reconcile these differences by evaluating UMP in one model and by reporting multiple distributional statistics alongside welfare. This clarifies how conclusions can differ when studies focus on different parts of the distribution or on different measures, and it delivers welfare effects that reduced form studies cannot measure.[7]

Another contribution is to the HANK literature, which uses frameworks with heterogeneous agents to study economies with rich variation across households and firms. Much of this work examines how heterogeneity or inequality influences aggregate dynamics, monetary transmission, or preferences over policy rules, for example in Kaplan et al. (2018), Bayer et al. (2020), and Gornemann et al. (2016). In contrast, the primary focus here is the distributional effects of monetary policy. The modeling framework is built for this purpose and combines a two-asset structure, a frictional labor market with wage rigidity, and a banking sector so that policy operates through employment, wages, profits, and asset prices. The specification is chosen to represent cross-sectional patterns of income and wealth while keeping aggregate responses to policies and shocks consistent with the data. Relative to standard New Keynesian setups, which tend to generate countercyclical profits after demand expansions, the estimated model produces profits that rise with demand due to this combination of features, which helps capture how a demand driven policy boost raises business and asset income for wealthy households.[8]

This paper also contributes to the structural modeling of UMP by extending representative agent analyses, such as Gertler and Karadi (2011), to a heterogeneous agent setting and by quantifying both distributional and welfare effects of different forms of UMP, namely QE and forward guidance. Two closely related papers are Cui and Sterk (2021) and Sims et al. (2022). Cui and Sterk (2021) study QE in a HANK environment with assets of differing liquidity and highlight a liquidity composition channel through which asset purchases affect aggregate demand. Sims et al. (2022) use a one asset HANK model to ask whether responses to a QE shock differ from those in a representative agent model and find limited differences in aggregate dynamics. The present paper instead takes the banking and QE transmission mechanism of Gertler and Karadi (2011) as a starting point and provides a quantitative assessment of UMP in an estimated heterogeneous agent model in which unemployment, wages, profits, and asset prices all adjust endogenously and in which QE and forward guidance are both active policy tools. This allows me to study the aggregate, distributional, and welfare consequences of QE and forward guidance during the 2009–2015 ELB episode and to document how the transmission

[7]For conventional policy, tightening is found to raise inequality in Coibion et al. (2017), Mumtaz and Theophilopoulou (2017), and Furceri et al. (2018), while neutral or opposite effects are reported in Inui et al. (2017), Davtyan (2017), and Hafemann et al. (2018).

[8]For empirical evidence on the procyclicality of profits or business income in response to monetary policy shocks, see Christiano et al. (2005) and Coibion et al. (2017). The Online Appendix also reports SVAR impulse responses showing that profits increase after an expansionary monetary policy shock.

mechanism varies across the wealth distribution.[9]

Lastly, this paper relates to work that estimates HANK models, for example in Bayer et al. (2020) and Auclert et al. (2020b). The contribution here is to estimate a HANK model with an occasionally binding lower bound on the policy rate together with UMP. The likelihood is evaluated using an inversion filter, as in Guerrieri and Iacoviello (2017), Cuba-Borda et al. (2019), and Atkinson et al. (2020). Following Jones et al. (2022), forward guidance is modeled as exogenous ELB durations, which avoids treating the duration of ELB periods as a latent regime path during estimation and substantially reduces the computational burden of estimation.[10] To the best of my knowledge, this is among the first papers to estimate a HANK model with an occasionally binding lower bound for the 2009 to 2015 period.

The remainder of the paper proceeds as follows. Section 2 presents the model. Section 3 describes the parametrization and estimation strategy and reports the estimation results. Section 4 studies how different types of monetary policy operate in the model using impulse responses to monetary policy shocks at the steady state. Section 5 conducts counterfactual experiments to evaluate the aggregate and welfare effects of UMP during the ELB episode. Section 6 investigates the impact of UMP on inequality measures. Section 7 decomposes the effects of UMP into the contributions from QE and forward guidance and compares them with an unconstrained conventional policy rule. Section 8 concludes.

2 Model

The model is a medium-scale New Keynesian economy with heterogeneous households who face uninsurable income risk and hold two assets, liquid deposits and illiquid equity. Goods producers face price rigidity and fixed operating costs under monopolistic competition, the labor market features search and matching frictions with wage rigidity, and banks intermediate deposits into equity subject to an incentive constraint. The monetary authority sets the nominal interest rate using a

[9]In Cui and Sterk (2021), liquid payouts from the mutual fund are fixed by employment status and unemployment is exogenous. In Sims et al. (2022), unemployment transitions are exogenous. Hence unemployment and cash flows from asset revaluation do not adjust endogenously to QE in these frameworks, which narrows the scope for distributional movements relative to the environment studied here. I also compare the baseline heterogeneous agent model to a corresponding representative agent version and find that QE is less effective in the representative agent model, while forward guidance is relatively more powerful there, indicating that heterogeneity interacts differently with these two UMP tools.

[10]In Guerrieri and Iacoviello (2017), endogenous ELB durations are found by iterating over exogenous shocks during estimation, which can be computationally costly in a HANK model with a large idiosyncratic state space. Featuring exogenous ELB durations, as in Jones et al. (2022), not only allows the model to include an additional UMP tool, that is, forward guidance, but also keeps estimation tractable without needing to find endogenous durations.

Taylor type rule that is subject to an effective lower bound. When the lower bound binds, the central bank uses two instruments, asset purchases in the spirit of Gertler and Karadi (2011) and forward guidance implemented as exogenous lower bound durations following Jones et al. (2022). A period is a quarter. The main text lays out the agents' problems and the economic environment. Detailed derivations, market clearing conditions, the equilibrium definition, and the solution method are presented in the Online Appendix.

2.1 Households

Time is discrete and denoted by $t = 0, 1, 2, ...$ The economy is populated by a unit mass of households who face uninsurable income risk due to the evolution of idiosyncratic productivity s_t and employment status e_t, and hold two assets, deposits b_{t+1} and equity a_{t+1}. Each period households die with an exogenous probability ζ and are replaced by newborns with the same s_t and e_t and zero wealth.[11] Surviving households choose consumption c_t, hours n_t, and next period assets b_{t+1} and a_{t+1} to maximize expected lifetime utility with discount factor $\beta \in (0, 1)$

$$\mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t (1-\zeta)^t \left\{ u(c_t, n_t | s_t, e_t) - \chi_t \mathbb{1}_{\{a_{t+1} \neq a_t\}} \right\} \right], \tag{1}$$

subject to the budget constraint

$$c_t + q_t a_{t+1} + b_{t+1} = (1-\tau) y_t + (q_t + r_t^a) a_t + (1 + r_t^b) b_t + T_t \ , \tag{2}$$

$$a_{t+1} \geq 0 \ , \ \ b_{t+1} \geq \underline{\mathrm{b}}. \tag{3}$$

The period utility function follows Greenwood et al. (1988),

$$u(c_t, n_t | s_t, e_t) = \frac{\left[c_t - \psi s_t \frac{n_t^{1+\xi}}{1+\xi} \right]^{1-\sigma}}{1-\sigma}, \tag{4}$$

where σ is the inverse of the elasticity of intertemporal substitution (IES), ξ is the inverse Frisch elasticity, and ψ is a scale parameter.

Household earnings y_t depend on e_t. Households can be employed, unemployed, or business owners who receive profits without supplying labor. The job finding probability is endogenous and depends on the aggregate state of the economy, while separations and transitions into and out of business ownership are exogenous. Employed households earn $w_t s_t n_t$. Unemployed households receive benefits $vw \min\{s_t, \bar{s}\}$,

[11] As in Kaplan et al. (2018), the wealth of the deceased is redistributed to surviving households in proportion to their asset holdings.

where w is the steady state real wage, v is the replacement ratio, and $\bar{s}$ is the average idiosyncratic productivity. Business owners receive profit income equal to a fraction ν of aggregate profits Π_t.

Households allocate savings between deposits and equity. Deposits earn the real return $r_t^b = \frac{1+i_t}{\pi_t}$, where i_t is the nominal policy rate and π_t is gross inflation. Households may borrow up to an exogenous limit $\underline{b}$ at a nominal premium $\iota_b \geq 0$, so $r_t^b = \frac{1+i_t+\iota_b}{\pi_t}$ when $b_t < 0$. Equity is purchased at price q_t and pays real dividends r_t^a. Adjusting equity holdings incurs a stochastic utility cost χ_t, drawn from a logistic distribution with location μ_χ and scale σ_χ, which makes equity illiquid and concentrates equity holdings among higher income and higher wealth households. Household earnings are taxed at rate τ, and households receive lump sum transfers T_t from the government and from a money market mutual fund (MMMF), both described below.

2.2 Firms

The production side is standard. Goods producers produce intermediate goods by renting capital and labor services according to a Cobb-Douglas technology.

$$Y_{jt} = Z_t K_{jt}^{\theta} L_{jt}^{1-\theta}, \tag{5}$$

where Y_{jt}, K_{jt}, and L_{jt} are firm j's output, capital input, and labor input, respectively. Total factor productivity Z_t follows an AR(1) process with persistence ρ_z and innovation standard deviation σ_z. Goods producers operate under monopolistic competition and set prices subject to adjustment costs à la Rotemberg (1982). In a symmetric equilibrium in which all firms set the same price and produce the same output, optimal price setting implies a standard New Keynesian Phillips curve.

$$\log\left(\frac{\pi_t}{\pi_{t-1}^{\iota_p}\pi^{1-\iota_p}}\right) = \mathbb{E}_t\left[\Lambda_{t,t+1}\log\left(\frac{\pi_{t+1}}{\pi_t^{\iota_p}\pi^{1-\iota_p}}\right)\right] + \kappa\left(\mathrm{MC}_t - \frac{1}{\Psi_t^p}\right), \tag{6}$$

where κ is the Phillips curve slope, $\Lambda_{t,t+1}$ is the discount factor used by firms, MC_t is real marginal cost, and ι_p is the weight on lagged inflation. Firms face a time-varying markup shock $\Psi_t^p = \frac{\eta_t}{\eta_t - 1}$, where η_t is the elasticity of substitution across goods, and the markup shock follows an AR(1) process with autocorrelation ρ_p and innovation standard deviation σ_p. Goods producers also face a fixed cost shock Ψ_t^F with autocorrelation ρ_F and standard deviation σ_F. Real profits of the goods producing sector are

$$\Pi_t^I = \left(1 - \frac{\mathrm{MC}_t}{\Psi_t^p}\right)Y_t - \Phi_t^P(\pi_t, \pi_{t-1}) - \Psi_t^F Y,$$

where $\Phi_t^P(\pi_t, \pi_{t-1}) = \frac{\eta_t}{2\kappa}\left(\log \pi_t - \log \pi_{t-1}^{\iota_p}\pi^{1-\iota_p}\right)^2 Y_t$ is the price adjustment cost and Y is steady state output.

There is also a representative capital firm, which converts equity investment into capital subject to adjustment costs and rents capital services to goods producers with variable depreciation. The optimality condition for capital accumulation implies the following capital price.

$$\log\left(\frac{\pi_t}{\pi_{t-1}^{\iota_p}\pi^{1-\iota_p}}\right) = \mathbb{E}_t\left[\Lambda_{t,t+1}\log\left(\frac{\pi_{t+1}}{\pi_t^{\iota_p}\pi^{1-\iota_p}}\right)\right] + \kappa\left(\mathrm{MC}_t - \frac{1}{\Psi_t^p}\right). \tag{7}$$

where K_t is the aggregate capital stock, ϕ governs adjustment costs, and Ψ_t^k is a capital production efficiency shock akin to an investment specific technology or marginal efficiency of investment shock in Justiniano et al. (2011). The shock follows an AR(1) process with autocorrelation ρ_k and innovation standard deviation σ_k.

2.3 Labor agencies

The labor market features search and matching frictions, with vacancy posting and a matching function, in the spirit of Mortensen and Pissarides (1994) and Pissarides (2000). The timing of events is summarized in Figure 1. Labor agencies post vacancies to hire households and supply labor services to goods producers. Let $J^L(s_t)$ denote the present value to a labor agency of a filled match with a household of productivity s_t

$$J^L(s_t) = (r_t^l - w_t - \Xi^L)\, s_t n_t + \mathbb{E}_t\left[\Lambda_{t,t+1}(1-\zeta)(1-\lambda)(1-P_e)\, J^L(s_{t+1})\right], \tag{8}$$

where r_t^l is the labor rental rate, w_t is the real wage, Ξ^L is the per match maintenance cost, λ is the exogenous separation rate, and P_e is the transition probability into business ownership.[12]

In the standard approach, wages are set through Nash bargaining. Given the large idiosyncratic state space, solving household-specific bargaining is computationally burdensome. Following Gornemann et al. (2016), I therefore use a wage rule of the form

$$\frac{w_t}{w} = \left\{\Psi_t^w\left(\frac{r_t^l}{r^l}\right)\right\}^{(1-\rho_w)}\left\{\frac{w_{t-1}}{w}\left(\frac{\pi_{t-1}}{\pi_t}\right)^{\iota_w}\left(\frac{\pi}{\pi_t}\right)^{1-\iota_w}\right\}^{\rho_w} \tag{9}$$

[12] Ξ^L is introduced to estimate the vacancy posting cost without affecting the steady state of the economy. For a given vacancy posting cost, the maintenance cost is adjusted to fix the expected value of vacancies at the steady state.

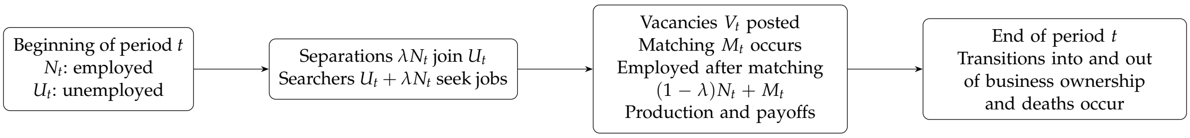


Figure 1: Timeline of events in the labor market.

with $0 < \rho_w < 1$ and $0 < \iota_w < 1$, where Ψ_t^w is an i.i.d. wage markup shock.[13] The wage rule implies nominal wage rigidity, with ρ_w governing its degree. When $\rho_w = 0$, real wages move one for one with the labor rental rate and with the wage markup shock. If $\rho_w > 0$, the nominal wage responds only partially to current inflation and instead is indexed to a weighted average of past and steady state inflation, with ι_w representing the degree of indexation to the previous period's inflation rate and $1 - \iota_w$ represents the degree of indexation to the steady state inflation rate. When current inflation is higher than this weighted average rate used for indexation, it puts downward pressure on the real wage.

Vacancies V_t are pinned down by the free entry condition

$$\iota = \frac{M_t}{V_t} \int J^L(s_t)\, d\mu_t(s_t), \tag{10}$$

where ι is the vacancy posting cost, M_t is the number of matches formed in period t, and μ_t is the cross sectional distribution of households over productivity s_t. This condition equates the vacancy posting cost to the job filling rate, M_t/V_t, times the expected value of a match. The number of matches is given by the matching function specified in Haan et al. (2000). A positive demand disturbance raises the value of a match, with a magnitude that depends on wage rigidity, which stimulates vacancy posting, raises the job finding rate, and lowers unemployment.

2.4 Banks

Banks intermediate equity investment in the model. Following Gertler and Karadi (2011), a continuum of banks $j \in (0,1)$ takes deposits and purchases equity, with balance sheet and net worth given by

$$q_t A_{jt+1}^b = N_{jt} + B_{jt+1}^b, \tag{11}$$

$$N_{jt+1} = R_{t+1}^a\, q_t A_{jt+1}^b - R_{t+1} B_{jt+1}^b, \tag{12}$$

with $R_{t+1}^a \equiv (q_{t+1} + r_{t+1}^a)/q_t$ and $R_{t+1} \equiv 1 + r_{t+1}^b$. A fraction θ_b of banks survives each period. Exiting banks are replaced by entrants endowed with startup funds

[13] Although the shock is i.i.d., the wage rule makes its effects persistent because the real wage adjusts slowly.

equal to a fraction ω of last period's aggregate equity position.

Let $J^b(N_{jt})$ denote the value of a surviving bank. The problem is

$$J^b(N_{jt}) = \max_{\{A^b_{jt+1}, B^b_{jt+1}, N_{jt+1}\}} \mathbb{E}_t \left[\Psi^b_t \Lambda_{t,t+1} \{ (1-\theta_b) N_{jt+1} + \theta_b J^b(N_{jt+1}) \} \right] \tag{13}$$

subject to the balance sheet, the net worth law of motion, and the incentive constraint

$$J^b(N_{jt}) \geq \Delta\, q_t A^b_{jt+1}. \tag{14}$$

The risk premium shock follows

$$\log \Psi^b_t = \rho_b \log \Psi^b_{t-1} + \epsilon_{b,t}, \qquad \epsilon_{b,t} \sim \mathcal{N}(0, \sigma^2_b). \tag{15}$$

A positive innovation raises the valuation of future payoffs and tends to expand bank balance sheets.

In equilibrium the constraint binds and implies aggregate leverage Θ_t

$$q_t A^b_{jt+1} = \Theta_t N_{jt}, \qquad \Theta_t = \frac{\vartheta^n_t}{\Delta - \vartheta^a_t}, \tag{16}$$

where ϑ^a_t and ϑ^n_t are derived in the Online Appendix. Aggregation then yields the evolution of key banking quantities

$$\begin{aligned} \text{equity purchases} \quad & q_t A^b_{t+1} = \Theta_t N_t, & (17)\\ \text{deposits} \quad & B^b_{t+1} = q_t A^b_{t+1} - N_t = (\Theta_t - 1)\, N_t, & (18)\\ \text{aggregate net worth} \quad & N_t = \theta_b \{ (R^a_t - R_t)\Theta_{t-1} + R_t \} N_{t-1} + \omega\, q_{t-1} A^b_t. & (19) \end{aligned}$$

Bank profits are part of aggregate profits and ultimately accrue to the equity mutual fund and business owners.

2.5 Mutual funds

In the background, two mutual funds operate in the model. The equity mutual fund intermediates equity investment to the capital firm and distributes profits to business owners and equity holders. It operates under perfect competition, so it earns zero profits and holds no retained earnings, which implies that the equity price equals the price of capital and the dividend on equity equals after-tax aggregate profits per unit of equity,

$$r^a_t = (1-\tau)(1-\nu)\, \Pi_t / K_t, \tag{20}$$

where $\Pi_t = \Pi_t^f + \Pi_t^b$ is aggregate profits, the sum of profits from the firm side (including labor agencies) and from the banking sector.

The money market mutual fund provides liquidity to the banking sector by investing in bank deposits.[14] It receives a fraction of tax revenue from the government each period, invests in liquid assets, and returns the remainder to households as lump sum transfers. The fund also faces a liquidity preference shock Ψ_t^l that acts like a discount factor shock in its portfolio problem.

2.6 Government

Fiscal authority The fiscal authority collects taxes $\mathbb{T}_t$ and issues one-period government debt B_{t+1}^g to finance government purchases G_t, unemployment benefits D_t, lump sum transfers T_t^g, and contributions to the money market mutual fund C_t^g. Public debt is held by households and the money market mutual fund, and issuance follows

$$\frac{B_{t+1}^g}{B^g} = \left\{ \frac{(1+i_t)/\pi_t \, B_t^g}{(1+i)/\pi \, B^g} \right\}^{\rho_B}, \qquad 0 \leq \rho_B < 1, \tag{21}$$

where B_{t+1}^g is debt outstanding at the end of period t, B^g is its steady state level, and ρ_B is the speed of adjustment.[15]

The government budget constraint is

$$\mathbb{T}_t + B_{t+1}^g = G_t + D_t + \left(\frac{1+i_t}{\pi_t}\right) B_t^g + T_t^g + T_t^{CB} + C_t^g, \tag{22}$$

where $T_t^{CB} = (q_t + r_t^a) A_t^{CB} - (1+i_t)/\pi_t B_t^{CB}$ is the net remittance from the central bank. Lump sum transfers follow

$$T_t^g = \left(1 - \frac{1}{\Psi_t^g}\right) Y, \tag{23}$$

where Ψ_t^g is a transfer shock with autocorrelation ρ_g and standard deviation σ_g. The fiscal authority balances its budget each period by adjusting C_t^g so that (22) holds.[16]

[14] Households hold roughly a tenth of their balance sheets in liquid assets in the data, while the banking sector is highly leveraged and requires a large base of deposit funding. The money market mutual fund stands in for non-household liquidity providers and prevents unrealistically large household deposit shares.

[15] Households and the money market mutual fund treat bank deposits and government bonds as a single liquid asset. Aggregate demand for liquid assets is met by the sum of bank deposits and the bonds issued by the fiscal authority and by the central bank.

[16] The money market mutual fund smooths the transfers to households, so adjustments in C_t^g translate into muted period-by-period changes in household income relative to a setup without the fund. Consequently, the fiscal response has a smaller impact on households than in a standard setup.

Monetary authority The monetary authority sets the nominal interest rate on bonds and deposits, following a Taylor rule with interest rate smoothing subject to an effective lower bound,

$$\frac{1+\hat{i}_{t+1}}{1+\hat{i}} = \left(\frac{1+\hat{i}_t}{1+\hat{i}}\right)^{\rho_R} \left[\left(\frac{\pi_t}{\pi}\right)^{\phi_\pi} \exp\{-\phi_u(u_t - u)\}\right]^{1-\rho_R} \exp(\epsilon_{R,t}), \tag{24}$$

$$i_{t+1} = \max\{0, \hat{i}_{t+1}\} \tag{25}$$

where $\epsilon_{R,t} \sim \mathcal{N}(0, \sigma_R^2)$ is a monetary policy shock. When the effective lower bound binds, the central bank resorts to unconventional monetary policies. First, when the nominal rate is constrained at zero, it purchases equity by issuing bonds as in Gertler and Karadi (2011),

$$A_{t+1}^{\mathrm{CB}} = \Psi_t^{\mathrm{QE}} A^{\mathrm{CB}}, \qquad B_{t+1}^{\mathrm{CB}} = q_t A_{t+1}^{\mathrm{CB}}, \tag{26}$$

$$\log \Psi_t^{\mathrm{QE}} = \rho_{\mathrm{QE}} \log \Psi_{t-1}^{\mathrm{QE}} + \epsilon_t^{\mathrm{QE}}, \tag{27}$$

where Ψ_t^{QE} is a QE shock with autocorrelation ρ_{QE} and innovation standard deviation σ_{QE}.[17] Outside the ELB, the central bank's asset holdings converge back gradually to their steady state level.

The other unconventional policy is forward guidance. Forward guidance operates only when the lower bound binds. I treat an ELB episode as a temporary regime and compute the transition path using an exogenous expected duration for the binding ELB, following Kulish et al. (2014) and Jones (2017). If the expected duration exceeds the endogenous duration implied by fundamentals under no future shocks, the effect is equivalent to anticipated expansionary policy rate innovations over the additional periods. I interpret this gap between exogenous and endogenous durations as forward guidance.

2.7 Solution method

I solve the model using the perturbation method of Reiter (2009). I first compute the steady state of the model using policy function iteration, then linearize the equilibrium conditions around the steady state and apply the perturbation method. Given

[17] Unlike Gertler and Karadi (2011), I model QE as an exogenous AR(1) process rather than through a feedback rule. When I estimate a Taylor-type feedback rule for QE with persistence, the data favor very high persistence, with implied autocorrelations for the QE process very close to one during the ELB episode. In practice, this makes the balance sheet behave much like a highly persistent exogenous shock in the estimated system. For the baseline, I therefore treat QE as an exogenous process with autocorrelation fixed at 0.99 and shut it down in counterfactual simulations by setting $\epsilon_t^{\mathrm{QE}} = 0$, so that the central bank balance sheet remains at its precrisis level. The Online Appendix reports robustness results for a specification in which QE instead follows a Taylor-type feedback rule.

the high dimensionality induced by the idiosyncratic states (asset holdings, earnings, and employment status), I reduce the state space following Bayer and Luetticke (2020), using Chebyshev polynomials for value function reduction and a fixed copula for the distribution reduction.[18] For the model solution at the ELB, I adopt the OccBin algorithm of Guerrieri and Iacoviello (2015) with exogenous expected ELB durations as in Jones (2017). Further details on the numerical method are presented in the Online Appendix.

3 Parametrization

Parametrization proceeds in two stages. First, a subset of parameters is calibrated so that the model's steady state matches moments from micro data on wealth and income. The remaining parameters are estimated with Bayesian methods using aggregate time series to discipline the model's dynamics. Estimation accounts for episodes when the effective lower bound binds and when UMP is active.

3.1 Calibration

Data For micro data on households' wealth and income, I use the 2007 Survey of Consumer Finances, which provides detailed balance sheets and income composition with strong coverage of the upper tail and is the last survey conducted before the Great Recession. Definitions follow the literature. Liquid assets include deposits, bonds, and credit card balances. Illiquid assets include business and financial assets and net housing wealth, with 40 percent of housing counted as illiquid asset, as in Kaplan et al. (2018). Consumer durables are excluded, and households with negative illiquid assets are omitted from the sample. Income is categorized into labor, capital, and transfer income, where labor income consists of wages, capital income includes business and asset income, and transfer income covers social security benefits and pensions. Aggregate targets and national accounting ratios are taken from standard national accounts sources.

Targets and mapping Parameters are chosen so that the steady state aligns with key distributional moments and standard aggregates in the data. In the household block, the calibration ensures that the model reproduces central features of portfolios and wealth concentration, for example the mass of wealthy hand-to-mouth households, the share of borrowers, infrequent adjustment of illiquid assets, and upper decile wealth shares. Labor market parameters are set to match the unemployment rate and

[18] In the Online Appendix, I also solve the model using a time varying copula and present results from that solution.

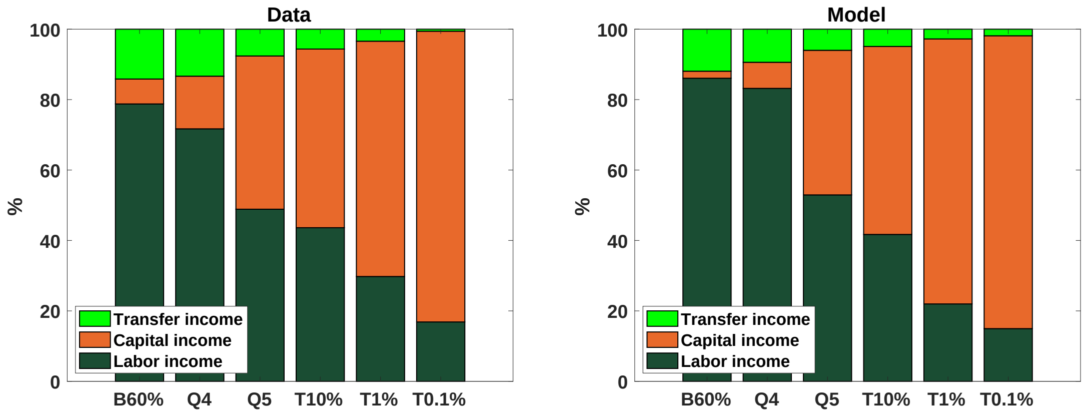
Table 1: Calibrated parameters

Parameter	Value	Description	Reference or targets
Households			
σ	1.5	Relative risk aversion	Standard value
β	0.9932	Household discount factor	Mass of wealthy hand-to-mouth
ξ	3	Inverse Frisch elasticity	Chetty et al. (2011)
ζ	1/180	Probability of death	Average life span 45 years
P_e	0.05%	Prob. of becoming business owner	Bayer et al. (2019)
Labor market			
λ	0.1	Separation rate	Haan et al. (2000)
w	1.2112	SS real wage	Labor share 60%
α	1.7127	Matching efficiency	Vacancy filling 70%
Goods producers			
η	3	Elasticity of substitution	Gornemann et al. (2016)
θ	0.27	Capital share	Capital share 40%
Banks			
Δ	0.3410	Limited enforcement	Bank leverage = 3
ν	0.2380	Profit share to business owners	Gini net worth
Government			
τ	0.30	Tax rate	Data
v	0.4	Replacement ratio	Standard value
ι_b	0.0253	Borrowing premium	Mass with zero assets
$\underline{b}$	1.3006	Borrowing limit	Mass with debt
Central bank			
π	1.0050	Inflation target	Fed target
$1+i$	1.0100	SS nominal rate	Liquid/illiquid ratio
A^{CB}/Y	0.05	CB assets over output	Data
ρ^{QE}	0.99	QE shock autocorrelation	See the Online Appendix

the vacancy filling rate while delivering a reasonable labor share under free entry. The production block yields a capital to output ratio consistent with national accounts and plausible wage and profit shares, with utilization calibrated to a standard depreciation rate. The financial sector calibration follows Gertler and Karadi (2011) and delivers a moderate level of bank leverage and a bank equity share consistent with the data. Government policy parameters are standard or set to match moments related to the distribution of liquid assets. Exact values and the parameter to moment mapping are presented in Table 1 and in the Online Appendix.

Income composition in the data and the model Figure 2 shows that the model reproduces the main pattern of income composition across wealth groups in the data. In both the SCF and the model, labor income is the primary source of income below the top of the wealth distribution, capital income dominates at the very top, and transfer income is small and concentrated in lower groups. For the top wealth quintile Q5, the labor income share is 53 percent in the model versus 49 percent in the data. For the top 10 percent, 1 percent, and 0.1 percent, the corresponding model shares are 42, 22, and 15 percent, compared with 44, 29, and 17 percent in the data. The model therefore captures the sharp decline in labor income share and the rise in capital income share as wealth increases, with some understatement of labor income at the very top. Overall, the model tracks income composition across the wealth distribution reasonably well in a way that is relevant for the transmission of policy.

Figure 2: Income composition in the data and the model



Notes: The figure shows the composition of household income by wealth group in the data (left panel) and in the model (right panel). Each bar reports the shares of labor income, capital income, and transfer income in percent. Groups are ordered by net worth on the horizontal axis: B60% denotes the bottom 60 percent of the wealth distribution, Q4 the fourth quintile, Q5 the top quintile excluding the top 10 percent, T10% the top 10 percent, T1% the top 1 percent, and T0.1% the top 0.1 percent. The data are from the 2007 Survey of Consumer Finances.

3.2 Estimation

To discipline the model dynamics with the data, I estimate the remaining model parameters using Bayesian methods. A key challenge in estimating a HANK model is the time required to update the solution for each parameter draw due to the large idiosyncratic state space. To address this, I follow Bayer and Luetticke (2020) and estimate only parameters that do not affect steady state objects, such as households' value functions and the time invariant distribution. This approach allows me to update only a small subset of Jacobian entries of the linearized system at each iteration, which makes estimation feasible. In addition, I evaluate the likelihood using the inversion filter as in Cuba-Borda et al. (2019), which speeds up estimation.[19]

Another challenge in estimation is the occasionally binding effective lower bound. The solution at the bound depends on the expected duration of the ELB, so these durations must be determined while the policy rate is constrained. Because jointly identifying durations and shocks is difficult and time consuming, I adopt the approach in Kulish et al. (2014), Jones (2017), and Jones et al. (2022) and assume a sequence of exogenous expected ELB durations during estimation.[20] I estimate these durations jointly with the structural shocks and the other parameters using the randomized blocking scheme of Chib and Ramamurthy (2010), updating structural parameters and expected durations in separate blocks.

[19]Even with state space reduction, the model is large and constructing the state transition matrix for the Kalman filter is time consuming.

[20]Alternatively, one can infer endogenous expected ELB durations in each quarter of the ELB episode as in Guerrieri and Iacoviello (2017), Atkinson et al. (2020), and Cuba-Borda et al. (2019), but this requires repeated inversion of very large matrices and is computationally burdensome.

Table 2: Prior and posterior distributions of friction parameters

Symbol	Description	Prior			Posterior		
		Density	Mean	Std	Mode	10%	90%
		Frictions					
κ	Slope of Phillips curve	Gamma	0.10	0.02	0.0525	0.0340	0.0765
ι_p	Price indexation	Gamma	0.50	0.15	0.1219	0.0670	0.2069
ρ_w	Wage autocorrelation	Beta	0.50	0.20	0.7982	0.7065	0.8654
ι_w	Wage indexation	Beta	0.50	0.15	0.1835	0.1132	0.2639
ϕ	Capital adjustment cost	Normal	30.00	5.00	50.017	49.193	51.184
ι	Vacancy posting cost	Gamma	0.05	0.02	0.0317	0.0189	0.0495

Notes: Posterior modes and 10–90 percent credible intervals shown.

Priors are constructed from the New York Fed Primary Dealer Survey.[21] In the priors, the mode of the expected ELB duration is higher in 2012 and 2013. The posteriors show only slight increases during these years. Overall, posterior modes range from 6 to 8 quarters through 2014 and then fall to one or two quarters in the final year of the episode, reflecting tight priors.

For the estimation of structural parameters, I use the following set of variables spanning from Q1 1992 to Q4 2018 as observables.

$$\left[\Delta \log Y_t,\ \Delta \log C_t,\ \Delta \log \tilde{I}_t,\ \log \pi_t,\ \log(1+i_t),\ \log u_t,\ \Delta \log w_t,\ \Delta \log T_t^g,\ \Delta \log \Pi_t,\ \Delta \log A_t^{\mathrm{CB}}\right], \tag{28}$$

which correspond, respectively, to output growth, consumption growth, investment growth, the inflation rate, the gross nominal interest rate, the unemployment rate, real wage growth, growth in lump sum transfers, growth in corporate profits, and growth in central bank assets. Data are from BEA NIPA, BLS, and the Federal Reserve Board, compiled via FRED.

To apply the inversion filter, the number of structural shocks must equal the number of observables. These shocks are the MMMF liquidity preference shock Ψ_t^l, total factor productivity Z_t, the price markup shock Ψ_t^p, the wage shock Ψ_t^w, the investment technology shock Ψ_t^k, the banks' risk premium shock Ψ_t^b, the transfer shock Ψ_t^g, the monetary policy shock $\epsilon_{R,t}$, the fixed cost shock Ψ_t^F, and the QE shock Ψ_t^{QE} to central bank asset holdings.

I use standard priors from the literature for structural parameters, following Smets and Wouters (2007) and Justiniano et al. (2011). Table 2 reports priors and posteriors for parameters that characterize frictions in the model. The posterior indicates high price and wage rigidity, low vacancy posting costs, and high capital adjustment costs.

[21] The survey asks primary dealers about the timing of liftoff from the effective lower bound and reports the distribution of responses. Results before January 2011 are unavailable, so I use the January 2011 distribution for 2009–2010 and truncate the priors at four years.

The Phillips curve slope has a posterior mode near 0.05, implying an average price duration of about six quarters. The estimated wage rigidity is substantial, with about one fifth of the real wage adjusting to productivity changes at the posterior mode. Vacancy posting costs are estimated to be low relative to the prior. The model's high wage rigidity is consistent with stable real wages and their weak correlation with output, while profits are volatile and positively correlated with output due to high wage rigidity and low vacancy posting costs. The posterior mode for capital adjustment costs is higher than the prior mode, possibly due to the presence of banks and a financial accelerator that amplifies investment volatility.

Additional data construction details, the full set of priors and posteriors, and the complete estimation algorithm are provided in the Online Appendix.

4 Monetary policy in the model

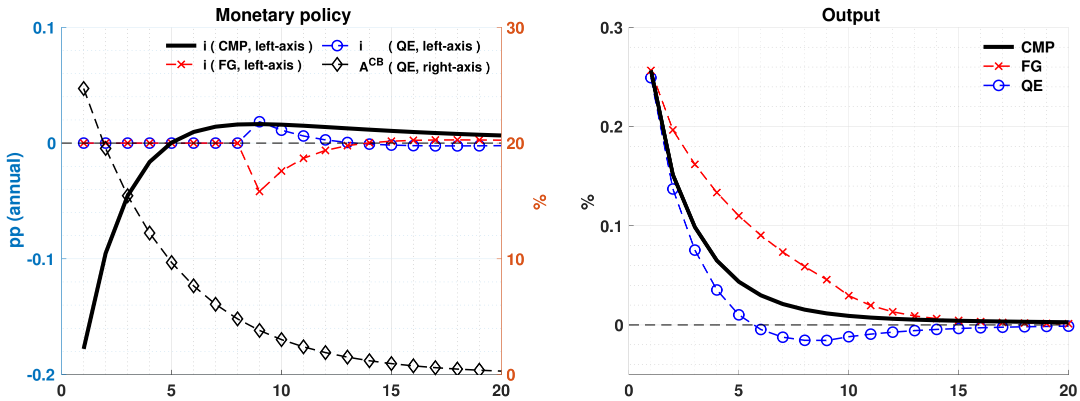
Before turning to the counterfactual analysis for the ELB episode, I first examine how different types of monetary policy, both conventional and unconventional, work in the model. Specifically, this section studies the transmission of QE, and FG in the estimated HANK model using impulse responses of aggregates and of consumption across the wealth distribution. The goal is to provide a transparent benchmark for the mechanisms that underlie the counterfactual exercises in the next sections.

As a reference point for comparison, I show the effects of a CMP shock, which is a contemporaneous shock to the policy rate of 25 basis points at an annual rate.[22] For QE and FG, I choose the size of the shocks so that the response of output on impact matches the response to the CMP shock. The QE shock is a contemporaneous expansion of the central bank balance sheet. Central bank assets jump up on impact and then follow an exogenous first order autoregressive process with persistence equal to the interest rate smoothing parameter in the Taylor rule.[23] In both QE and FG, the nominal policy rate is held at its steady state level for eight quarters, after which the estimated Taylor rule is allowed to operate. The forward guidance shock is implemented as an anticipated negative deviation from the Taylor rule in quarter

[22]The model generates empirically plausible responses of aggregate variables to conventional monetary policy shocks. One caveat is that it does not produce the hump-shaped impulse responses often found in the data. Auclert et al. (2020b) show how hump-shaped aggregate responses can arise in models with sticky expectations and how to implement such mechanisms using the sequence-space Jacobian method. By contrast, the present paper relies on a state-space solution method, so embedding such mechanisms and generating hump-shaped responses is not a trivial extension.

[23]The Online Appendix reports impulse responses for a more persistent QE process with autoregressive parameter of 0.99, which is the value used in the counterfactual analysis, and for an alternative specification in which QE follows a rule that is analogous to the Taylor rule with interest rate smoothing.

Figure 3: CMP, QE, and FG shocks in the model



Notes: The left panel shows impulse responses of the annualized nominal policy rate and central bank assets to three monetary policy shocks. The black solid line plots the policy rate after a conventional monetary policy shock (CMP). The red crosses show the policy rate under forward guidance (FG), and the blue circles show the policy rate under quantitative easing (QE). The black diamonds (right axis) plot the path of central bank assets A^{CB} in the QE case. In the QE and FG shocks, the policy rate is kept at its steady state level for eight quarters, while central bank assets remain at their steady state level in CMP and FG. The right panel plots the corresponding impulse responses of output under CMP (black solid), FG (red crosses), and QE (blue circles). All responses are reported as deviations from steady state, in percentage points at an annual rate for the policy rate and in percent for output and central bank assets.

nine, so that agents expect a one time cut in the policy rate two years ahead.[24] In the CMP and FG analysis, the central bank balance sheet is kept at its steady state level,

4.1 Aggregate effects of CMP, QE, and forward guidance

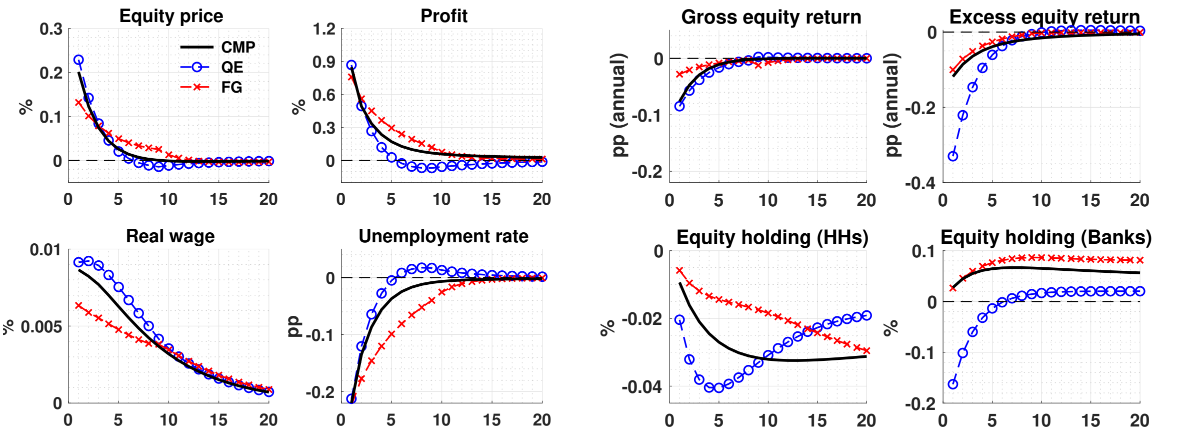
Figure 3 plots the paths of the monetary policy instruments and the corresponding responses of output. The CMP shock lowers the policy rate on impact and the rate then gradually converges back to its steady state level as the Taylor rule brings policy back to normal. Under QE and FG, the policy rate is initially fixed at the steady state level and starts to move only after the peg for eight quarters ends. In the QE case, the balance sheet expansion raises inflation and output, so the Taylor rule implies a small increase in the policy rate once the peg is lifted. In the FG case, there is an anticipated negative shock to the policy rate in quarter nine, so the policy rate falls when the peg ends and agents internalize this future cut throughout the horizon.

The right panel of Figure 3 shows that, despite the normalization of shock sizes to deliver very similar output responses on impact, the dynamics of output differ across policy tools. CMP generates a standard pattern in which output rises on impact and then monotonically returns to steady state.[25] FG produces a more persistent expan-

[24]In the estimation and the counterfactual analysis, forward guidance is modeled as an exogenous sequence of expected ELB durations rather than as an anticipated future policy rate shock. When the estimated Taylor rule and the state of the model economy would imply lift off absent the ELB constraint, an extension of the expected ELB duration is equivalent to an expected future cut in the policy rate, so the two implementations generate similar dynamics for the model economy.

[25]The effects of conventional monetary policy in the model are comparable to existing estimates.

Figure 4: Aggregate effects of CMP, QE, and FG shocks



Notes: The figure shows impulse responses of asset prices, returns, labor market variables, and equity holdings to the three monetary policy shocks. The black solid line corresponds to a conventional monetary policy shock (CMP), the blue circles to a quantitative easing shock (QE), and the red crosses to a forward guidance shock (FG). The top row displays the responses of the equity price, firm profits, the gross equity return, and the excess equity return. The bottom row displays the responses of the real wage, the unemployment rate, equity holdings of households, and equity holdings of banks. All variables are shown as deviations from steady state, with equity prices, profits, wages, and equity holdings in percent, unemployment in percentage points, and gross and excess equity returns in percentage points at an annual rate.

sion. Output rises by roughly the same amount on impact as under CMP, but the response decays slowly because the expected future cut in the policy rate keeps real rates low for an extended period. QE instead yields a front loaded expansion with a modest reversal. Output increases on impact when central bank purchases raise aggregate demand, but the response becomes slightly negative at medium horizons.

These differences in output responses across monetary policy tools are mainly due to their asymmetric effects on the financial sector in the model. The first order direct effect of a QE operation is on the equity price. As the equity price goes up, the valuation of the financial sector's net worth rises, which is positive for banks. At the same time, however, the higher equity price lowers the gross rate of return on equity and compresses the excess return on bank assets, that is, the gross rate of return on equity minus the real rate on deposits, which reduces banks' equity investment. Over time, this crowding out of private intermediation propagates and puts downward pressure on real activity as the central bank balance sheet normalizes back to its steady state level. Thus, the expansion induced by the monetary authority is short-lived. In contrast, in the absence of general equilibrium effects, a cut in the nominal rate raises the excess equity return. The banking sector then expands its balance sheet, and this expansion propagates through the financial accelerator channel incorporated in the model. An anticipated future cut operates in the same way through the

For example, McKay and Wolf (2023) find that a monetary policy shock that leads to an average 1 percent increase in the level of output over two years raises consumption by about 0.9 percent on average over the same horizon. In my model, a shock of similar size leads to an average 0.8 percent increase in consumption over two years.

expectation channel. CMP and FG therefore crowd in private investment, in contrast to QE, and this crowding in leads to more persistent expansionary effects through private intermediation.

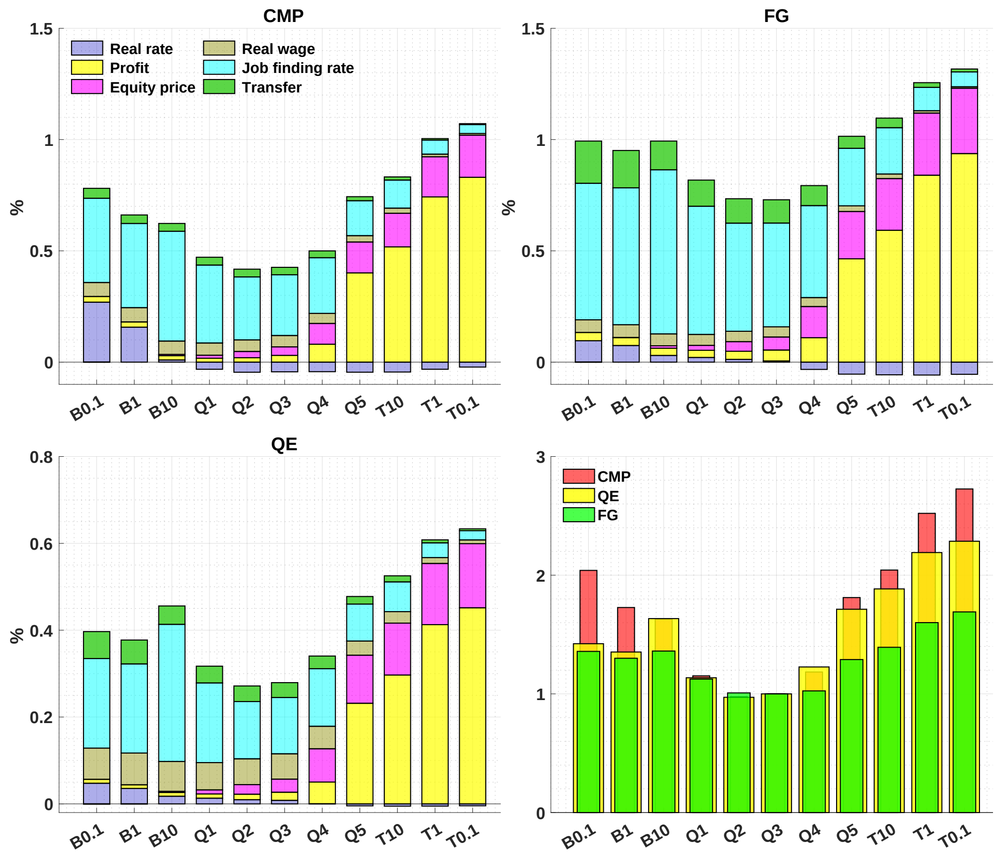
Figure 4 shows the resulting aggregate general equilibrium effects of the mechanisms described above. On impact, QE generates the largest increase in the equity price. As shown in the right panel of the figure, this rise in the equity price, combined with the fixed nominal rate, lowers the excess equity return by a large amount and crowds out banks' equity investment.[26] This dynamic pattern shows up in the responses of other aggregate variables, such as output, profits, and the unemployment rate. The positive short run response is short-lived and reverses sign at medium horizons. The real wage response is an exception because it features a high degree of rigidity, so the short run response persists. In the case of CMP and FG, the equity price and the excess equity return also fall, but these movements are not first order direct effects of the policies and instead reflect the expansion of private intermediation by banks. Because banks respond strongly to an increase in the spread induced by a fall in the short term rate, they raise their equity holdings and the equity price. These expansions propagate persistently and lead to persistent responses of aggregate variables, such as profits and the unemployment rate. In the case of FG, where the expansion operates through the expectation channel, the responses are even more drawn out.

Households play a limited role in these shock propagations in the model. QE directly affects only households that hold equity, and in the calibration this is a small group at the top of the wealth distribution that adjusts infrequently. A significant fraction of households are hand-to-mouth and respond mainly through income channel rather than through asset return channels, as in other HANK models.[27] The main amplification therefore comes from the interaction between asset returns and bank balance sheets, in line with the financial accelerator mechanism embedded in the banking sector.

[26]This mechanism is consistent with evidence that bank equity valuations and profitability are very sensitive to the level and slope of the yield curve and to monetary policy shocks that affect term premia and risk spreads. See Altavilla et al. (2018) for evidence on the response of bank equity prices and profitability to monetary policy shocks, and Bhattarai and Neely (2022) for a survey of the empirical literature on unconventional monetary policy.

[27]See Kaplan et al. (2018) for a discussion of the role of hand-to-mouth households in the transmission of monetary policy. In the Online Appendix I report impulse responses in variants of the model that shut down the direct effects of UMP on households by fixing either the equity price and equity return or the real interest rate faced by households. Output responses change only slightly in these exercises. In contrast, when I shut down the transmission to banks by fixing bank equity holdings at their steady state level, the aggregate responses change markedly. The Online Appendix also shows that in a RANK version of the model, the role of households is larger in the transmission of UMP. Shutting down households' response in the RANK version leads to bigger differences in aggregate responses to UMP shocks.

Figure 5: Distributional effects of CMP, QE, and FG shocks



Notes: The figure shows cumulative consumption responses over twenty quarters across the household wealth distribution for three monetary policy shocks. Each bar corresponds to households grouped by wealth on impact: B0.1 denotes the bottom 0.1 percent of the wealth distribution, B1 the bottom 1 percent, B10 the bottom 10 percent, Q1–Q5 the first to fifth quintiles, and T1 and T0.1 the top 1 percent and top 0.1 percent. The top row and the bottom left panel report the cumulative responses in CMP, FG, and QE and decompose them into contributions from the real interest rate, profits, equity prices, real wages, the job finding rate, and transfers. The bottom right panel reports the cumulative responses in CMP, FG, and QE relative to the response of the middle quintile Q3, so values are interpreted as multiples of the Q3 response. In all other panels the units are percent deviations from steady state consumption.

4.2 Distributional effects across the wealth distribution

Distributional effects depend both on these aggregate responses and on the heterogeneity of households in the model. To focus on the short and medium run, I measure, for each wealth group, the cumulative response of consumption over the first twenty quarters following the shock. This measure is useful for interpreting the welfare results from the counterfactual exercises in the next section, which consider sequences of shocks during the ELB episode rather than a single disturbance. Figure 5 reports the cumulative consumption responses across the wealth distribution for CMP, QE, and FG, together with a decomposition into contributions from different aggregate variables.[28]

[28] The decomposition is constructed by feeding households the path of one variable at a time, solving their problem backward, and moving the distribution forward using the optimal policy rules and the transition matrix while keeping all other variables at their steady state levels.

All three policy tools raise consumption for every wealth group. QE has the smallest overall effect, reflecting its short-lived aggregate expansion, while FG has the largest effect because of its very persistent impact on output and labor market conditions. Across the wealth distribution, the cumulative responses are broadly U-shaped. Households at the bottom and at the top of the wealth distribution gain more than the middle. This pattern reflects the fact that households at the very top benefit from higher profits and equity prices, while households near the bottom gain mainly through improved labor market prospects and, for borrowers, through lower real interest rates. The decomposition in Figure 5 shows that for households in the top decile and above, the main drivers of consumption gains are higher profits and equity prices, whereas for the rest of the distribution the dominant contribution comes from the job finding rate. Real wages play a minor role because they are highly rigid and respond only modestly to monetary policy shocks.

Although the qualitative U-shaped pattern is common across tools, the degree of unevenness differs. Policies with more persistent aggregate effects generate relatively flatter distributional profiles.[29] FG produces the most even pattern, with only a mild U shape, so the gains in cumulative consumption are relatively similar across the wealth distribution. CMP and QE deliver more pronounced U-shaped profiles. For QE, the strong but short run increase in equity prices and profits raises consumption of wealthy households by more than in the FG case, while the smaller and less persistent improvement in labor market conditions limits the gains for the rest of the distribution. For CMP, the cut in the real interest rate benefits borrowers at the bottom of the distribution and implies a redistribution away from wealthy savers. At the same time, higher bank net worth and profitability strengthen the response of consumption at the top. The result is the most pronounced U-shaped profile among the three tools, with relatively large gains for both poor and rich households and the smallest gains for the middle.

The analysis in this section is based on one time monetary policy shocks at the steady state. In the next sections, I turn to the counterfactual exercises that remove UMP during the actual ELB episode and evaluate the effects of the full sequence of estimated shocks and policies on aggregates, welfare, and inequality.

5 Effects of UMP during the ELB

This section quantifies the effects of unconventional monetary policy (UMP) during the effective lower bound (ELB) episode in the model. I compare the estimated baseline economy, which features asset purchases and forward guidance as observed

[29] See the Online Appendix for welfare effects of QE shocks under alternative persistence.

in the data period, to a counterfactual economy with the same realized shocks and the same initial distribution but without UMP during the ELB episode. In the counterfactual, the central bank keeps its asset holdings at their precrisis level and does not provide forward guidance, so expected ELB durations are determined endogenously by the model state and shocks each period. The policy rule resumes as soon as fundamentals warrant liftoff.[30]

5.1 Simulation of baseline and counterfactual economies

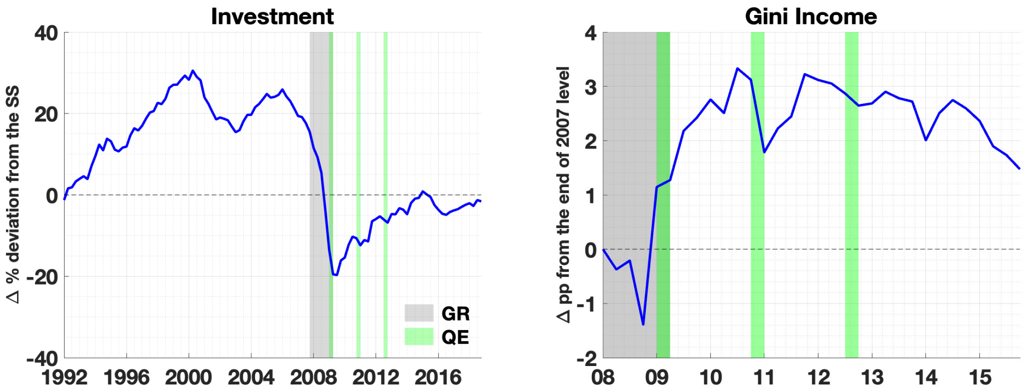
Before turning to results, I summarize how the distributional and welfare objects are constructed and how the baseline and counterfactual paths over the ELB episode are obtained. At the start of the ELB episode in 2009 Q1, I initialize the aggregate state and the cross sectional distribution using the baseline estimation. For each subsequent quarter, in both the baseline economy and the counterfactual that shuts off UMP, I take the current aggregate state and the current cross sectional distribution from the previous step of the simulation. To construct households' optimal choices and welfare for that quarter, I form a deterministic continuation path with no future shocks of length 500 quarters.[31] Given that path for the variables that enter households' problem, I solve the household problem by backward iteration along the no shock path back to the current quarter, which delivers current period policy functions. I then apply those policy functions to the current distribution and update the distribution one period forward using the realized shocks. Because UMP is shut off only in the counterfactual economy, the aggregate states and distributions in the two simulations gradually diverge even though they are hit by the same realizations of the structural shocks.

I carry out this procedure for each quarter of the ELB episode and for each posterior draw. Distributional objects such as group level consumption, income, inequality measures, and welfare are computed from the simulated cross sectional distributions in the baseline and counterfactual economies. Aggregate effects are computed as differences between the baseline and counterfactual time series implied by the estimated model when UMP is removed, holding all other shocks at their realized values in the baseline economy. Welfare is measured in consumption equivalents by comparing lifetime utility in the baseline and counterfactual for households grouped by their wealth in 2009 Q1, holding group composition fixed over the episode. A more detailed description of the algorithm, including examples of no future shock paths, is presented in the Online Appendix.

[30]Endogenous ELB durations are computed using a shooting algorithm, following Guerrieri and Iacoviello (2015).

[31]The procedure assumes the economy converges back to the steady state within 500 periods in the absence of future shocks.

Figure 6: Great Recession and the evolution of investment



Notes: The left panel reports investment as a percent deviation from its steady state level. The right panel displays an inequality series for reference; distributional analysis appears in later sections. The gray bar marks the Great Recession.

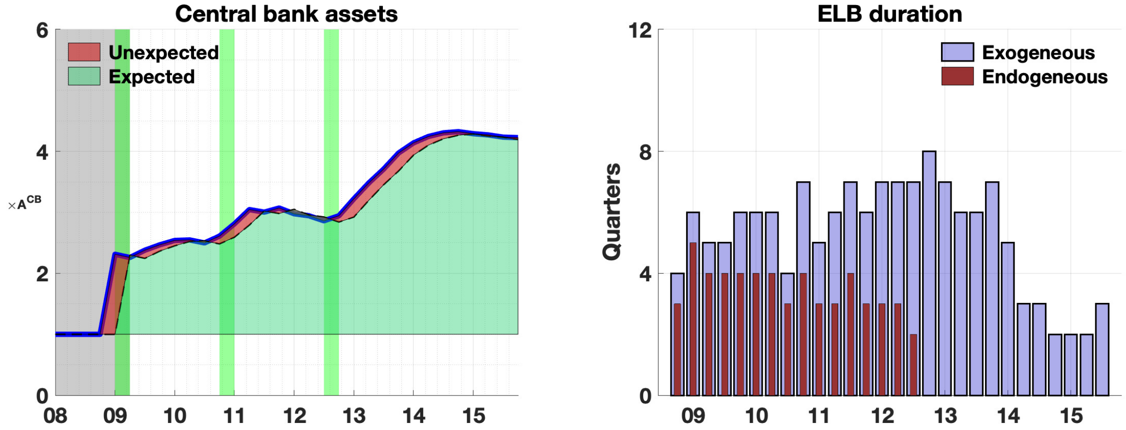
5.2 Great Recession in the model

At the posterior mode, the downturn is driven by a sequence of large adverse risk premium and productivity shocks in 2008 and 2009, as shown in Figure OA4. These shocks depress intermediation, equity values, investment, and output. Investment falls by more than 20 percent relative to its precrisis level at the trough in the model, as shown in Figure 6. The economy then recovers gradually during the period when the central bank expands its balance sheet and maintains the policy rate at the ELB. Inequality, measured by the income Gini index, rises over the same period. The simulation exercises below investigate how UMP has affected the evolution of aggregate variables and inequality.

5.3 UMP in the model

Figure 7 summarizes how UMP is implemented during the ELB episode in the simulations. The balance sheet expansion in the left panel shows the scale and timing of the asset purchases relative to the precrisis level, together with the expected path implied by the estimated QE process. The right panel reports expected ELB durations at the posterior mode, distinguishing between the exogenous durations used in estimation and the endogenous durations implied by model fundamentals under the same policy rule without forward guidance. Periods in which the exogenous expected duration exceeds the endogenous duration correspond to additional expected accommodation. From mid 2012 onward, the policy rate remains at the bound even though fundamentals would permit liftoff. In the model, this is equivalent to expected future expansionary interest rate shocks that are studied in the previous section.

Figure 7: UMP in the model



Notes: The left panel shows the central bank balance sheet as a ratio to its precrisis level at the end of 2007. The blue line traces the realized path of asset holdings, with the red area representing the contribution of unexpected contemporaneous shocks and the green area representing the expected level implied by the contribution of past realizations of shocks. The right panel reports expected ELB durations, displaying both the exogenous expected durations used in the estimation and the endogenous durations implied by model fundamentals under the same policy environment. All series, including the exogenous and endogenous expected ELB durations, are evaluated at the posterior mode.

5.4 Aggregate effects of UMP

Unconventional monetary policy in the model raises aggregate activity during the ELB episode. Figure 8 summarizes the model implied effects of UMP on aggregate variables by comparing the baseline to the counterfactual without UMP. Each panel plots the difference between the two simulated paths, in percent for level variables and in percentage points for the unemployment rate. The black solid line shows the effect at the posterior mode, and the gray shaded band shows the tenth to ninetieth percentile range across posterior draws.[32] Positive values indicate that the variable is higher under UMP than under no UMP.

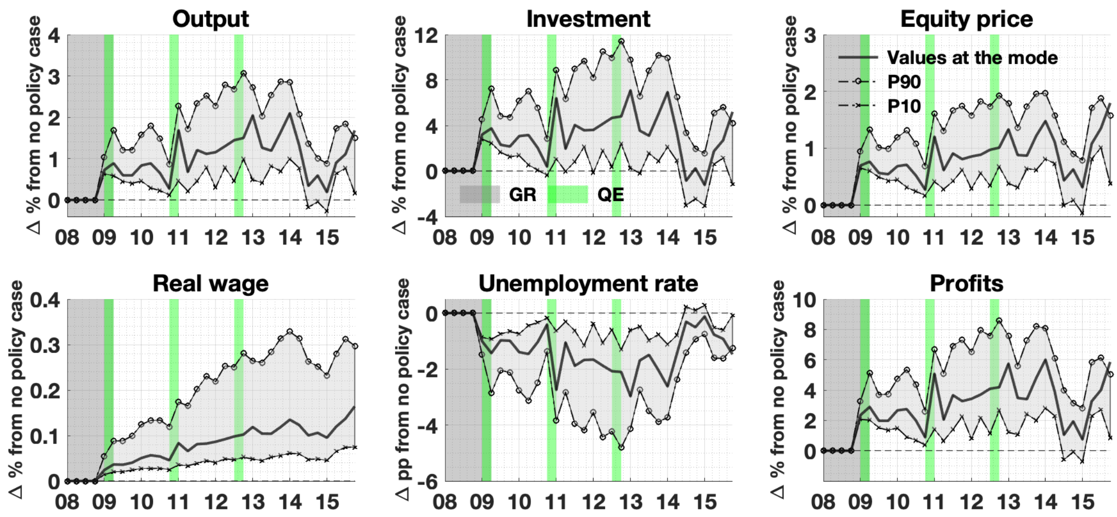
Unconventional monetary policy interventions operate first through asset markets and the financial sector. Balance sheet expansions reallocate portfolios from bonds and deposits toward productive capital, raising equity valuations and increasing investment. Forward guidance adds inflationary pressure that lowers real rates, stimulating investment via portfolio reallocation as well.[33] On average over the ELB episode, investment is about 3 percent higher under UMP than in the counterfactual without UMP. Equity prices are about 1 percent higher on average.

However, the positive effects of UMP are not limited to financial markets. Through

[32] To construct these series, for each posterior draw I recover structural shocks using the inversion filter and simulate two economies, the baseline with the estimated monetary policy and a counterfactual without UMP. At each date the effect reported in Figure 8 is the difference between the two simulated paths for the same variable.

[33] High frequency evidence shows that, even at the ELB, medium and longer maturities remained responsive to policy communications, so shifts in the expected path of short rates through forward guidance can lower real rates and support demand. See Swanson and Williams (2014).

Figure 8: Aggregate effects of UMP



Notes: For level variables X, each panel reports $100 \times (X^{\text{UMP}}/X^{\text{NoUMP}} - 1)$ in percent. For the unemployment rate I report percentage point differences $u^{\text{UMP}} - u^{\text{NoUMP}}$. The black solid line is the posterior mode. The gray shaded band is the tenth to ninetieth percentile range across posterior draws. Positive values mean higher outcomes under UMP than under no UMP.

general equilibrium, they raise aggregate demand and production, with average effects on output of about 1 percent.[34] A boost in aggregate demand and production increases labor demand, which leads either to higher real wages or to a fall in the unemployment rate. Because the posterior implies a high degree of wage rigidity, real wages move little, with an average increase near 0.1 percent, so higher labor demand shows up mainly as lower unemployment. In the simulations, the unemployment rate is lower by about 1.4 percentage points on average over the ELB episode. As labor input increases with lower unemployment and little change in the real wage, profits rise substantially, by roughly 3 to 4 percent on average, with peaks near 8 percent in some quarters.

The magnitudes of these aggregate effects are in line with empirical assessments of unconventional policy during the ELB, once one accounts for differences in metrics and policy scopes. For example, Chung et al. (2011) estimate real GDP nearly 3 percent higher by late 2012 and the unemployment rate about 1.5 percentage points lower following asset purchases, which is comparable to the unemployment effects in the model. Engen et al. (2015) attribute to the combination of asset purchases and forward guidance a peak reduction in unemployment of about 1.2 percentage points and a peak boost to real GDP growth of about 0.5 percentage point. The average level

[34] Higher equity investment translates into more capital accumulation in the model. On the household side, an increase in equity prices generates wealth effects for households that adjust their equity holdings. A fall in the gross rate of return on equity and lower real rates induced by UMP also raise household consumption via intertemporal substitution. The magnitudes of these effects are relatively small in HANK models, as discussed in the previous section, because most assets in the model are held by a small fraction of wealthy households at the top of the distribution.

effect on output in the model over 2009–2015 is of a similar order after translating their growth rate metric to levels. Focusing on QE in isolation, Kim et al. (2023) find that a $500 billion purchase shock raises real output by about 1.2 percent and lowers unemployment by about 0.5 percentage point at the peak. Some of the average level effects in the model are larger than QE only estimates because the counterfactual removes both balance sheet policy and forward guidance during the ELB. A formal decomposition by tool is presented later in the paper.

5.5 Welfare effects of UMP

This subsection investigates how the aggregate effects presented above translate into welfare across the household distribution during the ELB episode. Welfare is measured in consumption equivalent units at the beginning of the episode. For a given household state $x = (a, b, s, e)$ at 2009Q1, the consumption equivalent is the constant fraction $\lambda(x)$ that scales the counterfactual path of consumption in the economy without UMP so that the household is indifferent between living in the baseline with UMP and living in the counterfactual without UMP.[35] Formally,

$$\lambda(x) = \left\{ \frac{V^{\text{UMP}}_{2009\text{Q1}}(x)}{V^{\text{NoUMP}}_{2009\text{Q1}}(x)} \right\}^{\frac{1}{1-\sigma}} - 1, \tag{29}$$

where $V^{P}_{2009\text{Q1}}(x) = \sum_{\tau=0}^{T} \beta^{\tau}\, \mathbb{E}\left[u\left(c^{P}_{2009\text{Q1}+\tau}(x)\right)\right]$ for $P \in \{\text{UMP}, \text{NoUMP}\}$.[36] I report group means in percent for wealth groups defined from the 2009Q1 cross section. Positive values indicate that the baseline with UMP is preferred to the counterfactual without UMP.

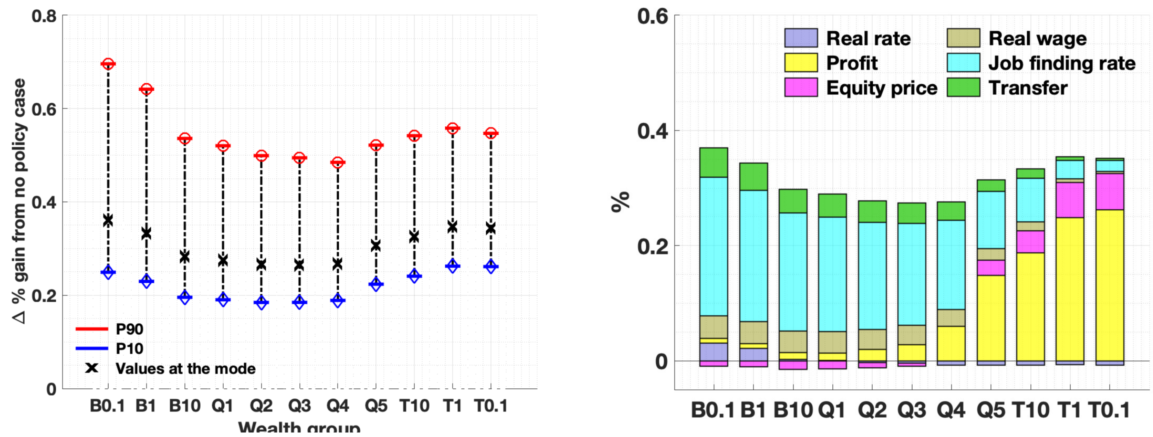
Before turning to results, I summarize how the objects that enter the consumption equivalent calculation are constructed. The evaluation date is 2009Q1 and the comparison is made at that date for each posterior draw. For each policy environment $P \in \{\text{UMP}, \text{NoUMP}\}$ and each calendar date τ in the sample, I first compute a path for the aggregate variables that enter the household problem with the assumption of no future shock (the no shock path), with length sufficient to ensure convergence. Using this no shock path I solve the household problem at date τ by backward iteration along the path and obtain current policy functions $c^{P}_{\tau}(x)$, $a'^{,P}_{\tau}(x)$, and $b'^{,P}_{\tau}(x)$.

I then evaluate the value function at the same date using those policies together with the realized $V^{P}_{\tau+1}$ that has already been computed at date $\tau + 1$. This two step recursion starts at 2018Q4 and proceeds backward one quarter at a time to 2009Q1.

[35] Under GHH preferences, "consumption" denotes the consumption–hours composite that enters period utility.

[36] I set $T = 500$ periods to ensure convergence back to the steady state in the no shock computation.

Figure 9: Welfare effects of UMP: Consumption equivalents



Notes: The figure displays welfare gains from UMP in consumption equivalent units evaluated at 2009Q1. The left panel shows the tenth to ninetieth percentile range across posterior draws for each wealth group, with the black line marking the posterior mode. The right panel reports a decomposition at the posterior mode; colored boxes indicate additive contributions by channels and sum to the dot for each group. Group labels are: B0.1, B1, B10 for the bottom 0.1 percent, 1 percent, and 10 percent; Q1 to Q5 for wealth quintiles; T10, T1, T0.1 for the top 10 percent, 1 percent, and 0.1 percent. Units are percent; positive values indicate that the baseline with UMP is preferred to the counterfactual without UMP.

At each step the current policies are computed along the no shock path from date τ, while the value calculation uses the realized next period value that embeds the shocks that occur in the next period. The same construction is performed under the baseline policy environment and under the counterfactual without UMP. At 2009Q1 this delivers $V^{\text{UMP}}_{\text{2009Q1}}(x)$ and $V^{\text{NoUMP}}_{\text{2009Q1}}(x)$ for every household state in the cross section. I then convert the utility difference into a consumption equivalent $\lambda(x)$ using the definition above and report averages by wealth group defined at 2009Q1. Further details are provided in the Online Appendix.

Figure 9 summarizes the distribution of welfare gains across wealth groups. The left panel shows the full posterior distribution for each group, with the black line denoting the posterior mode and the dotted band indicating the 10th to 90th percentile range. At the posterior mode, welfare gains are positive for all groups, with an average gain of about 0.27 percent of lifetime consumption.[37] The cross sectional profile is modestly U-shaped.[38] Households at the bottom and at the top of the

[37]For comparison, classic estimates of the welfare cost of business cycles in representative-agent models are typically below 0.1 percent of lifetime consumption (Robert E. Lucas, 1987), while heterogeneous-agent analyses such as Storesletten et al. (2001) obtain average welfare costs on the order of 0.6 percent and larger losses for households with little savings. The 0.27 percent welfare gain from UMP in this paper is thus economically meaningful but well within the range found in the literature for macroeconomic policies.

[38]McKay and Wolf (2023) survey the empirical evidence on the incidence of monetary policy across households and conclude that the overall consumption effects of monetary easing are fairly uniform across the distribution. This perspective is consistent with the mild U shaped profile of consumption equivalent welfare gains in the model: UMP delivers broad based gains, and the cross sectional dispersion in gains is small relative to the aggregate benefit.

wealth distribution enjoy above average gains, while the middle quintiles have the smallest gains. The difference between the bottom 1 percent and the middle 60 percent is about 0.06 percentage points, and the difference between the top 1 percent and the middle 60 percent is of similar magnitude. These patterns are robust across posterior draws, as indicated by the dotted band.[39]

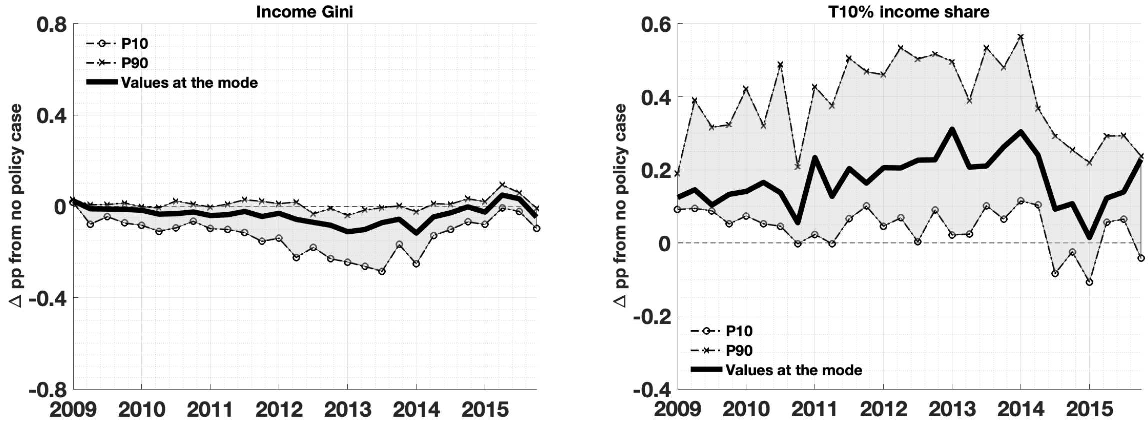
The channels behind these welfare differences align with the aggregate responses documented earlier. At the bottom of the wealth distribution, a larger share of households are unemployed, so they benefit more from the improvement in job finding that follows from higher aggregate demand. In the middle of the distribution, welfare gains are more muted because real wages move little in the short run and these households rely heavily on labor income. At the top of the distribution, welfare gains are stronger because profits and equity prices rise under UMP and these households hold a larger share of equity and business income. The right panel of Figure 9 makes this mapping explicit by showing the additive contributions of these channels at the posterior mode, which sum to the total consumption equivalent for each group.

Compared to the welfare effects of UMP shocks at the steady state shown in Figure OA15, two differences stand out. First, the U-shaped responses are less pronounced in the counterfactual simulations, with more muted effects at the top. Second, the contribution of lump sum transfers is smaller. In contrast to the IRFs for a one time shock at the steady state, the counterfactuals face a sequence of policy shocks during the ELB episode. Unconventional monetary policy is therefore more persistent, which, as discussed in the previous section, makes the distributional effects flatter. In the counterfactual simulations, the no UMP path is also far from the steady state. In the absence of UMP, the unemployment rate remains well above its steady state level, whereas profits recover relatively quickly and stay above steady state even without UMP. For poor households, UMP improves outcomes precisely when unemployment is high. For wealthy households, profits are already high in the no UMP economy, so additional profit gains from UMP have a smaller marginal value than in the steady state. As a result, the positive impact of UMP on labor market conditions is more important for poor households than in the steady state analysis, while the positive welfare effects from higher profits are less important for the top in the counterfactual simulations.

The contribution of lump sum transfers to welfare also differs between the two cases. In the model, a rise in government surplus induced by UMP is passed to households only through the MMMF component of lump sum transfers. During the ELB episode, positive lump sum transfer shocks raise the overall level of trans-

[39]By working status, business owners see the largest gains, about 0.82 percent of lifetime consumption. The unemployed gain about 0.35 percent, and the employed gain about 0.27 percent at the posterior mode.

Figure 10: Effects of UMP on inequality measures



Notes: The left panel shows the difference between the baseline with UMP and the counterfactual without UMP for the income Gini, reported as index-point deviations from the no UMP case on the 0 to 100 scale. The right panel shows the corresponding difference for the top 10 percent income share, reported as percentage-point deviations from the no UMP case. The black solid line denotes the value at the posterior mode, and the gray shaded band spans the 10th to 90th percentile range across posterior draws. Lines labeled P10 and P90 trace the 10th and 90th percentiles. Positive values indicate higher inequality under UMP than under no UMP for the measure shown.

fers, so the share coming from MMMF is smaller, roughly half of its steady state share. MMMF also faces positive liquidity preference shocks at the beginning of the ELB episode, whose effects persist, so part of the additional government surplus is directed toward liquid assets rather than transfers to households. As a result, the welfare effect attributed to transfers is smaller in the counterfactual simulations than in the steady state IRF analysis. The welfare computation based on no shock paths with backward and forward iterations captures these differences in the relative importance of aggregate variables such as the unemployment rate, profits, and transfers.

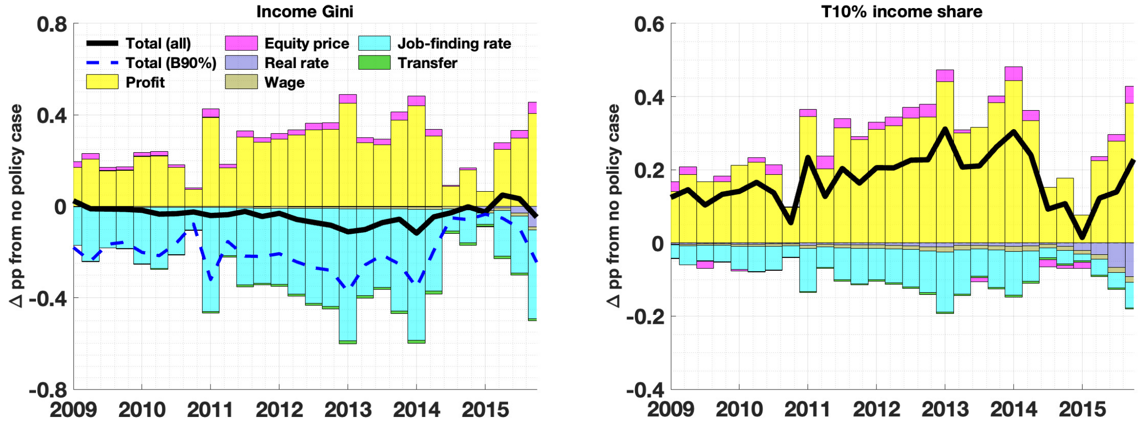
6 Effects of UMP on inequality measures

A large empirical literature evaluates distributional consequences of monetary policy using summary inequality statistics such as the Gini index and top income shares.[40] To relate the model's implications to that body of work and to shed light on why results sometimes appear to conflict, I examine the same objects in the simulations during the ELB episode.

Figure 10 reports the model implied effects of UMP on two standard income inequality measures by comparing the baseline to the counterfactual without UMP.

[40]For example, Mumtaz and Theophilopoulou (2017) estimate UK impulse responses for the Gini of disposable income, consumption, and net wealth. Saiki and Frost (2014) study Japan's QQE and document increases in the income Gini. Ampudia et al. (2018) use euro area HFCS micro data to track the Gini of income, consumption, and net wealth around monetary shocks. Bivens (2015) assesses U.S. unconventional policy with a focus on top income shares and related income components.

Figure 11: Effects of UMP on inequality measures: decomposition



Notes: The left panel shows the differences in the model implied income Gini index (0 to 100) between the baseline and the counterfactual case, while the right panel shows the income share of the top 10 percent as the percentage point difference compared to the corresponding levels in the counterfactual case. The dotted blue line represents the difference in the income Gini index among the bottom 90 percent wealthiest households at the posterior mode. The colored bars represent the contributions of different variables to the effects on each inequality measure.

For the Gini index, I plot $\text{Gini}^{\text{UMP}} - \text{Gini}^{\text{NoUMP}}$ in index points, and for the top 10 percent share I plot $\text{Top10}^{\text{UMP}} - \text{Top10}^{\text{NoUMP}}$ in percentage points. Negative values for the Gini difference mean lower overall inequality under UMP, while positive values for the top 10 percent share difference mean a larger top share under UMP.

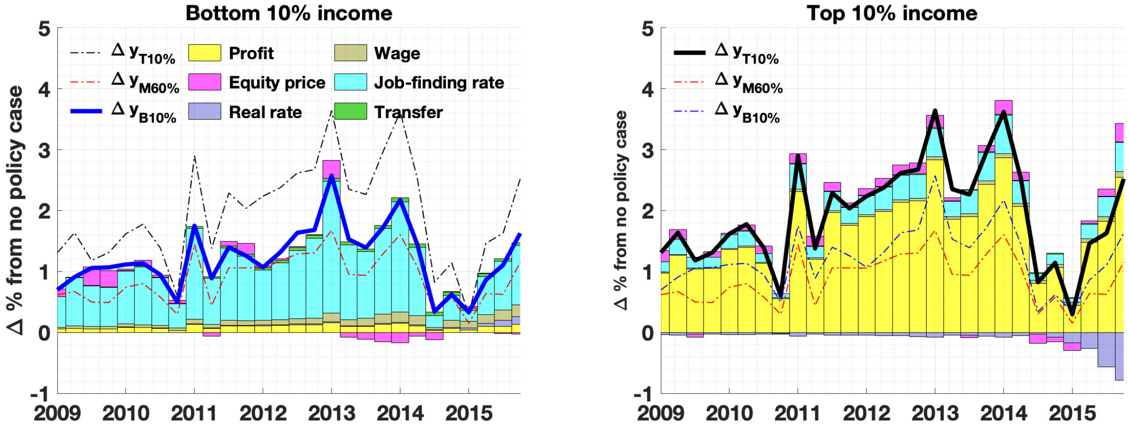
The two measures deliver different perspectives. The left panel of Figure 10 shows a modest decline in the overall income Gini under UMP across posterior draws, with trough effects around 0.04 index point at the posterior mode and an average improvement that is small but negative during the ELB episode.[41] The mechanism is straightforward. UMP raises aggregate demand and improves the job finding probability. Because labor income is the primary source of income for most households, especially away from the very top of the wealth distribution, the reduction in unemployment compresses the income distribution from the middle down. Consistent with that logic, the dotted blue line in the left panel of Figure 12 shows that the Gini computed within the bottom 90 percent declines more, closely tracking the contribution of improvement in employment. These dynamics are in line with central bankers' emphasis on the labor market channel during the recovery.[42]

The right panel in Figure 10 shows that the top 10 percent income share increases under UMP across posterior draws, with an average increase of roughly 0.2 per-

[41] Lenza and Slacalek (2018) study the Eurosystem's asset purchase programme using euro area household micro data. In their baseline calibration, a QE shock that lowers the term spread by 30 basis points reduces the gross income Gini from about 43.15 to 43.07 four quarters after the shock, while the wealth Gini changes very little. The changes in the income Gini implied by the ELB episode experiment in this paper are even smaller, consistent with the view that QE has modest effects on standard inequality measures relative to its aggregate impact.

[42] See, for example, Bernanke (2015) and Draghi (2016).

Figure 12: Distributional effects of UMP on group incomes



Notes: Each panel plots income differences between the baseline and the counterfactual for the top 10 percent (black), the middle 60 percent (red), and the bottom 10 percent (blue) wealth groups. Units are percent differences relative to the counterfactual group income at each date. Colored bars show the contribution of proximate income components to the group-level effect.

centage point at the posterior mode. As Figure 12 shows, the increase is driven by higher profits and equity payouts when demand and investment rise, which raise the income of households at the top because they hold a disproportionate share of equity and business ownership. Lower unemployment works in the opposite direction but does not fully offset the profit and equity channels at the very top. This divergence between the Gini and the top share helps explain why empirical studies can reach different conclusions depending on which summary statistic they emphasize and which parts of the distribution the underlying data measure well. The Gini is more sensitive to broad improvements in employment across the lower and middle parts of the distribution, whereas top income shares are dominated by movements in capital income.

Figure 12 complements these results by showing the evolution of group income directly. Lines report $100 \times (y^{\text{UMP}}_{\text{group}} / y^{\text{NoUMP}}_{\text{group}} - 1)$ so units are percent differences relative to the no-UMP path for each group at each date. The top 10 percent experiences the largest gains in income under UMP. On average the top group's income is about 2 percent higher than in the counterfactual. The decomposition highlights profits and equity prices as the dominant contributors for this group. The bottom 10 percent also gains, with income more than 1 percent higher on average. The channel is the improvement in employment prospects, which matters more at the bottom because unemployment is more prevalent there at the onset of the episode.[43] The middle 60 percent records the smallest gains because real wages move little in the short run and this group relies heavily on labor income from already employed households.

[43] At 2009Q1 in the model, 8.75 percent of households in the bottom 10 percent of the wealth distribution are unemployed compared with 6.54 percent in the middle quintile. An increase in job finding therefore moves more households from unemployment to employment at the bottom.

7 Decomposition and CMP comparison

Central banks face two practical questions when the policy rate is constrained by the effective lower bound. One concerns instrument choice, namely whether large scale asset purchases or forward guidance should carry more of the burden.[44] The other concerns the cost of the binding ELB constraint in aggregate and distributional terms, which bears on whether policymakers should seek more conventional policy space through framework choices such as a higher steady inflation objective or level type strategies.[45] To provide insights on these questions, this section first compares the two instruments within unconventional monetary policy, quantitative easing (QE) and forward guidance (FG), for the ELB episode studied here. The results above reflect their combined effects, so I decompose the total UMP effect into the distinct contributions of QE and FG to identify which instrument drives outcomes and when during the episode each is more important. I then compare unconventional and conventional policy by asking what would have occurred if the ELB constraint were not binding and the central bank did not need to rely on unconventional tools. This comparison sheds light on the cost of the binding ELB in the aftermath of the Great Recession. Unless noted, simulations use posterior mode parameters and the baseline sequence of exogenous expected ELB durations.

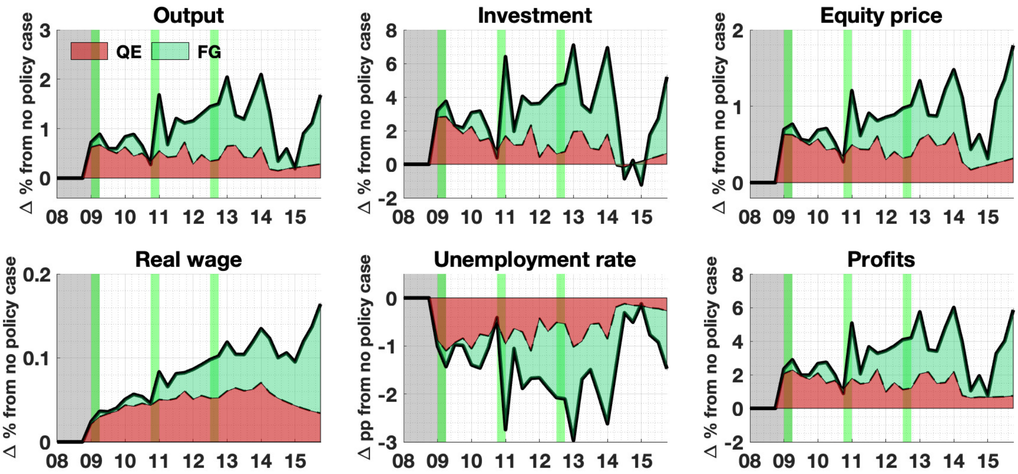
7.1 Decomposition of UMP: QE versus FG

This section isolates the contribution of QE and FG during the ELB episode by simulating an additional counterfactual economy in which ELB durations are determined endogenously by the state of the economy and realized shocks. In this counterfactual, the policy rate lifts off as soon as the estimated rule prescribes a positive rate, while QE is conducted as in the baseline economy. Figure 13 shows the decomposition of total UMP effects during the ELB episode. On average, QE accounts for about 55 percent of the total effect on key aggregates, with FG providing the remaining 45 percent. Although the average split is close to half and half, the timing differs and their roles are not interchangeable. QE's impact is concentrated early in the episode when the first expansion of the central bank balance sheet arrives as a large, unexpected intervention in the model. This early move generates sizable stimulus because it raises the central bank's balance sheet sharply relative to expectations. Even though the balance sheet later reaches nearly four times its precrisis size by

[44] Theoretical work shows that history dependent commitments can be powerful at the lower bound, see Eggertsson and Woodford (2003). Event study and term structure evidence indicates that UMP announcements moved longer term yields, see Gagnon et al. (2011) and Bhattarai and Neely (2016).

[45] See Blanchard et al. (2010), Williams (2016), and Bernanke (2017).

Figure 13: Aggregate effects of QE and forward guidance



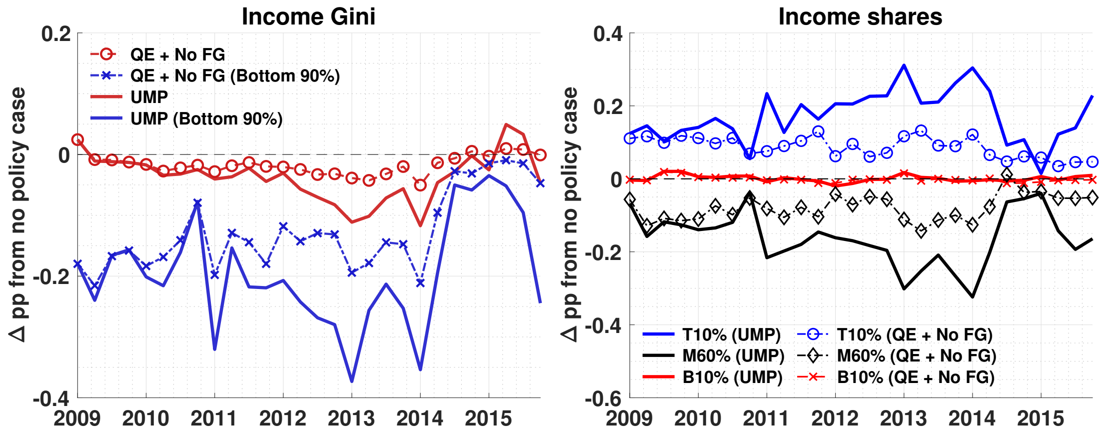
Notes: Black line, baseline UMP. Red shading, contribution of QE holding ELB durations endogenous. Green shading, additional effect from FG.

2014, the marginal aggregate effects of additional purchases are smaller because subsequent expansions are more gradual and partly anticipated.[46] FG plays a smaller role in the early period because the estimated exogenous ELB durations are shorter while the endogenous durations implied by the severe contraction are relatively long, so the gap between them is small. Beginning around 2011, exogenous durations lengthen through active guidance and endogenous durations shorten as conditions stabilize, so the gap widens. With liftoff warranted by the Taylor rule in the absence of guidance, keeping the policy rate at zero becomes more effective in sustaining stimulus. In this sense, QE provides the initial impulse, and FG prolongs and amplifies it.

Forward guidance also strengthens the distributional impact of QE. With QE alone, the income Gini index for the full distribution barely moves over the episode because gains at the bottom and at the top roughly offset each other. Early in the ELB episode, the Gini index measured for the bottom 90 percent of the wealth distribution falls by about 0.2 percentage points as unemployment declines. However, the overall Gini index changes very little because the top 10 percent benefits from higher profits and higher equity prices, and the top 10 percent income share rises by roughly 0.1 percentage point, as shown in the right panel of Figure 14. Once FG is added, the distributional consequences become more pronounced. Around 2013, the overall Gini index is about 0.05 percentage points lower under UMP than under QE alone,

[46]Existing work points in the same direction. The first round of large scale asset purchases lowered longer term yields by roughly 1 to 2 percentage points, whereas later rounds such as QE2 and the Maturity Extension Program moved yields by only about 0.2 to 0.4 percentage points, see Bhattarai and Neely (2016). Event study and term structure analyses also find that later purchases had smaller incremental effects as markets internalized the possibility of renewed interventions. See, for example, Gagnon et al. (2011), D'Amico and King (2013), and Li and Wei (2013).

Figure 14: Distributional effects of QE and FG: inequality measures



Notes: Left, differences in the income Gini index relative to the no-policy counterfactual. Right, differences in group income shares relative to the counterfactual. Solid lines, baseline UMP. Dotted lines, QE with endogenous ELB durations.

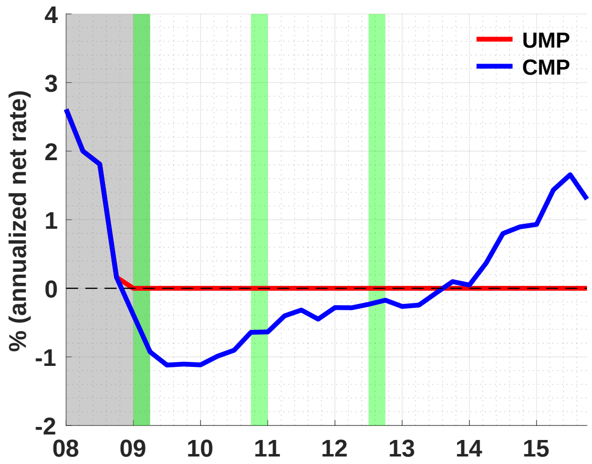
and the bottom 90 percent Gini is about 0.2 percentage points lower. The guidance induced extension of the ELB also boosts profits by about 2 to 4 percent and equity prices by about 1 percent relative to the QE only path between 2011 and 2014, which pushes the top 10 percent income share higher and compresses the middle sixty share further. In short, FG scales up the effect of QE. It improves labor market outcomes in a way that lowers inequality among the bottom 90 percent, and, at the same time, it raises asset income at the top. The result is a more clearly U-shaped profile of gains across the wealth distribution.

Taken together, the model points to a division of labor between the two tools. QE is especially powerful at the start of the crisis, when the central bank can deliver a large and unexpected balance sheet expansion. FG becomes more powerful later, when the policy rate is already expected to lift off and additional purchases are less surprising, because extending the expected ELB duration prolongs and amplifies the initial stimulus. On distributional effects, neither instrument delivers uniformly progressive or regressive effects. Both QE and FG support lower income households through the labor market, and both raise the top 10 percent income share through higher profits and equity values. FG mainly increases the scale of both channels.

7.2 UMP versus CMP

I next assess the cost of the effective lower bound when unconventional tools are available by contrasting the baseline UMP with a CMP counterfactual in which the policy rate follows the estimated Taylor rule without the ELB constraint and the

Figure 15: Policy rate paths: baseline UMP vs unconstrained CMP



Notes: The red line shows the policy rate under UMP, while the blue line shows the policy rate under the CMP counterfactual with no ELB and no UMP.

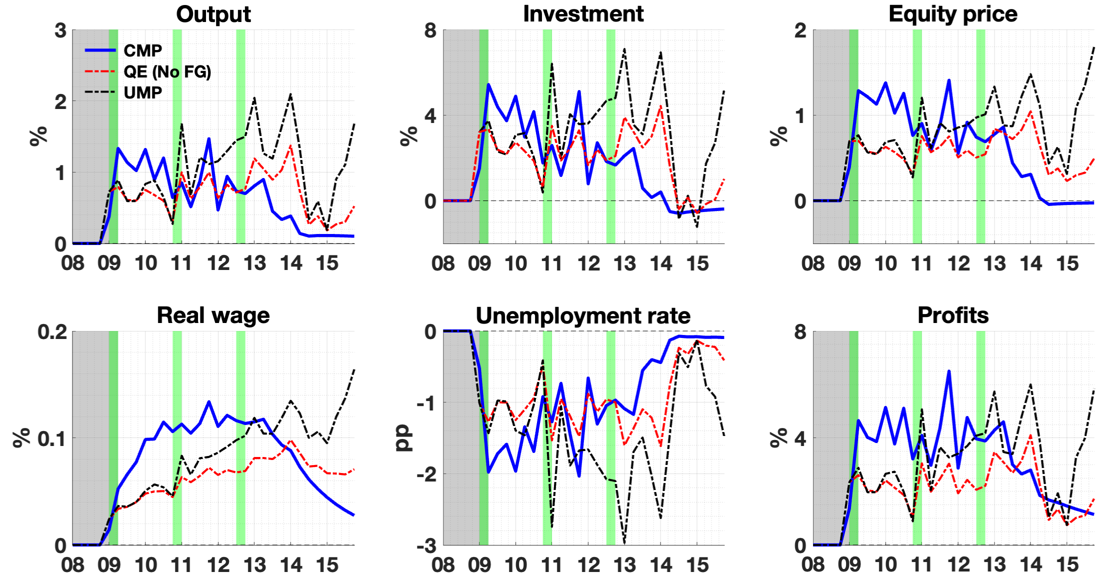
central bank does not use unconventional tools.[47] As shown in Figure 15, under CMP the policy rate falls to about minus one percent in 2009, remains below zero until about 2014, then rises. Under UMP, by construction, the policy rate stays at the effective lower bound over the episode.

Early in the episode CMP delivers a somewhat stronger aggregate boost than UMP. In 2009–2010, output under CMP is roughly 0.3 to 0.4 percent above the UMP path, and profits and equity prices are higher as well. The mechanism is straightforward. A lower policy rate reduces banks' funding costs and crowds in private investment. Under UMP, the initial QE surprise raises equity valuations but compresses expected returns, which crowds out bank investment in general equilibrium.[48] As shown in Figure 17, CMP raises banks' equity holdings noticeably early in the ELB episode by lowering funding costs. Under UMP, banks' equity holdings do not increase much at first. Instead, the central bank's equity position rises in the model, supporting aggregate investment through QE. Consequently, CMP also delivers a larger and more persistent increase in banks' net worth. QE lifts banks' net worth through direct valuation effects, but the general equilibrium impact of cheaper funding and the associated expansion of banks' balance sheets under CMP

[47] In the simulation the nominal policy rate may fall below zero. I do not interpret this as the effect of negative policy rates in practice, where deposit rate passthrough is imperfect. Instead I treat it as CMP in an economy with a higher steady state nominal rate and inflation rate such that real rates match those in the baseline while the central bank has more room to lower the nominal rate.

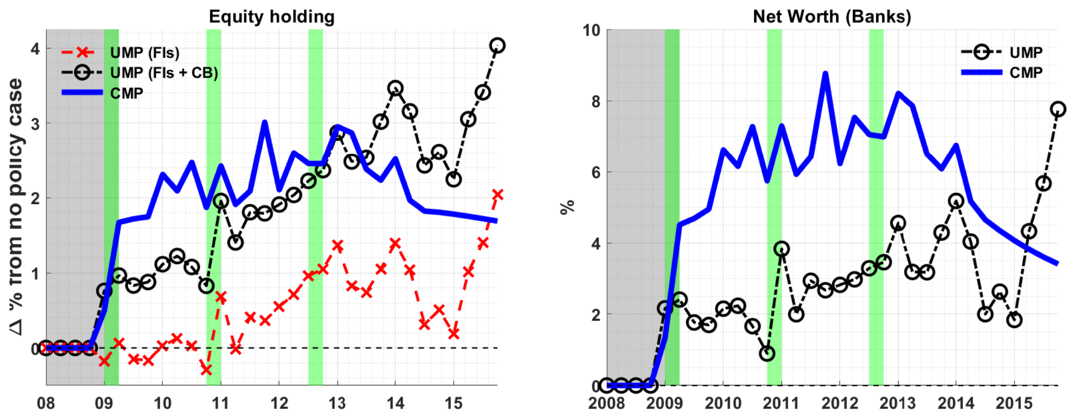
[48] Several strands of evidence support the channels emphasized here. When policy space is ample, as in our CMP counterfactual, rate cuts can steepen the yield curve and raise net interest margins by lowering funding costs, see English et al. (2018) and Borio et al. (2017). In contrast, large scale asset purchases reduce long yields mainly via term premium and portfolio balance channels and tend to flatten the curve, see Gagnon et al. (2011), Krishnamurthy and Vissing-Jorgensen (2011), and D'Amico and King (2013). This compression can weigh on intermediation margins for deposit funded banks.

Figure 16: Aggregate effects of UMP and CMP



Notes: Red and blue lines are UMP and CMP. All variables, except unemployment, are percent differences from the no-policy counterfactual. Unemployment is in percentage point differences.

Figure 17: Banking sector responses: UMP vs CMP

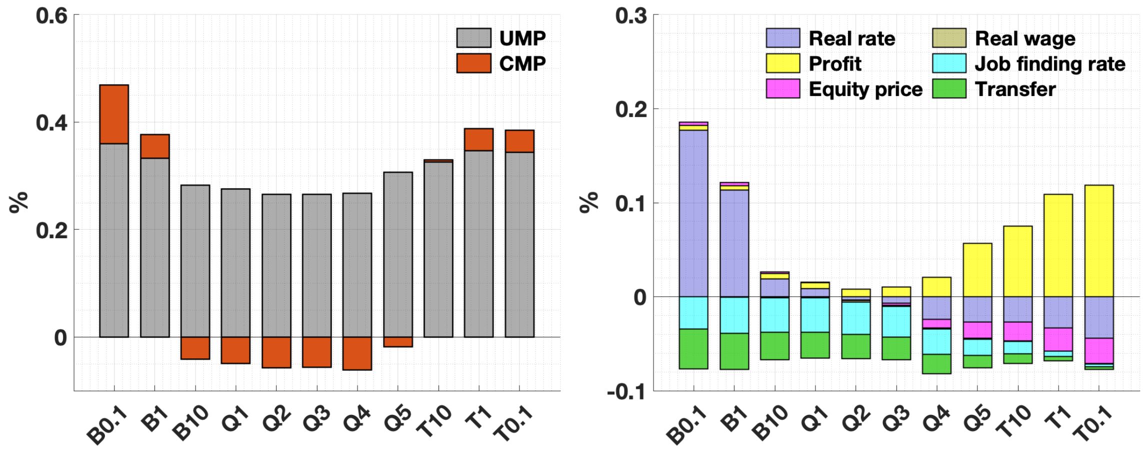


Notes: Left, equity holdings. Right, banks' net worth. Blue, CMP. Red lines isolate banks under UMP. Black dotted lines include both financial institutions and the central bank where relevant.

is larger. Later in the episode, when forward guidance plays a more active role, UMP provides greater support because CMP stimulus wanes after liftoff, while guidance keeps policy accommodative. Overall, there is a modest early aggregate cost from the lower bound, reflecting the more favorable bank funding channel under CMP, but the sustained support from forward guidance later in the episode closes most of this gap, so average UMP and CMP effects are comparable over the full episode.

Turning from aggregate outcomes to distributional consequences, UMP and CMP display slight differences. Contrary to the common view that UMP, especially QE, inevitably worsens inequality, the simulations indicate that CMP produces the more

Figure 18: Welfare gain comparison: UMP vs CMP



Notes: Gray bars, welfare gains from UMP in consumption equivalents. Dark orange bars, additional gains from CMP relative to UMP. The right panel decomposes the additional CMP gains relative to UMP.

pronounced U-shaped pattern of welfare gains.[49] As shown in the left panel of Figure 18, households at the very bottom and the very top of the wealth distribution gain more under CMP than under UMP, while the middle gains less relative to UMP. The decomposition in the right panel of Figure 18 explains why. Lower real funding costs under CMP directly benefit borrowers at the bottom through cheaper debt service and easier credit access. For households at the top, lower real rates reduce returns on safe liquid assets, but that effect is more than offset by stronger bank profitability and equity income when the policy rate is unconstrained. These financial sector channels, shown by the yellow bars, tilt CMP's gains toward the top. By contrast, UMP delivers broad based improvements with a flatter profile across the distribution.

In summary, these aggregate results suggest that, once UMP is available, the ELB is not very costly in aggregate terms. Over the full episode, UMP and CMP provide similar overall stimulus, with CMP somewhat stronger early on and UMP stronger later through forward guidance. The distributional implications slightly differ. CMP concentrates incremental gains in the tails and compresses the middle, whereas UMP spreads gains more evenly across households.

[49] Franconi and Rella (2023) provide complementary evidence on wealth inequality using the U.S. Distributional Financial Accounts. They show that expansionary interest rate cuts and asset purchases both raise wealth across the distribution, but have distinct implications for wealth inequality. Over the medium term, conventional rate cuts lead to a persistent increase in top wealth shares, whereas the inequality effect of asset purchases is short lived and largely fades out. Their finding that conventional policy has more persistent and uneven effects on the wealth distribution than QE points in the same direction as the comparison between UMP and conventional policy in this section.

8 Conclusion

This paper quantifies the aggregate and distributional consequences of the Federal Reserve's unconventional monetary policy during 2009 to 2015 in a medium-scale heterogeneous agent New Keynesian model with portfolio heterogeneity, frictional labor markets, and a banking sector. UMP is modeled as quantitative easing through central bank asset purchases and forward guidance through exogenous extensions of the effective lower bound, and the model is estimated with Bayesian methods on U.S. aggregates while explicitly accounting for the lower bound. The calibration matches key cross sectional features of the 2007 Survey of Consumer Finances, so that differences in wealth, income composition, and unemployment risk shape the transmission of policy across households.

In the estimated model, UMP provided meaningful macroeconomic stabilization during the ELB episode. Relative to a counterfactual without asset purchases and forward guidance, unconventional policy raised output by about 1 percent on average, lowered the unemployment rate by around one and a half percentage points, and increased profits and equity prices while leaving real wages nearly unchanged. These aggregate gains translated into broad based welfare improvements. Measured in consumption equivalents at the start of the ELB episode, the average welfare gain is about 0.27 percent of lifetime consumption and is mildly U-shaped across the wealth distribution. Households at the bottom benefit mainly through lower unemployment and improved job finding, while households at the top benefit mainly through higher profits and equity valuations, with somewhat smaller gains for the middle that relies heavily on labor income.

Distributional effects appear mixed when summarized by standard inequality measures, which helps reconcile divergent findings in empirical work. During the ELB episode the income Gini index falls slightly on average because UMP compresses the lower and middle parts of the distribution through lower unemployment, while the top 10 percent income share rises because higher profits and equity prices raise capital income at the top. The analysis also shows that QE and forward guidance play complementary roles, with QE providing about half of the total stimulus through early balance sheet expansions and forward guidance sustaining and amplifying these effects later in the episode. A comparison with a counterfactual unconstrained Taylor rule suggests that, once UMP is available, the lower bound is not very costly in aggregate terms, while conventional policy is somewhat more adverse distributionally because it channels relatively larger gains to the financial sector.

References

Altavilla, Carlo, Miguel Boucinha, and José-Luis Peydró (2018) "Monetary policy and bank profitability in a low interest rate environment", *Economic Policy*, 33 (96), pp. 531–586.

Alves, F., G. Kaplan, B. Moll, and Giovanni L. Violante (2019) "A further look at the propagation of monetary policy shocks in hank".

Ampudia, Miguel, Dimitris Georgarakos, Jiri Slacalek, Oreste Tristani, Philip Vermeulen, and Giovanni L. Violante (2018) "Monetary policy and household inequality", ECB Working Paper 2170, European Central Bank.

Anderson, Eric, Sergio Rebelo, and Arlene Wong (2018) "Markups across space and time", Working Paper 24434, National Bureau of Economic Research.

Atkinson, Tyler, Alexander W. Richter, and Nathaniel A. Throckmorton (2020) "The zero lower bound and estimation accuracy", *Journal of Monetary Economics*, 115, pp. 249–264.

Auclert, Adrien, Bence Bardóczy, and Matthew Rognlie (2020a) "Mpcs, mpes and multipliers: A trilemma for new keynesian models", Working Paper 27486, National Bureau of Economic Research.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub (2020b) "Micro jumps, macro humps: Monetary policy and business cycles in an estimated hank model", Working Paper 26647, National Bureau of Economic Research.

Bayer, Christian, Benjamin Born, and Ralph Luetticke (2020) "Shocks, frictions, and inequality in us business cycles", CEPR Discussion Papers 14364.

Bayer, Christian and Ralph Luetticke (2020) "Solving heterogeneous agent models in discrete time with many idiosyncratic states by perturbation methods", *Quantitative Economics*, forthcoming.

Bayer, Christian, Ralph Luetticke, Lien Pham-Dao, and Volker Tjaden (2019) "Precautionary savings, illiquid assets, and the aggregate consequences of shocks to household income risk", *Econometrica*, 87 (1), pp. 255–290.

Bernanke, Ben S. (2015) "Monetary policy and inequality", *Brookings Blog*.

Bernanke, Ben S. (2017) "Monetary policy in a new era", Paper prepared for the conference on Rethinking Macroeconomic Policy.

Bhattarai, Saroj and Christopher J. Neely (2016) "A survey of the empirical literature on u.s. unconventional monetary policy", Working Paper 2016-021, Federal Reserve Bank of St. Louis.

Bhattarai, Saroj and Christopher J. Neely (2022) "An analysis of the literature on international unconventional monetary policy", *Journal of Economic Literature*, 60 (2), pp. 527–597.

Bilbiie, Florin and Diego R. Känzig (2023) "Greed? profits, inflation, and aggregate demand", NBER Working Paper 31618, National Bureau of Economic Research, NBER Working Paper No. 31618. Available at SSRN: https://ssrn.com/abstract=4553621.

Bilbiie, Florin O. (2009) "Nonseparable preferences, fiscal policy puzzles, and inferior goods", *Journal of Money, Credit and Banking*, 41 (2/3), pp. 443–450.

Bivens, Josh (2015) "Gauging The Impact Of The Fed On Inequality During The Great Recession", *Hutchins Center on Fiscal and Monetary Policy at Brookings* (12), pp. 1 – 33.

Blanchard, Olivier, Giovanni Dell'Ariccia, and Paolo Mauro (2010) "Rethinking macroeconomic policy", IMF Staff Position Note SPN/10/03, International Monetary Fund.

Borio, Claudio, Leonardo Gambacorta, and Boris Hofmann (2017) "The influence of monetary policy on bank profitability", *International Finance*, 20 (1), pp. 48–63.

Broer, Tobias, Niels-Jakob Harbo Hansen, Per Krusell, and Erik Öberg (2019) "The New Keynesian Transmission Mechanism: A Heterogeneous-Agent Perspective", *The Review of Economic Studies*, 87 (1), pp. 77–101.

Carroll, Christopher D. (2006) "The method of endogenous gridpoints for solving dynamic stochastic optimization problems", *Economics Letters*, 91 (3), pp. 312 – 320.

Casiraghi, Marco, Eugenio Gaiotti, Lisa Rodano, and Alessandro Secchi (2018) "A "reverse robin hood"? the distributional implications of non-standard monetary policy for italian households", *Journal of International Money and Finance*, 85, pp. 215 – 235.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber (2011) "Are micro and macro labor supply elasticities consistent? a review of evidence on the intensive and extensive margins", *American Economic Review*, 101 (3), pp. 471–75.

Chib, Siddhartha and Srikanth Ramamurthy (2010) "Tailored randomized block MCMC methods with application to DSGE models", *Journal of Econometrics*, 155 (1), pp. 19–38.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans (1999) "Chapter 2 monetary policy shocks: What have we learned and to what end?", 1 of Handbook of Macroeconomics, Elsevier, pp. 65 – 148.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans (2005) "Nominal rigidities and the dynamic effects of a shock to monetary policy", *Journal of Political Economy*, 113 (1), pp. 1–45.

Christiano, Lawrence J., Martin S. Eichenbaum, and Mathias Trabandt (2016) "Unemployment and business cycles", *Econometrica*, 84 (4), pp. 1523–1569.

Chung, Hess, Jean-Philippe Laforte, David Reifschneider, and John C. Williams (2011) "Have we underestimated the likelihood and severity of zero lower bound events? (box: The estimated effects of the fed's asset purchase program)", *FRBSF Economic Letter* (2011-01).

Cohan, William D. (2014) "How quantitative easing contributed to the nation's inequality problem", *The New York Times Dealbook*.

Coibion, Olivier, Yuriy Gorodnichenko, Lorenz Kueng, and John Silvia (2017) "Innocent bystanders? monetary policy and inequality", *Journal of Monetary Economics*, 88, pp. 70 – 89.

Cuba-Borda, Pablo, Luca Guerrieri, Matteo Iacoviello, and Molin Zhong (2019) "Likelihood evaluation of models with occasionally binding constraints", *Journal of Applied Econometrics*, 34 (7), pp. 1073–1085.

Cui, Wei and Vincent Sterk (2021) "Quantitative easing with heterogeneous agents", *Journal of Monetary Economics*, 123, pp. 68–90.

D'Amico, Stefania and Thomas B. King (2013) "Flow and stock effects of large-scale treasury purchases: Evidence on the importance of local supply", *Journal of Financial Economics*, 108 (2), pp. 425–448.

Davtyan, Karen (2017) "The distributive effect of monetary policy: The top one percent makes the difference", *Economic Modelling*, 65, pp. 106 – 118.

Domanski, Dietrich, Michela Scatigna, and Anna Zabai (2016) "Wealth inequality and monetary policy", *BIS Quarterly Review*.

Draghi, Marion (2016) "Stability, Equity, and Monetary Policy", 2nd DIW Europe Lecture, German Institute for Economic Research (DIW).

Eggertsson, Gauti B. and Michael Woodford (2003) "The zero bound on interest rates and optimal monetary policy", *Brookings Papers on Economic Activity* (1), pp. 139–233.

Engen, Eric M., Thomas Laubach, and David Reifschneider (2015) "The macroeconomic effects of the federal reserve's unconventional monetary policies", Finance and Economics Discussion Series 2015-5, Board of Governors of the Federal Reserve System.

English, William B., Skander J. van den Heuvel, and Egon Zakrajšek (2018) "Interest rate risk and bank equity valuations", *Journal of Monetary Economics*, 98, pp. 80–97.

Franconi, Alessandro and Giacomo Rella (2023) "Monetary policy across the wealth distribution", Working Paper 70, Stone Center on Socio-Economic Inequality, Stone Center on Socio-Economic Inequality Working Paper Series.

Furceri, Davide, Prakash Loungani, and Aleksandra Zdzienicka (2018) "The effects of monetary policy shocks on inequality", *Journal of International Money and Finance*, 85, pp. 168 – 186.

Gagnon, Joseph, Matthew Raskin, Julie Remache, and Brian Sack (2011) "The financial market effects of the federal reserve's large-scale asset purchases", *International Journal of Central Banking*, 7 (1), pp. 3–43.

Gertler, Mark and Peter Karadi (2011) "A model of unconventional monetary policy", *Journal of Monetary Economics*, 58 (1), pp. 17 – 34, Carnegie-Rochester Conference Series on Public Policy: The Future of Central Banking April 16-17, 2010.

Gornemann, Nils, Keith Kuester, and Makoto Nakajima (2016) "Doves for the Rich, Hawks for the Poor? Distributional Consequences of Monetary Policy", International Finance Discussion Papers 1167, Board of Governors of the Federal Reserve System (U.S.).

Gould, Elise (2019) "Decades of rising economic inequality in the u.s. testimony before the u.s. house of representatives ways and means committee", *Economic Policy Institute*.

Greenwood, Jeremy, Zvi Hercowitz, and Gregory W. Huffman (1988) "Investment, capacity utilization, and the real business cycle", *The American Economic Review*, 78 (3), pp. 402–417.

Guerrieri, Luca and Matteo Iacoviello (2015) "Occbin: A toolkit for solving dynamic models with occasionally binding constraints easily", *Journal of Monetary Economics*, 70, pp. 22 – 38.

Guerrieri, Luca and Matteo Iacoviello (2017) "Collateral constraints and macroeconomic asymmetries", *Journal of Monetary Economics*, 90, pp. 28–49.

den Haan, Wouter J., Garey Ramey, and Joel Watson (2000) "Job destruction and propagation of shocks", *American Economic Review*, 90 (3), pp. 482–498.

Hafemann, Lucas, Paul Rudel, and Jörg Schmidt (2018) "Moving closer or drifting apart: Distributional effects of monetary policy", *The Manchester School*, 86 (S1), pp. 110–136.

Heathcote, Jonathan, Fabrizio Perri, and Giovanni L. Violante (2010) "Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006", *Review of Economic Dynamics*, 13 (1), pp. 15–51.

Inui, Masayuki, Nao Sudo, and Tomoaki Yamada (2017) "Effects of Monetary Policy Shocks on Inequality in Japan", Bank of Japan Working Paper Series 17-E-3, Bank of Japan.

Jones, Callum (2017) "Unanticipated shocks and forward guidance at the zlb".

Jones, Callum, Mariano Kulish, and Daniel M. Rees (2022) "International spillovers of forward guidance shocks", *Journal of Applied Econometrics*, 37 (1), pp. 131–160.

Justiniano, Alejandro, Giorgio E. Primiceri, and Andrea Tambalotti (2011) "Investment shocks and the relative price of investment", *Review of Economic Dynamics*, 14 (1), pp. 102 – 121, Special issue: Sources of Business Cycles.

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante (2018) "Monetary policy according to hank", *American Economic Review*, 108 (3), pp. 697–743.

Kim, Taeyoung, Thomas Laubach, and Min Wei (2023) "Macroeconomic effects of large-scale asset purchases and forward guidance", Finance and Economics Discussion Series 2020-047r1, Board of Governors of the Federal Reserve System.

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

Kulish, Mariano, James Morley, and Tim Robinson (2014) "Estimating the Expected Duration of the Zero Lower Bound in DSGE Models with Forward Guidance", Melbourne Institute Working Paper Series wp2014n16, Melbourne Institute of Applied Economic and Social Research, The University of Melbourne.

Lee, Donggyu (2019) "The Effects of Monetary Policy on Consumption and Inequality", *Working paper*.

Lenza, Michele and Jiri Slacalek (2018) "How does monetary policy affect income and wealth inequality? Evidence from quantitative easing in the euro area", Working Paper Series 2190, European Central Bank.

Li, Canlin and Min Wei (2013) "Term structure modeling with supply factors and the federal reserve's large-scale asset purchase programs", Finance and Economics Discussion Series 2013-39, Board of Governors of the Federal Reserve System.

McKay, Alisdair and Christian K. Wolf (2023) "Monetary policy and inequality", *Journal of Economic Perspectives*, 37 (1), p. 121–44.

Monacelli, Tommaso and Roberto Perotti (2008) "Fiscal Policy, Wealth Effects, and Markups", NBER Working Papers 14584, National Bureau of Economic Research, Inc.

Montecino, Juan A. and Gerald Epstein (2015) "Did Quantitative Easing Increase Income Inequality?", Working Papers Series 28, Institute for New Economic Thinking.

Mortensen, Dale T. and Christopher A. Pissarides (1994) "Job creation and job destruction in the theory of unemployment", *Review of Economic Studies*, 61 (3), pp. 397–415.

Mumtaz, Haroon and Angeliki Theophilopoulou (2017) "The impact of monetary policy on inequality in the uk: An empirical analysis", *European Economic Review*, 98, pp. 410–423.

Nekarda, Christopher J. and Valerie A. Ramey (2020) "The cyclical behavior of the price-cost markup", *Journal of Money, Credit and Banking: 50th Anniversary Issue*, Forthcoming.

Pissarides, Christopher A. (2000) *Equilibrium Unemployment Theory*, Cambridge MA, MIT Press, second edition edition.

Reiter, Michael (2009) "Solving heterogeneous-agent models by projection and perturbation", *Journal of Economic Dynamics and Control*, 33 (3), pp. 649 – 665.

Robert E. Lucas, Jr. (1987) *Models of Business Cycles*, Oxford, Basil Blackwell.

Rotemberg, Julio J. (1982) "Sticky prices in the united states", *Journal of Political Economy*, 90 (6), pp. 1187–1211.

Saez, Emmanuel and Gabriel Zucman (2016) "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data", *The Quarterly Journal of Economics*, 131 (2), pp. 519–578.

Saiki, Ayako and Jon Frost (2014) "Does unconventional monetary policy affect inequality? evidence from japan", *Applied Economics*, 46 (36), pp. 4445–4454.

Schwartz, Nelson D. (2013) "Recovery in u.s. is lifting profits, but not adding jobs", *The New York Times*.

Sims, Eric R., Jing Cynthia Wu, and Ji Zhang (2022) "Unconventional monetary policy according to hank", NBER Working Paper 30329, National Bureau of Economic Research.

Smets, Frank and Rafael Wouters (2007) "Shocks and frictions in us business cycles: A bayesian dsge approach", *American Economic Review*, 97 (3), pp. 586–606.

Storesletten, Kjetil, Chris I. Telmer, and Amir Yaron (2001) "The welfare cost of business cycles revisited: Finite lives and cyclical variation in idiosyncratic risk", *European Economic Review*, 45 (7), pp. 1311–1339.

Storesletten, Kjetil, Chris I. Telmer, and Amir Yaron (2004) "Cyclical dynamics in idiosyncratic labor market risk", *Journal of Political Economy*, 112 (3), pp. 695–717.

Swanson, Eric T. and John C. Williams (2014) "Measuring the effect of the zero lower bound on medium- and longer-term interest rates", *American Economic Review*, 104 (10), pp. 3154–3185.

Tauchen, George (1986) "Finite state markov-chain approximations to univariate and vector autoregressions", *Economics Letters*, 20 (2), pp. 177 – 181.

Williams, John C. (2016) "Monetary policy in a low r world", *FRBSF Economic Letter* (2016-23).

Woodford, Michael (1995) "Price-level determinacy without control of a monetary aggregate", *Carnegie-Rochester Conference Series on Public Policy*, 43, pp. 1–46.

Yellen, Janet L. (2014) "Perspectives on Inequality and Opportunity from the Survey of Consumer Finances", Speech at the Conference on Economic Opportunity and Inequality, Federal Reserve Bank of Boston.

Young, Eric R. (2010) "Solving the incomplete markets model with aggregate uncertainty using the krusell–smith algorithm and non-stochastic simulations", *Journal of Economic Dynamics and Control*, 34 (1), pp. 36 – 41, Computational Suite of Models with Heterogeneous Agents: Incomplete Markets and Aggregate Uncertainty.

Zimmerman, Kasper (2019) "The influence of monetary policy on bank profitability".

Online Appendix

A Further details on the model description

A.1 Households

Time is discrete, $t = 0, 1, \ldots$. Households face the evolution of idiosyncratic productivity s_t and employment status $e_t \in \{E, U, B\}$ (employed, unemployed, business owner). They hold two assets, liquid deposits b_{t+1} and illiquid equity a_{t+1} purchased at price q_t that pays real dividends r^a_t. Each period a household dies with probability $\zeta \in (0,1)$ and is replaced by a newborn with the same (s_t, e_t) and zero wealth. The wealth of the deceased is redistributed in proportion to asset holdings among remaining households.

Per period utility follows Greenwood et al. (1988). The specification is

$$u(c_t, n_t \mid s_t, e_t) = \frac{\left[c_t - \psi\, s_t\, \frac{n_t^{1+\xi}}{1+\xi}\right]^{1-\sigma}}{1-\sigma}, \qquad \sigma > 0,\ \xi > 0,\ \psi > 0. \tag{30}$$

This form removes wealth effects on labor supply under employment. Let $\beta \in (0,1)$ be the discount factor. The household problem is

$$\max_{\{c_t, n_t, a_{t+1}, b_{t+1}\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t (1-\zeta)^t \Big\{ u(c_t, n_t \mid s_t, e_t) - \chi_t \mathbb{1}\{a_{t+1} \neq a_t\} \Big\} \tag{31}$$

s.t.

$$c_t + q_t a_{t+1} + b_{t+1} = (1-\tau)\, y_t + (q_t + r^a_t) a_t + (1 + r^b_t) b_t + T_t, \tag{32}$$

$$a_{t+1} \geq 0, \qquad b_{t+1} \geq \underline{b}. \tag{33}$$

Under $e_t = E$, GHH preferences imply uniform hours across employed households,

$$w_t = \psi\, n_t^{\xi} \quad \Rightarrow \quad n_t = \left(\frac{w_t}{\psi}\right)^{1/\xi}. \tag{34}$$

This condition pins down hours for workers, and $n_t = 0$ if $e_t \in \{U, B\}$. Earnings depend on working status through

$$y_t = \begin{cases} w_t s_t n_t, & e_t = E, \\ \upsilon\, w\, \min\{s_t, \bar{s}\}, & e_t = U, \\ \nu\, \Pi_t, & e_t = B, \end{cases} \tag{35}$$

and here w_t is the real wage per efficiency unit, w is its steady state value, υ is the replacement ratio with cap $\bar{s}$, $\nu \in (0,1)$, and Π_t are aggregate profits.

Asset structure Liquid returns depend on borrower versus saver status due to a premium $\iota_b \geq 0$,

$$1 + r^b_t = \begin{cases} \dfrac{1 + i_t}{\pi_t}, & b_t \geq 0, \\ \dfrac{1 + i_t + \iota_b}{\pi_t}, & b_t < 0, \end{cases} \tag{36}$$

and i_t is the nominal policy rate while π_t is gross inflation. Equity is illiquid, so changing a_{t+1} incurs a stochastic utility cost χ_t. Costs are i.i.d. across households and time with logistic CDF $F(\chi) = \left(1 + e^{-(\chi-\mu_\chi)/\sigma_\chi}\right)^{-1}$ with location μ_χ and scale $\sigma_\chi > 0$. This F governs both adjustment probabilities and the average cost conditional on adjustment.

Timing and status transitions At the start of t, employed households separate exogenously with probability λ and become unemployed, and business owners lose their status with probability $\widetilde{P}^e$ and become unemployed. Unemployed households search with endogenous job finding probability f_t. At the end of t, a fraction P^e of non business owners transition into business ownership. Idiosyncratic productivity s_t follows a finite state Markov chain. These flows determine the evolution of the idiosyncratic state.

Portfolio adjustment stage Define cash on hand before portfolio choices as

$$x_t \equiv (q_t + r_t^a)a_t + (1 + r_t^b)b_t + (1-\tau)y_t + T_t. \tag{37}$$

This x_t summarizes resources available for consumption and next period asset positions. Let $V(a_t, b_t \mid s_t, e_t)$ be the ex ante value at the portfolio decision point. Because of adjustment costs, the household either adjusts equity (A) or does not (N),

$$V(a_t, b_t \mid s_t, e_t) = P(a_t, b_t \mid s_t, e_t)\, V^A(a_t, b_t \mid s_t, e_t) + \left[1 - P(a_t, b_t \mid s_t, e_t)\right] V^N(a_t, b_t \mid s_t, e_t). \tag{38}$$

Here $P(a_t, b_t \mid s_t, e_t) = \Pr\{\chi_t \leq \Delta V(a_t, b_t \mid s_t, e_t)\}$ and $\Delta V(a_t, b_t \mid s_t, e_t) \equiv V^A(a_t, b_t \mid s_t, e_t) - V^N(a_t, b_t \mid s_t, e_t)$, which defines the endogenous adjustment decision.

Adjusters If the household adjusts, the associated problem is

$$V^A(a_t, b_t \mid s_t, e_t) = \max_{c_t, n_t, a_{t+1}, b_{t+1}} \; u(c_t, n_t \mid s_t, e_t) - \chi_t + \beta(1-\zeta)\,\mathbb{E}_t\left[V(a_{t+1}, b_{t+1} \mid s_{t+1}, e_{t+1})\right] \tag{39}$$

$$\text{s.t.} \quad c_t + q_t a_{t+1} + b_{t+1} = x_t, \qquad a_{t+1} \geq 0, \qquad b_{t+1} \geq \underline{b}. \tag{40}$$

This program determines $(c_t, n_t, a_{t+1}, b_{t+1})$ for adjusters.

Non adjusters If equity is not adjusted, $a_{t+1} \equiv a_t$ and the problem is

$$V^N(a_t, b_t \mid s_t, e_t) = \max_{c_t, n_t, b_{t+1}} \; u(c_t, n_t \mid s_t, e_t) + \beta(1-\zeta)\,\mathbb{E}_t\left[V(a_t, b_{t+1} \mid s_{t+1}, e_{t+1})\right] \tag{41}$$

$$\text{s.t.} \quad c_t + b_{t+1} = x_t - q_t a_t, \qquad b_{t+1} \geq \underline{b}. \tag{42}$$

This program determines (c_t, n_t, b_{t+1}) for non adjusters.

Adjustment probability The household adjusts if $\chi_t \leq \Delta V(a_t, b_t \mid s_t, e_t)$. The implied probability and conditional mean cost are

$$P(a_t, b_t \mid s_t, e_t) = F\big(\Delta V(a_t, b_t \mid s_t, e_t)\big), \qquad \mathbb{E}_t[\chi_t \mid \text{adjust}] = \frac{\int_{-\infty}^{\Delta V(a_t, b_t \mid s_t, e_t)} \chi f(\chi)\, d\chi}{F(\Delta V(a_t, b_t \mid s_t, e_t))}. \tag{43}$$

Here f is the logistic density and F is its CDF.

Optimality conditions Let u_c denote $\partial u / \partial c$. For either margin (A or N) the deposit condition is

$$u_c(c_t, n_t \mid s_t, e_t) \geq \beta(1-\zeta)\,\mathbb{E}_t[V_b(a_{t+1}, b_{t+1} \mid s_{t+1}, e_{t+1})], \quad \text{with equality if } b_{t+1} > \underline{b}. \tag{44}$$

Conditional on adjustment, the first order condition with respect to equity is

$$u_c(c_t, n_t \mid s_t, e_t)\, q_t = \beta(1-\zeta)\,\mathbb{E}_t[V_a(a_{t+1}, b_{t+1} \mid s_{t+1}, e_{t+1})]\,, \quad \text{with equality if } a_{t+1} > 0. \tag{45}$$

The envelope conditions imply

$$V_b(a_t, b_t \mid s_t, e_t) = (1 + r_t^b)\, u_c(c_t, n_t \mid s_t, e_t), \tag{46}$$

$$V_a(a_t, b_t \mid s_t, e_t) = \begin{cases} (q_t + r_t^a)\, u_c(c_t^A, n_t^A \mid s_t, e_t), & \text{if adjust,} \\ r_t^a\, u_c(c_t^N, n_t^N \mid s_t, e_t) + \beta(1-\zeta)\,\mathbb{E}_t[V_a(a_t, b_{t+1} \mid s_{t+1}, e_{t+1})]\,, & \text{if not adjust.} \end{cases} \tag{47}$$

A.2 Firms

The firm sector has three parts. A competitive final good producer aggregates intermediate goods using a CES technology. A continuum of intermediate goods firms produce intermediate goods by renting capital and labor services and set prices subject to nominal rigidity. A representative capital firm supplies capital services and produces new capital financed by equity investment. In the main text, the final good and intermediate goods producers are referred to as goods producers.

A.2.1 Final good

The final good is a constant elasticity of substitution (CES) aggregator of differentiated intermediate goods. Formally,

$$Y_t = \left[\int_0^1 Y_{jt}^{\frac{\eta_t - 1}{\eta_t}}\,, dj\right]^{\frac{\eta_t}{\eta_t - 1}}, \tag{48}$$

where Y_{jt} is firm j's output and $\eta_t > 1$ is the time-varying elasticity of substitution. Profit maximization by the final good firm delivers the standard demand schedule and the associated price index,

$$Y_{jt} = \left(\frac{P_{jt}}{P_t}\right)^{-\eta_t} Y_t \tag{49}$$

$$P_t = \left[\int_0^1 P_{jt}^{1-\eta_t}\,, dj\right]^{\frac{1}{1-\eta_t}}, \tag{50}$$

where P_{jt} is the nominal price of good j and P_t is the aggregate price level. These expressions imply that a higher η_t makes individual demand more elastic and lowers desired markups.

A.2.2 Intermediate goods producers

There is a continuum of intermediate goods firms that hire capital and labor rental services to produce differentiated goods. Production follows a Cobb–Douglas technology,

$$Y_{jt} = Z_t, K_{jt}^{\theta}, L_{jt}^{1-\theta}, \tag{51}$$

where Z_t is total factor productivity, K_{jt} is the flow of capital services, L_{jt} is labor services, and $\theta \in (0,1)$ is the capital share. Given the demand (49) and technology (51), firm j chooses prices and inputs to maximize the expected present value of profits,

$$\max_{P_{jt},K_{jt},L_{jt}} \sum_{t=0}^{\infty} \mathbb{E}_0 \left[\Lambda_{0,t}, \Pi^I_{jt} \right] \quad \text{with} \quad \Pi^I_{jt} \equiv \frac{P_{jt}}{P_t} Y_{jt} - r^l_t L_{jt} - r^k_t K_{jt} - \Phi^P(P_{jt}, P_{j,t-1}) - \Psi^F_t Y. \tag{52}$$

Here r^l_t and r^k_t are the rental rates of labor and capital services, respectively, Ψ^F_t is a fixed operating cost shock that follows an AR(1) process, and $\Lambda_{t,t+1}$ denotes the average marginal rate of substitution of business owners used to discount profits. Price setting is subject to Rotemberg adjustment costs with indexation,

$$\Phi^P(P_{jt}, P_{j,t-1}) = \frac{\eta_t}{2\kappa} \left\{ \log \frac{P_{jt}}{P_{j,t-1}} - \log \left(\pi^{\iota_p}_{t-1}, \pi'^{1-\iota_p} \right) \right\}^2 Y_t, \tag{53}$$

where $\kappa > 0$ governs the slope of the Phillips curve and $\iota_p \in [0, 1]$ measures the degree of backward-looking indexation.

Under a symmetric equilibrium, all adjusting firms set the same price and the sector aggregates to a New Keynesian Phillips curve,

$$\log \pi_t - \log \left(\pi^{\iota_p}_{t-1}, \pi'^{1-\iota_p} \right) = \mathbb{E}_t \Big[\Lambda_{t,t+1}, \left(\log \pi_{t+1} - \log(\pi^{\iota_p}_t, \pi'^{1-\iota_p}) \right) \Big] + \kappa, \left(\mathrm{MC}_t - \frac{1}{\Psi^p_t} \right), \tag{54}$$

where real marginal cost is

$$\mathrm{MC}_t = \frac{(r^k_t)^\theta (r^l_t)^{1-\theta}}{Z_t} \left(\frac{1}{\theta} \right)^\theta \left(\frac{1}{1-\theta} \right)^{1-\theta}, \qquad \Psi^p_t \equiv \frac{\eta_t}{\eta_t - 1} \tag{55}$$

collects the markup shock implied by time-varying elasticity η_t. Equation (54) embeds both forward-looking price setting via $\Lambda_{t,t+1}$ and backward-looking indexation via ι_p.

A.2.3 Capital firm

A representative capital firm chooses utilization and supplies capital services demanded by investors. For a given physical stock K_t, period profits from utilization are

$$r^k_t v_t K_t - \delta(v_t) K_t, \tag{56}$$

where v_t is the utilization rate and $\delta(\cdot)$ is the depreciation function. The first order condition with respect to utilization equates the rental rate of capital to the marginal depreciation,

$$r^k_t = \delta'(v_t). \tag{57}$$

Depreciation rises convexly with utilization according to

$$\delta(v_t) = \delta_0, v^{\delta_1}_t, \qquad \delta_1 > 1, \tag{58}$$

where δ_0 is the depreciation rate at full utilization.

The firm's investment department converts final goods into new capital subject to adjustment costs and an efficiency shock. Its period payoff is

$$q_t K_{t+1} - \Psi^k_t \left\{ K_{t+1} + \frac{\phi}{2} \left(\log \frac{K_{t+1}}{K_t} \right)^2 K_{t+1} \right\}, \tag{59}$$

where q_t is the price of new capital, $\phi > 0$ controls adjustment costs, and Ψ_t^k captures investment-specific efficiency (MEI). Capital firm's intertemporal optimality condition yields the asset pricing condition for new capital,

$$q_t = \Psi_t^k \left\{ 1 + \phi \log \frac{K_{t+1}}{K_t} + \frac{\phi}{2} \left(\log \frac{K_{t+1}}{K_t} \right)^2 \right\} - \mathbb{E} * t \left[\Lambda_{t,t+1}, \Psi_{t+1}^k, \phi \left(\log \frac{K_{t+2}}{K_{t+1}} \right) \frac{K_{t+2}}{K_{t+1}} \right]. \quad (60)$$

Finally, the implied investment expenditure is

$$\tilde{I}_t = \Psi_t^k K_{t+1} \left\{ 1 + \frac{\phi}{2} \left(\log \frac{K_{t+1}}{K_t} \right)^2 \right\} - \{1 - \delta(v_t)\} K_t. \quad (61)$$

A.3 Labor agencies

Labor agencies intermediate between households and goods producers. They post vacancies, hire households, and supply labor services. The value of a filled match for an agency with a household of productivity s_t is

$$J^L(s_t) = (r_t^l - w_t - \Xi^L)\, s_t n_t + \mathbb{E}_t \Big[\Lambda_{t,t+1} \,(1-\zeta)\,(1-\lambda)\,(1-P^e)\, J^L(s_{t+1}) \Big]. \quad (62)$$

Here r_t^l is the labor rental rate, w_t the real wage, Ξ^L the per match maintenance cost, ζ the death probability, λ the exogenous separation probability, P^e the transition into business ownership, $\Lambda_{t,t+1}$ the business owners' discount factor, and n_t the hours supplied by the matched household.

Wages follow the rule described in the main text,

$$\frac{w_t}{w} = \left\{ \Psi_t^w \left(\tfrac{r_t^l}{r^l} \right) \right\}^{1-\rho_w} \left\{ \frac{w_{t-1}}{w} \left(\tfrac{\pi_{t-1}}{\pi_t} \right)^{\iota_w} \left(\tfrac{\pi}{\pi_t} \right)^{1-\iota_w} \right\}^{\rho_w}, \quad (63)$$

with $0 < \rho_w < 1$ and $0 < \iota_w < 1$, where Ψ_t^w is a wage markup shock and π_t is gross inflation. Vacancy entry is pinned down by free entry,

$$\iota = \frac{M_t}{V_t} \int J^L(s_t)\, d\mu_t(s_t), \quad (64)$$

where ι is the vacancy posting cost, V_t vacancies, M_t matches, and μ_t the cross sectional distribution over s_t.

Let U_t be unemployed at the beginning of period t, N_t employed at the beginning of period t, and define the search pool $S_t \equiv U_t + \lambda N_t$. The number of matches is

$$M_t = \frac{S_t\, V_t}{\left(S_t^\alpha + V_t^\alpha \right)^{1/\alpha}}, \qquad \alpha > 0. \quad (65)$$

This matching function is homogeneous of degree one and bounded above by $\min\{S_t, V_t\}$. Market tightness is $\theta_t \equiv V_t / S_t$. The job finding rate and the vacancy filling rate are

$$f_t = \frac{M_t}{S_t} = \frac{\theta_t}{\left(1 + \theta_t^\alpha\right)^{1/\alpha}}, \qquad p_t^v = \frac{M_t}{V_t} = \frac{1}{\left(1 + \theta_t^{-\alpha}\right)^{1/\alpha}}. \quad (66)$$

These definitions are consistent with the period timing and feed into employment transitions.

A.4 Banks

As long as the expected equity premium $R^a_{t+i} - R_{t+i}$ is positive, a bank's optimal choice is to expand assets. Without a borrowing limit, either assets would grow without bound or the premium would be driven to zero. To discipline leverage, I assume a moral hazard/costly enforcement friction as in Gertler and Karadi (2011). At the beginning of the period, a bank can divert the fraction Δ of its assets and transfer it to business owners. If diversion occurs, depositors force bankruptcy but recover only the remaining $1-\Delta$ fraction of assets. Taking this incentive problem into account, investors supply deposits only up to the point at which the following constraint holds,

$$J^b(N_{jt}) \geq \Delta q_t A^b_{jt+1} \tag{67}$$

where the left-hand side is the franchise value the bank would forfeit by diverting, and the right-hand side is the private benefit from diversion. To make the condition operational, I compute the bank's value. Using a standard guess and verify approach, the value of bank j is linear in assets and net worth,

$$J^b(N_{jt}) = \vartheta^a_t q_t A^b_{jt+1} + \vartheta^n_t N_{jt} \tag{68}$$

with coefficients

$$\vartheta^a_t = \mathbb{E}_t\Big[(1-\theta_b)\Psi^b_t\Lambda_{t,t+1}(R^a_{t+1} - R_{t+1}) + \theta_b\Psi^b_t\Lambda_{t,t+1}x_{t,t+1}\nu_{t+1}\Big] \tag{69}$$

$$\vartheta^n_t = \mathbb{E}_t\Big[(1-\theta_b)\Psi^b_t\Lambda_{t,t+1}R_{t+1} + \theta_b\Psi^b_t\Lambda_{t,t+1}z_{t,t+1}\eta_{t+1}\Big] = (1-\theta_b) + \mathbb{E}_t\Big[\theta_b\Psi^b_t\Lambda_{t,t+1}z_{t,t+1}\eta_{t+1}\Big] \tag{70}$$

where $x_t = q_{t+1}A^b_{jt+2}/q_tA^b_{jt+1}$ is the gross growth rate of assets between t and $t+1$, and $z_t = N_{jt+1}/N_{jt}$ is the gross growth rate of net worth. The aggregate risk premium shock Ψ^b_t follows an AR(1) process,

$$\log\Psi^b_t = \rho_b \log\Psi^b_{t-1} + \epsilon_{b,t} \ , \ \epsilon_{b,t} \sim N(0,\sigma^2_b) \tag{71}$$

so that a positive innovation raises the value of future payoffs and tends to expand bank balance sheets.

With the value function in hand, the incentive constraint becomes

$$\vartheta^a_t q_t A^b_{jt+1} + \vartheta^n_t N_{jt} \geq \Delta q_t A^b_{jt+1} \tag{72}$$

and, when it binds, the banker's asset position is tied to net worth,

$$q_t A^b_{jt+1} = \frac{\vartheta^n_t}{\Delta - \vartheta^a_t} N_{jt} = \Theta_t N_{jt} \tag{73}$$

where Θ_t is the leverage ratio (assets to net worth).[50] When the constraint binds, net worth evolves according to

$$N_{jt+1} = \big\{(R^a_{t+1} - R_{t+1})\Theta_t + R_{t+1}\big\}N_{jt} \tag{74}$$

[50] Given $N_{jt} > 0$, the constraint binds only if $0 < \vartheta^a_t < \Delta$. Under the parameterizations used in this paper, the constraint always binds.

and the corresponding gross growth rates satisfy

$$z_{t,t+1} = N_{jt+1}/N_{jt} = \{(R^a_{t+1} - R_{t+1})\Theta_t + R_{t+1}\} \tag{75}$$

$$x_{t,t+1} = q_{t+1}A^b_{jt+2}/q_tA^b_{jt+1} = \Theta_{t+1}N_{jt+1}/\Theta_tN_{jt} = (\Theta_{t+1}/\Theta_t)z_{t,t+1} \tag{76}$$

which will be useful below.

Since Θ_t does not depend on bank-specific states, I can aggregate across banks to obtain

$$q_tA^b_{t+1} = \Theta_tN_t \tag{77}$$

where A^b_{t+1} is aggregate bank equity holdings and N_t is aggregate bank net worth. For aggregate dynamics, write N_t as the sum of net worth for surviving banks, N_{ot} (old), and for entrants, N_{et} (new). I assume the startup funds for new banks equal the value of assets previously intermediated by exiting banks, $(1-\theta_b)q_{t-1}A^b_t$, and that each entrant receives the fraction $\omega/(1-\theta_b)$ of this value from the equity mutual fund. It follows that

$$N_t = N_{ot} + N_{et} = \theta_b\{(R^a_t - R_t)\Theta_{t-1} + R_t\}N_{t-1} + \omega q_{t-1}A^b_t \tag{78}$$

and profits distributed by the financial sector are net of startup funds,

$$\Pi^b_t = (1-\theta_b)\{(R^a_t - R_t)\Theta_{t-1} + R_t\}N_{t-1} - \omega q_{t-1}A^b_t \tag{79}$$

which completes the characterization of the banking sector.

A.4 Mutual funds

Two mutual funds operate in the background. The equity mutual fund channels equity issuance to the capital firm and pays out all profits to business owners and equity holders under perfect competition. The money market mutual fund provides liquidity to banks by investing in deposits using small government contributions and investment income, and returns the remainder to households as smoothed lump sum transfers. It is subject to a liquidity preference shock.

A.4.1 Equity mutual fund

There exists a hypothetical mutual fund that owns all firms in the model. To distinguish it from the other mutual fund, call it the equity mutual fund. Its roles are to collect profits from firms, pay dividends to shareholders, and issue new equity for capital accumulation. The fund operates in perfect competition, so there are no retained earnings and all profits are paid as dividends. Funds raised by issuing equity are transferred to the capital firm to purchase new capital. The period cash-flow constraint is

$$(1-\tau)(1-\nu)\Pi_t - q_t\left(K_{t+1} - K_t\right) + q_t\left(A_{t+1} - A_t\right) = r^a_tA_t, \tag{80}$$

where Π_t is total profits, ν is the share of profits paid directly to business owners,[51] and τ is the tax rate on profits. Since aggregate capital equals aggregate equity in the model, the equity price equals the price of new capital and the dividend rate is

$$r^a_t = (1-\tau)(1-\nu)\,\Pi_t/K_t. \tag{81}$$

[51]The fund is owned by business owners, so a fraction of profits is distributed to them regardless of their equity holding.

Aggregate profits decompose into the firm side (which includes labor agencies) and the banking sector,

$$\Pi_t = \Pi_t^f + \Pi_t^b. \tag{82}$$

A.4.2 Money market mutual fund

The money market mutual fund provides liquidity to the financial sector. It receives contributions from the government and invests in liquid assets; with these funds and investment proceeds it makes lump-sum transfers to households and smooths the flow of those transfers. Its objective is

$$\max_{\{T_t^m, B_{t+1}^m\}} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \Psi_t^l \beta_m^t \frac{(T_t^m)^{1-\sigma}}{1-\sigma} \right] \tag{83}$$

subject to the period budget constraint

$$T_t^m + B_{t+1}^m = C_t^g + (1 + r_t^b)\, B_t^m, \tag{84}$$

where T_t^m is the lump-sum transfer to households, B_{t+1}^m the fund's liquid asset position, and C_t^g the government contribution. The fund discounts with its own factor β_m and faces an AR(1) liquidity preference shock Ψ_t^l.[52]

A.5 Government

Fiscal authority. The fiscal authority collects taxes and issues bonds to finance government purchases, unemployment benefits, lump sum transfers, and contributions to the money market mutual fund. To ensure price level determinacy, I assume that the fiscal authority controls its debt according to a simple autoregressive rule, as in Woodford (1995),

$$\frac{B_{t+1}^g}{B^g} = \left(\frac{R_t/\pi_t \times B_t^g}{R/\pi \times B^g} \right)^{\rho_B}, \; 0 \le \rho_B < 1, \tag{85}$$

where $\rho_B \in (0, 1)$ is the pace of debt adjustment.

Since agents form rational expectations, the government must satisfy the intertemporal budget constraint

$$B_t^g = \sum_{l \ge t}^{\infty} \left\{ \prod_{i=t}^{l} \left(\frac{\pi_i}{R_i} \right) \right\} \left\{ \mathbb{T}_l - (G_l + T_l^g + D_l + T_l^{CB} + C_l^g) \right\}, \tag{86}$$

where $\mathbb{T}$, G, T^g, D, and C^g denote tax revenues, government purchases, lump sum transfers (or taxes) to households, unemployment benefits, and contributions to the MMMF, respectively. The net remittance from the monetary authority is

$$T_t^{\mathrm{CB}} = q_t A_{t+1}^{\mathrm{CB}} - (q_t + r_t^a) A_t^{\mathrm{CB}} + R_t B_t^{\mathrm{CB}} - B_{t+1}^{\mathrm{CB}}. \tag{87}$$

Equation (86) implies that in each period the debt level equals the present discounted value of future primary surpluses. When the real value of government debt changes, at least one

[52] In principle the fund's intertemporal elasticity of substitution need not equal that of households; for parsimony they are set equal here. In steady state the fund's optimality condition implies a discount factor equal to the inverse of the real interest rate. With idiosyncratic risk, the average business-owner MRS is not equal to the inverse of the real rate in steady state. A liquidity-preference shock is used instead of a household preference shock to keep welfare comparisons tractable.

fiscal instrument adjusts. I assume the fiscal authority adjusts its contribution to the MMMF to balance the budget each period, while government purchases are fixed and lump sum transfers to households vary according to

$$T_t^g = \left(1 - \frac{1}{\Psi_t^g}\right) Y, \tag{88}$$

where Ψ_t^g is a lump sum transfer shock and Y is steady state output.[53]

Monetary authority. The monetary authority sets the policy rate by a Taylor rule with interest rate smoothing and an effective lower bound (ELB),

$$\frac{1+\hat{i}_{t+1}}{1+\hat{i}} = \left(\frac{1+\hat{i}_t}{1+\hat{i}}\right)^{\rho_R} \left[\left(\frac{\pi_t}{\pi}\right)^{\phi_\pi} \exp\{-\phi_u(u_t - u)\}\right]^{1-\rho_R} \exp(\epsilon_{R,t}), \qquad i_{t+1} = \max\{0, \hat{i}_{t+1}\}, \tag{89}$$

where $\epsilon_{R,t} \sim \mathcal{N}(0, \sigma_R^2)$ is a monetary policy shock, $0 < \rho_R < 1$ governs smoothing, and ϕ_π, ϕ_u are the inflation and unemployment-gap coefficients. When the ELB binds, unconventional tools become active.

Quantitative easing (QE). The central bank purchases illiquid assets from the private sector by issuing bonds. QE is modeled as an exogenous AR(1) process that scales the balance sheet toward a target level,

$$A_{t+1}^{\text{CB}} = \Psi_t^{\text{QE}} A^{\text{CB}} \;\;, \;\; B_{t+1}^{\text{CB}} = q_t A_{t+1}^{\text{CB}}, \tag{90}$$

$$\log \Psi_t^{\text{QE}} = \rho_{\text{QE}} \log \Psi_{t-1}^{\text{QE}} + \epsilon_{\text{QE},t}, \qquad \epsilon_{\text{QE},t} \sim \mathcal{N}(0, \sigma_{\text{QE}}^2). \tag{91}$$

This specification allows counterfactuals that shut down QE (setting $\epsilon_{\text{QE},t} = 0$), keeping the balance sheet unchanged during ELB episodes; a spread feedback rule would induce endogenous balance sheet movements even without shocks.

Forward guidance (FG). Forward guidance is active only when the ELB binds.[54] I treat an ELB episode as a temporary regime with an exogenous expected binding duration T_t^{FG}. During $k = 0, \ldots, T_t^{FG} - 1$, agents expect the policy rate to remain at the ELB, which shifts expected real rates and affects current demand via intertemporal substitution. If T_t^{FG} exceeds the endogenous binding duration implied by fundamentals under no future shocks, the extra periods are equivalent to anticipated expansionary policy rate innovations. The implementation follows the backward solution approach in Kulish et al. (2014) and Jones (2017). Endogenous ELB durations are computed following Guerrieri and Iacoviello (2015).

Model solution and forward guidance specification. The equilibrium conditions are log linearized around the steady state. The linearized system takes the form

$$A\, \mathbb{E}_t[X_{t+1}] + B\, X_t + C\, X_{t-1} + E\, \epsilon_t = 0 \quad \Longrightarrow \quad X_t = P\, X_{t-1} + Q\, \epsilon_t, \tag{92}$$

where X_t stacks endogenous variables (deviations from steady state) and ϵ_t stacks exogenous shocks.

When the ELB binds, I solve a piecewise linear system with a temporary ELB regime of

[53] Because markets are incomplete and households value liquidity, the model is non-Ricardian, so fiscal responses matter for distributional effects. With only short term government debt these effects can be strong Lee (2019). The assumption that the MMMF smooths lump sum transfers dampens these responses.

[54] In principle, QE can operate even when the ELB does not bind, though here it operates only at the ELB.

expected duration T_t^{FG} that transitions back to the normal regime. Over the ELB window the policy rate is fixed at the lower bound, and the terminal conditions at $t + T_t^{FG}$ pin down the return to the normal regime. Denoting ELB regime coefficients with tildes, the endpoint system satisfies

$$\tilde{A}\,\mathbb{E}_{t+T_t^{FG}}[X_{t+T_t^{FG}+1}] + \tilde{B}\,X_{t+T_t^{FG}} + \tilde{C}\,X_{t+T_t^{FG}-1} + \tilde{D} = 0, \qquad X_{t+T_t^{FG}+1} = P\,X_{t+T_t^{FG}}. \tag{93}$$

Backward substitution to period t then delivers reduced-form policy functions

$$X_t = P(T_t^{FG})\,X_{t-1} + J(T_t^{FG}) + Q(T_t^{FG})\,\epsilon_t,$$

and forward guidance corresponds to choosing T_t^{FG} above the endogenous binding duration implied by fundamentals under no future shocks.

A.6 Market clearing and equilibrium

This subsection collects the market-clearing conditions and the equilibrium definition. Let $C_t \equiv \int c_t\,d\mu_t$, $B_t \equiv \int b_t\,d\mu_t$, $A_t^h \equiv \int a_t\,d\mu_t$, and $N_t \equiv \int s_t n_t\,\mathbf{1}\{e_t = E\}\,d\mu_t$. The mass of unemployed and employed households at the beginning of t are U_t and N_t, and the search pool is $S_t = U_t + \lambda N_t$.

Goods market Final output equals absorption and resource costs,

$$Y_t \;=\; C_t \;+\; G_t \;+\; \tilde{I}_t \;+\; \Phi_t^P \;+\; F_t, \tag{94}$$

where $\tilde{I}_t$ is investment expenditure from (60)–(capital block), Φ_t^P is the Rotemberg price adjustment cost, and F_t is the fixed operating cost (specified in the firm block as a fraction of steady-state output).[55]

Labor services Labor services supplied by households through agencies equal the services rented by goods producers,

$$L_t^s \;=\; \int s_t n_t\,\mathbf{1}\{e_t = E\}\,d\mu_t \;=\; L_t^d. \tag{95}$$

Capital services and equity Capital utilization satisfies $K_t^d = v_t K_t$. Aggregate equity equals the capital stock,

$$A_t \;=\; K_t, \qquad A_{t+1}^h \;+\; A_{t+1}^b \;=\; A_{t+1}, \tag{96}$$

where A_{t+1}^b is banks' equity position and $A_{t+1}^h \equiv \int a_{t+1}\,d\mu_t$ is households' equity position. The equity mutual fund enforces $q_t A_{t+1} = q_t K_{t+1}$ via its cash-flow constraint.

Liquid asset market Liquid assets held by households and the money market mutual fund match liquid liabilities issued by banks, the fiscal authority, and the central bank,

$$B_{t+1} \;+\; B_{t+1}^m \;=\; B_{t+1}^b \;+\; B_{t+1}^g \;+\; B_{t+1}^{\mathrm{CB}}, \tag{97}$$

where $B_{t+1} \equiv \int b_{t+1}\,d\mu_t$ (household deposits), B_{t+1}^m (MMMF liquid assets), B_{t+1}^b (bank deposits outstanding), B_{t+1}^g (government bonds), and B_{t+1}^{CB} (central bank liabilities). As noted in the main text, households and the MMMF treat deposits and government bonds as a single liquid asset.

[55] The utility cost χ_t of equity adjustment is a taste cost and does not absorb final goods.

Government and central bank The fiscal budget constraint is given in (22). Central bank remittances to the fiscal authority are

$$T_t^{CB} = q_t A_{t+1}^{CB} - (q_t + r_t^a) A_t^{CB} + R_t B_t^{CB} - B_{t+1}^{CB}. \tag{98}$$

Equilibrium Given policy rules, exogenous processes, and initial conditions, a (competitive) equilibrium is sequences of allocations, prices, and distributions

$$\{c_t, n_t, a_{t+1}, b_{t+1}\}_{\text{hh}}, \ \{P_{jt}, Y_{jt}, K_{jt}, L_{jt}\}_{\text{firms}}, \ \{K_{t+1}, v_t, q_t\}_{\text{capital}}, \ \{A_{t+1}^b, B_{t+1}^b, N_{t+1}\}_{\text{banks}},$$

together with $\{A_{t+1}, B_{t+1}^g, B_{t+1}^{CB}, B_{t+1}^m, w_t, r_t^l, r_t^k, r_t^b, r_t^a, \pi_t, i_t\}$ and the law of motion for μ_t, such that (i) households, firms, labor agencies, banks, and funds satisfy their optimality conditions and constraints; (ii) policy rules for the monetary and fiscal authorities are satisfied; and (iii) the market-clearing conditions (94)–(98) hold.

B Numerical method

B.1 Solution method

For the calibration, I solve for the steady state of the model globally. Specifically, I use value function iteration combined with the endogenous grid method of Carroll (2006) to compute households' policy functions. Then, I find the invariant distribution using the non-stochastic simulation method of Young (2010) with the representation of the idiosyncratic distribution as histograms. The solution method captures the precautionary motive associated with idiosyncratic shocks as they are still present even though the model is at the steady state, and there are no aggregate shocks.

Once the steady state is found, I solve for the dynamics of the model using a perturbation method developed by Reiter (2009) with a state-space reduction technique proposed by Bayer and Luetticke (2020).[56] The methodology enables a fast solution that is necessary for Bayesian estimation. However, since the state-space is much larger compared to a representative model even after the reduction, estimating the model by solving the dynamics in full each time during the process is still not feasible.[57] Thus, one needs a way to accelerate the solution process.

On this regard, I follow Bayer et al. (2020) and update only a subset of the Jacobian during the estimation process. The system of equations that characterize an equilibrium can be expressed as follows.

$$\mathbb{E}_t \left[\mathcal{F}(X_{t+1}, Y_{t+1}, X_t, Y_t) \right] = 0 \tag{99}$$

where $\mathcal{F}$ is a non-linear function that consists of equilibrium conditions and laws of motion for relevant objects including the idiosyncratic distribution. $\mathbb{E}_t$ is the expectation operator conditional on the information available at period t. $X_{t+1} = (X_{1t+1}, X_{2t+1}, X_{3t+1}, \epsilon_{t+1})'$ is the vector of pre-determined or state variables. Specifically, X_{1t+1} is the vector of "idiosyncratic" state variables. In my model, X_{1t+1} consists of households' idiosyncratic state distribution at

[56] Bayer and Luetticke (2020) approximate the deviation of value functions from their steady state values using Chebyshev polynomials, and use a fixed copula for the approximation of changes in the idiosyncratic distributions.

[57] On a workstation computer with 10 cores (20 threads), it takes about 40 seconds to solve the dynamics model when 17,600 ($40 \times 40 \times 11$) points were used to represent the idiosyncratic state space.

the end of period t.[58] X_{2t+1} is the vector of "summary" variables, which includes aggregate bond and equity holding of households. Variables X_{2t+1} summarize the idiosyncratic decision of households into one scalar variable. Importantly, the relationship between idiosyncratic state and variables in X_{2t+1} is not affected by parameter values. X_{3t+1} is the vector of purely "aggregate" variables in the sense that idiosyncratic variables do not appear in the equations that define these variables. ϵ_{t+1} is the vector of all exogenous shocks. Y_t is the vector of endogenous control variables and further decomposed into Y_{1t+1}, Y_{2t+1}, and Y_{3t+1}. Y_{1t+1} is the vector of "idiosyncratic" control variables, which include the value functions and their derivatives. Y_{2t+1} is the vector of "summary" variables. Finally, Y_{3t+1} is the vector of "aggregate" variables.

The key idea of Bayer et al. (2020) is that one does not need to update the Jacobian with respect to "idiosyncratic" variables during the estimation if the estimated parameters are only relevant for the dynamics and do not affect households' problem. To this point more clearly, I write down the system of equations (99) as follows.

$$\mathbb{E}_t\left[\mathcal{F}(X_{t+1}, Y_{t+1}, X_t, Y_t)\right] = \left[\mathcal{F}_{1,t}, \mathcal{F}_{2,t}, \mathcal{F}_{3,t}, \mathcal{F}_{4,t}, \mathcal{F}_{5,t}, \mathcal{F}_{6,t}, \mathcal{F}_{7,t}\right]' \tag{100}$$

where $\mathcal{F}_{1,t}$ is the set of equations that describe relations among idiosyncratic state variables, i.e., between X_{1t} and X_{1t+1}. $\mathcal{F}_{2,t}$ is summary equations that aggregate individual variables into aggregate state variables. Note that $\mathcal{F}_{1,t}$ is affected only by parameters that alter households' optimal behaviors. Likewise, $\mathcal{F}_{2,t}$ is not affected by parameter choice as they are aggregation of individual variables over idiosyncratic state space. $\mathcal{F}_{3,t}$ is the set of equations for aggregate variables. Importantly, idiosyncratic state variables, i.e., ones in X_{1t}, do not appear in $\mathcal{F}_{3,t}$. Instead, variables in $X_{2,t}$ may appear in $\mathcal{F}_{3,t}$. $\mathcal{F}_{4,t}$ is the exogenous stochastic processes.

The remaining three sets of equations describe relations regarding control variables. $\mathcal{F}_{5,t}$ is the set of equations on idiosyncratic control variables. In the model, such variables include value functions and their derivatives. Again, parameters that are not relevant for households' problem do not affect these equations. $\mathcal{F}_{6,t}$ is summary equations regarding control variables.[59] Again, changes in parameters that are not relevant for households' problem do not affect these two sets of equations. Finally, $\mathcal{F}_{7,t}$ is the set of equations on aggregate variables. Note that idiosyncratic state and control variables appear in $\mathcal{F}_{7,t}$ only through summary variables.

From equation (100), we know that the Jacobian has the following form.

$$\mathcal{J}_t = \begin{bmatrix} \frac{\partial \mathcal{F}_{1,t}}{\partial X_{t+1}} & \frac{\partial \mathcal{F}_{1,t}}{\partial Y_{t+1}} & \frac{\partial \mathcal{F}_{1,t}}{\partial X_t} & \frac{\partial \mathcal{F}_{1,t}}{\partial Y_t} \\ \frac{\partial \mathcal{F}_{2,t}}{\partial X_{t+1}} & \frac{\partial \mathcal{F}_{2,t}}{\partial Y_{t+1}} & \frac{\partial \mathcal{F}_{2,t}}{\partial X_t} & \frac{\partial \mathcal{F}_{2,t}}{\partial Y_t} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \mathcal{F}_{7,t}}{\partial X_{t+1}} & \frac{\partial \mathcal{F}_{7,t}}{\partial Y_{t+1}} & \frac{\partial \mathcal{F}_{7,t}}{\partial X_t} & \frac{\partial \mathcal{F}_{7,t}}{\partial Y_t} \end{bmatrix} \tag{101}$$

where $\frac{\partial \mathcal{F}_{j,t}}{\partial X_l} = \left[\frac{\partial \mathcal{F}_{j,t}}{\partial X_{1l}}, \frac{\partial \mathcal{F}_{j,t}}{\partial X_{2l}}, \frac{\partial \mathcal{F}_{j,t}}{\partial X_{3l}}, \frac{\partial \mathcal{F}_{j,t}}{\partial \epsilon_l}\right]$, and $\frac{\partial \mathcal{F}_{j,t}}{\partial Y_l} = \left[\frac{\partial \mathcal{F}_{j,t}}{\partial Y_{1l}}, \frac{\partial \mathcal{F}_{j,t}}{\partial Y_{2l}}, \frac{\partial \mathcal{F}_{j,t}}{\partial Y_{3l}}\right]$ for $l = t$ and $t+1$. During Bayesian estimation, we need to update the Jacobian to compute a likelihood of the model for given data and for a given set of parameters. Since the dimension of the Jacobian is very large, updating the Jacobian is time-consuming. However, we do not need to update all the

[58] Note that the endogenous state variables for period $t+1$ are determined in period t.

[59] For instance, the aggregate consumption and saving are the sum of individual consumption and saving.

blocks in the Jacobian every time if we estimate parameters and shock processes that are only relevant for the dynamics of the model and do not directly affect households' optimal behaviors. Specifically, we only need to update the following derivatives: $\frac{\partial \mathcal{F}_{3,t}}{\partial X_{2t+1}}$, $\frac{\partial \mathcal{F}_{3,t}}{\partial X_{3t+1}}$, $\frac{\partial \mathcal{F}_{3,t}}{\partial \epsilon_{t+1}}$, $\frac{\partial \mathcal{F}_{3,t}}{\partial Y_{2t+1}}$, $\frac{\partial \mathcal{F}_{3,t}}{\partial Y_{3t+1}}$, $\frac{\partial \mathcal{F}_{3,t}}{\partial X_{2t}}$, $\frac{\partial \mathcal{F}_{3,t}}{\partial X_{3t}}$, $\frac{\partial \mathcal{F}_{3,t}}{\partial \epsilon_{t}}$, $\frac{\partial \mathcal{F}_{3,t}}{\partial Y_{2t}}$, $\frac{\partial \mathcal{F}_{3,t}}{\partial Y_{3t}}$, $\frac{\partial \mathcal{F}_{4,t}}{\partial \epsilon_{t}}$, $\frac{\partial \mathcal{F}_{7,t}}{\partial X_{2t+1}}$, $\frac{\partial \mathcal{F}_{7,t}}{\partial X_{3t+1}}$, $\frac{\partial \mathcal{F}_{7,t}}{\partial \epsilon_{t+1}}$, $\frac{\partial \mathcal{F}_{7,t}}{\partial Y_{2t+1}}$, $\frac{\partial \mathcal{F}_{7,t}}{\partial Y_{3t+1}}$. Then, the number of equations that we need to evaluate is close to the number of equations in a representative model with the same features. Thus, estimating the model using Bayesian method is possible.

B.2 Inversion filter

In this paper, I use an inversion filter to back out the structural shocks, following Guerrieri and Iacoviello (2017) and Cuba-Borda et al. (2019). Let $\mathbb{Y}_{\{1:T\}} = \{Y_1, Y_2, ..., Y_T\}$ denote the set of observables, where Y_j is the $n_y \times 1$ vector that contains the data on n_y observables in period j for $j = 1$, ... T. Also, denote the set of all the endogenous variables of the model in period t with the $n_x \times 1$ vector X_t. Similarly, $_t$ is the $n_\epsilon \times 1$ vector of structural shocks in period t. With these notations, one can describe a general form of the solution of the model in period t as follows.

$$X_t = P_t X_{t-1} + D_t + Q_t \epsilon_t \tag{102}$$

where P_t, D_t, and Q_t are the matrices of coefficients in the solution. As time subscripts imply, the coefficients in the solution can be time-varying. However, when the model is at the reference regime, i.e., when the ZLB is not binding in the data, these coefficients are not time-varying and one can compute them by applying a standard perturbation method. Specifically, we have

$$X_t = P X_{t-1} + Q \epsilon_t \tag{103}$$

when the ZLB is not binding. Let H_t be a $n_y \times n_x$ vector that selects the variables in the model that correspond to the observables.[60] Then,

$$Y_t = H_t X_t = H_t P X_{t-1} + H_t Q \epsilon_t \tag{104}$$

From the above equation, one can easily compute the set of structural shocks ϵ_t as follows given that the matrix $H_t Q$ is invertible.

$$\epsilon_t = (H_t Q)^{-1}(Y_t - H_t P X_{t-1}) \tag{105}$$

During the ELB periods, finding ϵ_t can be more demanding task since the matrices P_t, D_t, and Q_t depend not only on the state and structural shocks but also on the expectation on the duration of the ZLB episodes. However, if one assumes an exogenous duration of the ZLB, one can easily compute ϵ_t as follows.

$$\epsilon_t(\tilde{T}_t) = \{H_t Q(\tilde{T}_t)\}^{-1}\Big\{Y_t - H_t P(\tilde{T}_t) X_{t-1} - H_t D(\tilde{T}_t)\Big\} \tag{106}$$

where $\tilde{T}_t$ is the expected ZLB durations in period t. Note that the solution and the corresponding structural shocks are conditional on the duration T of the ZLB episodes. Once I find the series of shocks using the filter, I compute the likelihood of the model given the data

[60] As the data on the central bank's asset is only available since 2003, I include the variable as an observable only during those periods. Accordingly, I only introduce QE shocks during the same periods as well.

as follows.

$$\log p\big(\mathbb{Y}_{\{1:T\}}\big) = -\frac{Tn_y}{2}\log(2\pi) - \frac{T}{2}\log(\det(\Sigma)) - \frac{1}{2}\sum_{t=1}^{T}\epsilon_t'\Sigma^{-1}\epsilon_t + \sum_{t=1}^{T}\log\left(|\det\frac{\partial\epsilon_t}{\partial Y_t}|\right) \tag{107}$$

where $\frac{\partial\epsilon_t}{\partial Y_t} = \{H_t Q_t\}^{-1}$.[61]

C Further details on the calibration

C.1 Calibration targets and parametrization

I set the inverse of the elasticity of intertemporal substitution to 1.5, one of the standard values used in the literature. The discount factor is internally calibrated to match the mass of wealthy hand-to-mouth households, i.e., households with positive illiquid but zero liquid assets, which is 20 percent in the data.[62] The inverse of the Frisch elasticity of labor supply is set to 3, based on Chetty et al. (2011). The disutility of labor is set to ensure that employed households supply one unit of labor at the steady state. The probability of death implies an average working lifespan of 45 years as in Kaplan et al. (2018).

The distribution of illiquid asset adjustment costs affects the average adjustment frequency and the inequality of illiquid asset holdings in the model. The calibrated adjustment costs imply an average adjustment frequency of 6.7 percent per quarter at the steady state, close to 6.5 percent in Bayer et al. (2020). With these costs, the top 10 percent hold 65 percent of total illiquid assets in the model (76 percent in the data).

The income process, which is the ultimate source of inequality in the model, is specified to match asset holdings and wealth inequality in the data. I model the process for s_t as a standard AR(1) with three states, using the Tauchen (1986) method for discretization. I set the autocorrelation and the innovation standard deviation of the quarterly income process to 0.98 and 0.02, based on Storesletten et al. (2004). In addition to this standard part, I add two boundary states, one very low skill and one very high skill, to match wealth inequality in the data. I fix the probability of becoming a business owner P_e at 0.05 percent, which is similar to the value used in Bayer et al. (2019). I then calibrate the probability of leaving the business owner state, which captures income risk for top earners, to match the top 10 percent share of liquid assets. The resulting value for $\tilde{P}_e$ is 20.6 percent.

I set the exogenous job separation rate at 10 percent, following Haan et al. (2000). The steady state real wage is set to 1.2112 so that the ratio of labor income to output, net of fixed costs, is 60 percent at the steady state. I target a vacancy filling rate of 70 percent based on Haan et al. (2000) and Christiano et al. (2016). The target for the steady state unemployment rate is 5.5 percent, which is the average unemployment rate before the Great Recession in my sample. Matching these targets, for the given separation rate, the steady state real wage, and the vacancy posting cost imply a matching efficiency of 1.7127 and a match maintenance cost of 0.0056.[63]

For goods producers, I set the steady state elasticity of substitution to 3, following Gornemann et al. (2016). A relatively low elasticity of substitution implies a high steady state markup, which allows for a substantial fixed cost in production. Given labor agency profits

[61] The result is based on the local linearity of the solution. For more details, see Guerrieri and Iacoviello (2015).

[62] In the data, I define zero assets as assets whose value is less than $2,000.

[63] I estimate the vacancy posting cost and adjust Ξ^L to satisfy free entry for the given labor market parameters. The value reported in Table 1 corresponds to the vacancy posting cost at the posterior mode.

and other firms' profits, I set the fixed cost to match a capital to output ratio in the data.[64][65] The exponent on capital in the production function is set to 0.27, which implies a capital share, defined as profits of intermediate goods firms plus capital rental payments relative to output net of fixed costs, of 40 percent.

The parameters associated with variable capital utilization are calibrated to match two targets, the steady state utilization rate and the depreciation rate. Following standard practice, I set the steady state utilization rate to 1. I then target a steady state depreciation rate of 6 percent on an annualized basis. Matching these targets results in $\delta_0 = 0.015$ and $\delta_1 = 1.0025$.

For the financial sector parametrization, I follow Gertler and Karadi (2011). I target a steady state leverage ratio of 3, which implies $\Delta = 0.3410$. The survival rate of banks is 0.97, and ω is set to 0.0076 to match the bank equity share of 55 percent. The money market mutual fund discount factor is set to ensure that the steady state intertemporal optimality condition holds for a given real return on liquid assets.[66] The fraction of tax revenues given to the fund is set to 5.33 percent to ensure a tax rate of 30 percent while matching the share of lump sum transfers in the income of the bottom 80 percent. Finally, the fraction of firms' profits paid to business owners is set to 23.89 percent, which, together with the probability of becoming a business owner, contributes to the overall wealth inequality in the model.

For the government sector, I mostly use standard values. The replacement ratio is set to 40 percent. The tax rate is 30 percent. The levels of government purchases and lump sum transfers are set to match the share of transfer income in the bottom 80 percent of households' income and a tax rate of 30 percent. The borrowing premium of 2.53 percent is chosen to help match the mass of households with zero assets. The borrowing limit is set to match the fraction of households with debt in the data.

The central bank inflation target is set to 1.005, corresponding to an annual inflation rate of about 2 percent. The steady state policy rate is calibrated so that the model matches the household liquid to illiquid asset ratio in the data. Central bank assets are set equal to 5 percent of output in the steady state, based on the historical average before the implementation of QE. The autocorrelation of central bank assets is fixed at 0.99.[67]

[64] Aggregate capital is the current cost net stock of private fixed assets from the BEA; consumer durables are excluded.

[65] During estimation, the vacancy posting cost varies. To ensure free entry, I adjust Ξ^L. This alters the steady state value of labor agencies, which affects aggregate profits and the dividend rate. To maintain the steady state dividend rate, I also adjust the fixed cost for intermediate goods firms along with Ξ^L. Table 1 reports the fixed cost at the posterior mode of the vacancy posting cost.

[66] At the steady state, $1 = \beta_m R$ should hold, where R is the steady state gross real interest rate.

[67] If estimated, the persistence of central bank assets would be very close to one because the ELB period features a long sequence of balance sheet expansions. To avoid an essentially explosive process in a short sample, I fix the autocorrelation at 0.99.

Table OA1: Calibrated parameters

Parameter	Value	Description	Reference or targets
Households			
σ	1.5	Relative risk aversion	Standard value
β	0.9932	Household discount factor	Mass of wealthy hand-to-mouth
ξ	3	Inverse Frisch elasticity	Chetty et al. (2011)
ψ	0.8476	Disutility of labor	SS hours = 1
ζ	1/180	Probability of death	Average life span 45 years
μ_χ	9.0490	Mean of χ dist	SS adjustment prob 6.5%
σ_χ	3.4205	Scale of χ dist	Top 10% illiquid share
P_e	0.05%	Prob. of becoming business owner	Bayer et al. (2019)
$\tilde{P}_e$	20.6%	Prob. of becoming worker	Top 10% liquid share
Labor market			
λ	0.1	Separation rate	Haan et al. (2000)
w	1.2112	SS real wage	Labor share 60%
α	1.7127	Matching efficiency	Vacancy filling 70%
Ξ^L	0.0056	Match maintenance cost	Unemployment 5.5%
Goods producers			
η	3	Elasticity of substitution	Gornemann et al. (2016)
θ	0.27	Capital share	Capital share 40%
Ξ/Y	0.2013	Fixed cost over output	$K/Y = 3.03$
Capital firm			
δ_0	0.0150	Depreciation rate	6% annual
δ_1	1.0025	Depreciation elasticity	Utilization = 1
Banks			
$\tilde{\Lambda}$	$(1+i)/\bar{\pi}$	MMMF discount factor	SS optimality
τ_m	0.0533	Government contribution share	Transfers/Y = 0.1
Δ	0.3410	Limited enforcement	Bank leverage = 3
θ_b	0.97	Bank survival rate	Gertler and Karadi (2011)
ω	0.0076	New bank net worth	Banks' equity share 55%
ν	0.2380	Profit share to business owners	Gini net worth
Government			
τ	0.30	Tax rate	Data
υ	0.4	Replacement ratio	Standard value
ι_b	0.0253	Borrowing premium	Mass with zero assets
$\underline{b}$	1.3006	Borrowing limit	Mass with debt
Central bank			
π	1.0050	Inflation target	Fed target
$1+i$	1.0100	SS nominal rate	Liquid/illiquid ratio
A^{CB}/Y	0.05	CB assets over output	Data
ρ^{QE}	0.99	QE shock autocorrelation	-

C.2 Fit of the model

Targeted moments and model fit Tables OA4 and OA5 summarize key moments for the household wealth distribution in the data and in the calibrated model. The calibration targets a small set of moments, including the capital to output ratio, the liquid to illiquid asset ratio, the mass of households with negative or zero liquid assets, and top wealth shares for liquid and illiquid assets. The model reproduces these moments reasonably well. For example, the capital to output ratio is 3.02 in the model compared with 3.04 in the data, the liquid to illiquid asset ratio is 0.08 versus 0.11, and the net worth Gini is 0.80 versus 0.82. The fractions of households with negative liquid assets, with zero liquid and positive illiquid assets, and with zero holdings of both assets are also close, with model values of 0.20, 0.22, and 0.05 compared with 0.14, 0.20, and 0.08 in the data.

Table OA5 shows that the model captures the main features of wealth concentration by asset type. For liquid assets, the top 10 percent share is 87 percent in the model and 81 percent in the data, and the top 1 percent share is matched almost exactly at 43 percent. The model understates concentration at the very top, with a top 0.1 percent share of 11 percent compared with 19 percent in the data. For illiquid assets, the model generates substantial concentration

Table OA2: Productivities

Symbol	Value
s_1	0.1812
s_2	0.8962
s_3	1.0000
s_4	1.1159
s_5	5.4425
Owner	-

Table OA3: Transition matrix

		tomorrow					
		s_1	s_2	s_3	s_4	s_5	Owner
today	s_1	0.9054	0.0913	0.0020	0.0000	0.0050	0.0005
	s_2	0.0098	0.8988	0.0858	0.0000	0.0050	0.0005
	s_3	0.0020	0.0865	0.8195	0.0865	0.0050	0.0005
	s_4	0.0000	0.0000	0.0867	0.9078	0.0050	0.0005
	s_5	0.0395	0.0396	0.0395	0.0395	0.8415	0.0005
	Owner	0.0412	0.0412	0.0412	0.0412	0.0412	0.7938

Table OA4: Model fit: aggregate moments

	Data	Model
Capital to output ratio	3.04	3.02
Liquid to illiquid asset ratio	0.11	0.08
Gini net worth	0.82	0.80
Fraction with b $<$ 0	0.14	0.20
Fraction with b = 0 and a $>$ 0	0.20	0.22
Fraction with b = 0 and a = 0	0.08	0.05
Fraction with b = 0	0.28	0.27
Fraction with a = 0	0.10	0.17

Data : SCF 2007, NIPA

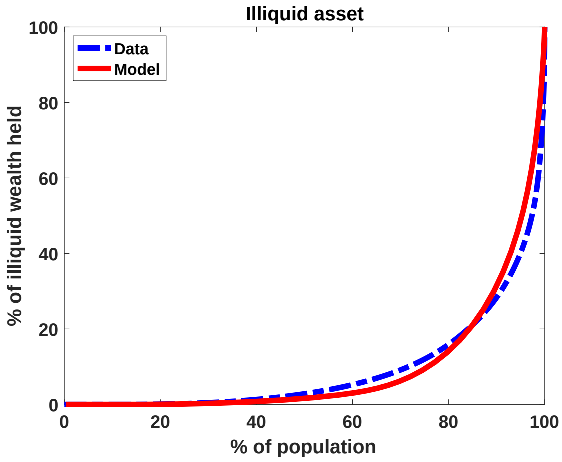
but again falls short at the very top of the distribution: the top 10 percent share is 65 percent in the model versus 76 percent in the data, the top 1 percent share is 15 percent versus 40 percent, and the top 0.1 percent share is 2 percent versus 16 percent, which is a well known challenge in this class of models. At the lower tail, liquid asset shares are negative in both the data and the model, reflecting net debt among low wealth households: the bottom 50 percent share is minus 3 percent in the data and minus 5 percent in the model, and the bottom 25 percent share is minus 4 percent and minus 5 percent. For illiquid assets, bottom shares are small and positive in the data and close to zero in the model. The Gini coefficients for liquid and illiquid assets are also close, at 0.94 and 0.98 for liquid assets in the data and model and 0.83 and 0.81 for illiquid assets. Overall, the model tracks the distribution of household wealth in the data reasonably well while exhibiting the usual difficulty in matching the extreme upper tail of illiquid assets.

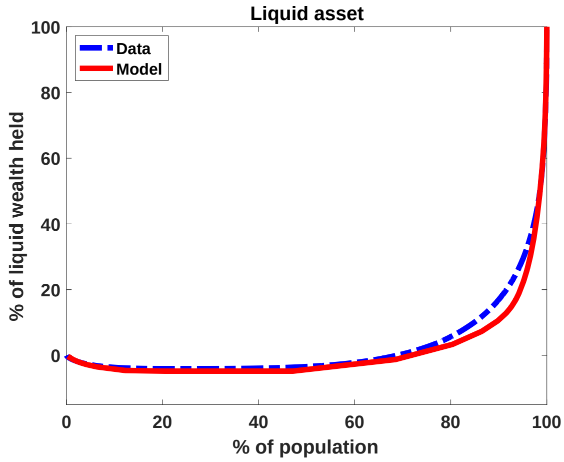
Table OA5: Model fit: wealth concentration by asset type

	Liquid Assets		*Illiquid Assets*	
Moments	Data	Model	Data	Model
Top 0.1 percent share	19	11	16	2
Top 1 percent share	43	43	40	15
Top 10 percent share	81	87	76	65
Bottom 50 percent share	-3	-5	3	2
Bottom 25 percent share	-4	-5	0.4	0.3
Gini Coefficient	0.94	0.98	0.83	0.81

Data : SCF 2007

Figure OA1: Lorenz curves in the data and the model





Notes: The figure shows assets holding inequality in the data and in the model using Lorenz curves. For the definition of liquid and illiquid asset in the data, see the main text.

D Further details on the estimation

D.1 Estimation procedure

During the estimation, I draw for two blocks, a structural parameters block and an expected ELB duration block, in isolation. When making draws for the structural parameters, the expected ELB durations are fixed at their previously accepted values, and vice versa. For the expected ELB duration draws, I first randomly sample the number of quarters to update from a discrete uniform distribution. Then, for the selected quarters, I draw new expected ELB durations from a discrete uniform proposal density and evaluate the likelihood. In this paper, I use a multinomial distribution with eight points adjacent to the existing expected ELB duration. That is, at each draw, I increase or decrease a subset of expected ELB durations by up to four quarters. The acceptance is determined based on the ratio of the likelihoods. For the other block with structural parameters, a standard Metropolis-Hastings algorithm is used. To speed up the estimation process, I use the inversion filter for likelihood evaluation instead of the Kalman filter. If the Kalman filter were used, I would need to continuously update the state transition matrix during the likelihood evaluation, which would be time-consuming given the large size of the equilibrium system.

D.2 Observables and a mapping between the data and the model

For the estimation, I use the following data. The most of the data were collected from FRED or BEA. The data period is from 1992 Q1 to 2018 Q4, except for the central bank's assets, of which data is only available since 2003.

1. Output
   - Model : $\tilde{Y}_t^{\text{obs}} = \log\left(\frac{Y_t}{Y_{t-1}}\right)$
   - Data : Nominal GDP (FRED, GDP), divided by GDP deflator (FRED, GDPDEF) and civilian non-institutionalized population (FRED, CNP16OV), log-transformed, first-differenced and de-meaned.

2. Consumption
   - Model : $\tilde{C}_t^{\text{obs}} = \log\left(\frac{C_t}{C_{t-1}}\right)$
   - Data : The sum of PCE on non-durable goods and services (BEA NIPA Table 2.3.5, item 8 & 13), divided by GDP deflator (FRED, GDPDEF) and civilian non-institutionalized population (FRED, CNP16OV), log-transformed, first-differenced and de-meaned.

3. Investment
   - Model : $\tilde{I}_t^{\text{obs}} = \log\left(\frac{I_t}{I_{t-1}}\right)$
   - Data : The sum of private fixed investment (BEA NIPA Table 5.3.5, all types) and PCE on durable goods (BEA NIPA Table 2.3.5, item 3), divided by GDP deflator (FRED, GDPDEF) and civilian non-institutionalized population (FRED, CNP16OV), log-transformed, first-differenced and de-meaned.

4. Inflation rate
   - Model : $\tilde{\pi}_t^{\text{obs}} = \log\left(\frac{\pi_t}{\pi}\right)$
   - Data : Log difference of GDP Implicit Price Deflator (FRED, GDPDEF) minus 0.5 percentage point.

5. Interest rate
   - Model : $\tilde{i}_t^{\text{obs}} = \log\left(\frac{R_t}{R}\right)$
   - Data : Effective Federal Funds Rate, divided by 400 to express in quarterly units minus logarithm of the model's steady state nominal rate.

6. Real wage
   - Model : $\tilde{w}_t^{\text{obs}} = \log\left(\frac{w_t}{w_{t-1}}\right)$
   - Data : Average hourly earnings of production and non-supervisory employees in total private sector (FRED, AHETPI), divided by GDP deflator (FRED, GDPDEF), log-transformed, first-differenced and de-meaned.

7. Unemployment rate
   - Model : $\tilde{u}_t^{\text{obs}} = \log\left(\frac{u_t}{u}\right)$

- Data : Unemployment as the number of unemployed as a percentage of the labor force (FRED, UNRATE) minus minus 5 percent divided by 100.

8. Lump-sum transfer

- Model : $\tilde{T}_t^{\text{obs}} = \log\left(\frac{T_t^g}{T_{t-1}^g}\right)$
- Data : The sum of government's current transfer payment (BEA NIPA table 3.2, item 26), capital transfer payments (item 22), net of current transfer receipts (item 19), capital transfer receipts (item 42), and unemployment benefit (NIPA underlying table 3.12U, item 7), divided by GDP deflator (FRED, GDPDEF) and civilian non-institutionalized population (FRED, CNP16OV), log-transformed, first-differenced and de-meaned.

9. Profits

- Model : $\tilde{\Pi}_t^{\text{obs}} = \log\left(\frac{\Pi_t}{\Pi_{t-1}}\right)$
- Data : Corporate profits after tax with inventory valuation adjustment and capital consumption adjustment (BEA account code: A551RC), divided by GDP deflator (FRED, GDPDEF), and civilian non-institutionalized population (FRED, CNP16OV), log-transformed, first-differenced and de-meaned.

10. Central bank's assets

- Model : $\tilde{A}_{t+1}^{\text{CB,obs}} = \log\left(\frac{A_{t+1}^{\text{CB}}}{A_{2007}^{\text{CB}}}\right)$
- Data : All Federal Bank's assets (FRED, WALCL), divided by GDP deflator (GDP deflator), civilian non-institutionalized population (CNP16OV), and its end of 2007 level. Log-transformed

D.3 Structural shocks

1. Total factor productivity shock

$$\log Z_t = \rho_z \log Z_{t-1} + \epsilon_{Z,t} \ , \epsilon_{Z,t} \ \sim \ N(0, \sigma^2_{\epsilon_Z,t}) \tag{108}$$

2. Risk premium shock (a shock to banks' discount factor)

$$\Lambda^b_{t,t+1} = \Psi^b_t \Lambda_{t,t+1} \tag{109}$$

$$\log\left(\frac{\Psi^b_t}{\Psi^b}\right) = \rho_b \log\left(\frac{\Psi^b_{t-1}}{\Psi^b}\right) + \epsilon_{b,t} \ , \epsilon_{b,t} \ \sim \ N(0, \sigma^2_{b,t}) \tag{110}$$

3. Price mark-up shock

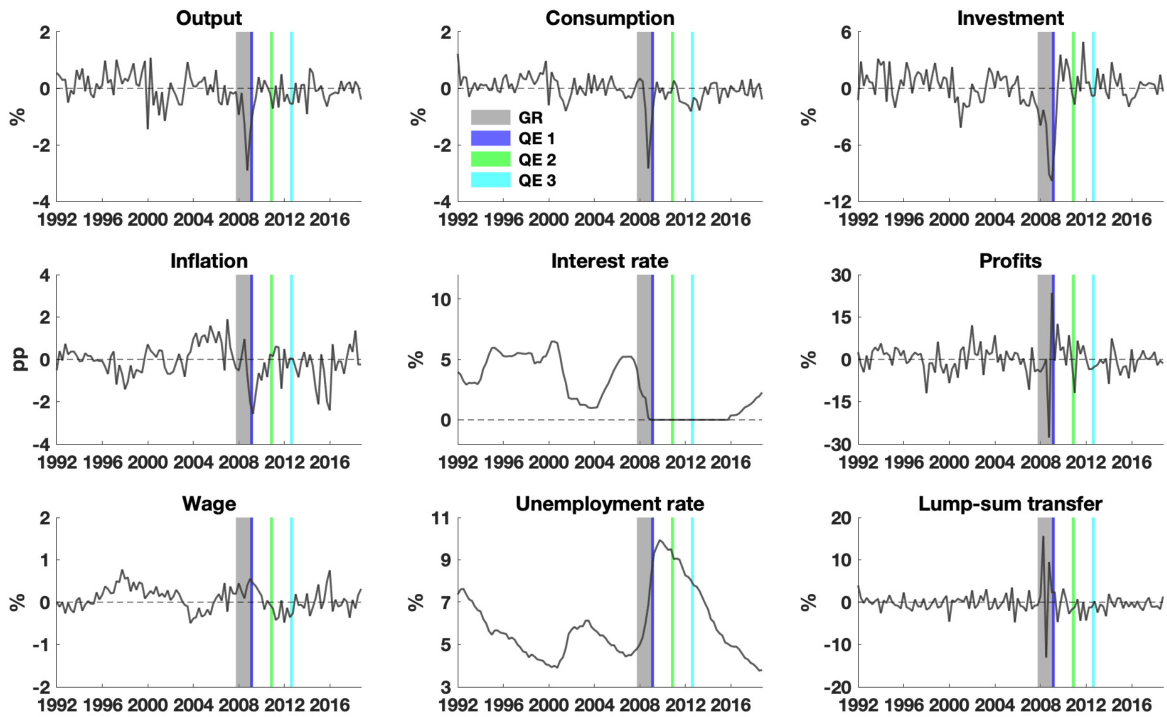
$$\Psi^p_t = \frac{\eta_t}{\eta_t - 1} \tag{111}$$

$$\log(\Psi^p_t) = \rho_p \log(\Psi^p_{t-1}) + \epsilon_{p,t} \ , \ \epsilon_{p,t} \ \sim \ N(0, \sigma^2_p) \tag{112}$$

4. Investment technology shock

$$\log(\Psi^k_t) = \rho_k \log(\Psi^k_{t-1}) + \epsilon_{k,t} \ \ , \sim \ N(0, \sigma^2_k) \tag{113}$$

Figure OA2: Observables



Notes: The figure shows de-meaned quarterly growth rates of output, consumption, investment, real wages, lump-sum transfers, and corporate profits. The inflation rate is shown as the percentage point deviation from its target of 2%. The nominal interest rate (annualized) and unemployment rate are shown as levels (percentage points). Green, blue, green, and sky blue areas depict the Great Recession period, and the periods in which QE 1, 2, and 3 are implemented, respectively.

5. Liquidity preference shock

$$\log(\Psi_t^l) = \rho_l \log(\Psi_{t-1}^l) + \epsilon_{l,t} \ , \ \epsilon_{l,t} \ \sim \ N(0, \sigma_l^2) \tag{114}$$

6. Wage shock

$$\frac{w_t}{w} = \left(\epsilon_{w,t}\frac{r_t^l}{r^l}\right)^{\vartheta_w(1-\rho_w)} \left\{\frac{w_{t-1}}{w} \times \left(\frac{\pi}{\pi_t}\right)\right\}^{\rho_w} \ , \ \ 0 < \rho_w < 1 \ , \ \vartheta_w > 0 \tag{115}$$

$$\tag{116}$$

7. Lump-sum transfer shock

$$T_t^g = \left(1 - \frac{1}{\Psi_t^g}\right) Y \tag{117}$$

$$\log\left(\frac{\Psi_t^g}{\Psi^g}\right) = \rho_g \log\left(\frac{\Psi_{t-1}^g}{\Psi^g}\right) + \epsilon_{g,t} \ , \epsilon_{g,t} \ \sim \ N(0, \sigma_g^2) \tag{118}$$

8. Monetary policy shock

$$1+\hat{i}_{t+1} = (1+\hat{i})\left(\frac{1+\hat{i}_t}{1+\hat{i}}\right)^{\rho_R}\left[\left(\frac{\pi_t}{\pi}\right)^{\phi_\pi}\left\{\exp(u_t-u)\right\}^{\phi_u}\right]^{1-\rho_R}\exp(\epsilon_{R,t}) \;\; , \; \epsilon_{R,t} \;\sim\; N(0,\sigma_R^2) \tag{119}$$

$$i_{t+1} = \max\{0, \hat{i}_{t+1}\} \tag{120}$$

9. Fixed cost shock

$$\Psi_t^F = \rho_F \Psi_{t-1}^F + (1-\rho_F)\Psi^F + \epsilon_{F,t} \;\; , \; \epsilon_{F,t} \;\sim\; N(0,\sigma_F^2) \tag{121}$$

$$\tag{122}$$

10. QE shock

$$A_{t+1}^{\mathrm{CB}} = \Psi_{\mathrm{QE},t} Y \; , \;\; \log(\Psi_t^{\mathrm{QE}}) = \rho_{\mathrm{QE}} \log(\Psi_{t-1}^{\mathrm{QE}}) + \epsilon_{\mathrm{QE},t} \;\; \epsilon_{\mathrm{QE},t} \;\sim\; N(0,\sigma_{\mathrm{QE}}^2) \tag{123}$$

D.4 Estimation results

Table OA6: Prior and posterior distributions of structural parameters

Symbol	Description	Prior			Posterior		
		Prior Density	Mean	Std	Mode	10%	90%
	Frictions						
κ	Slope of Phillips curve	Gamma	0.10	0.02	0.0525	0.0340	0.0765
ι_p	Price indexation	Gamma	0.50	0.15	0.1219	0.0670	0.2069
ρ_w	Wage autocorrelation	Beta	0.50	0.20	0.7982	0.7065	0.8654
ι_w	Wage indexation	Beta	0.50	0.15	0.1835	0.1132	0.2639
ϕ	Capital adjustment cost	Normal	30.00	5.00	50.017	49.193	51.184
ι	Vacancy posting cost	Gamma	0.05	0.02	0.0317	0.0189	0.0495
	Government policy						
ρ_B	Bond issuance rule	Beta	0.50	0.20	0.5058	0.3998	0.6047
ρ_g	Lump sum transfer shock AR	Beta	0.50	0.20	0.9986	0.9967	0.9995
σ_G	Lump sum transfer shock std dev	Inverse-Gamma	0.10	2.00	0.1991	0.1815	0.2172
ρ_R	Interest rate smoothing	Beta	0.50	0.20	0.7927	0.7567	0.8271
σ_R	Interest rate shock std dev	Inverse-Gamma	0.10	2.00	0.1693	0.1481	0.1953
ϕ_π	Taylor rule inflation gap response	Normal	1.70	0.30	1.3101	1.1551	1.5231
ϕ_y	Taylor rule unemployment gap response	Gamma	0.10	0.05	0.3748	0.3307	0.4276
	Structural Shocks						
ρ_l	Liquidity preference shock AR	Beta	0.50	0.20	0.9997	0.9993	0.9999
σ_l	Liquidity preference shock std dev	Inverse-Gamma	0.10	2.00	0.0483	0.0431	0.0551
ρ_z	TFP shock AR	Beta	0.50	0.20	0.9952	0.9933	0.9965
σ_z	TFP shock std dev	Inverse-Gamma	0.10	2.00	0.5782	0.5320	0.6333
ρ_p	Price markup shock AR	Beta	0.50	0.20	0.9608	0.9457	0.9720
σ_p	Price markup shock std dev	Inverse-Gamma	0.10	2.00	1.6344	1.3283	2.1629
ρ_k	Investment shock AR	Beta	0.50	0.20	0.9784	0.9645	0.9900
σ_k	Investment shock std dev	Inverse-Gamma	0.10	2.00	0.0714	0.0658	0.0778
ρ_b	Risk premium shock AR	Beta	0.50	0.20	0.9887	0.9815	0.9941
σ_b	Risk premium shock std dev	Inverse-Gamma	0.10	2.00	0.1601	0.1436	0.1796
ρ_Ξ	Fixed cost shock AR	Beta	0.50	0.20	0.9505	0.9355	0.9643
σ_Ξ	Fixed cost shock std dev	Inverse-Gamma	0.10	2.00	0.9203	0.8380	1.0163
σ_w	Wage shock std dev	Inverse-Gamma	0.10	2.00	0.8324	0.5000	1.3296

Notes The values for the standard deviations and the measurement error are multiplied by 100.

Table OA7: Prior and posterior distributions of expected ELB durations

	Prior			Posterior		
	Mode	10%	90%	Mode	10%	90%
2009 Q1	5	2	8	4	3	5
2009 Q2	5	2	8	6	3	8
2009 Q3	5	2	8	5	3	7
2009 Q4	5	2	8	5	3	7
2010 Q1	5	2	8	6	3	8
2010 Q2	5	2	8	6	3	8
2010 Q3	5	2	8	6	3	8
2010 Q4	5	2	8	4	3	6
2011 Q1	4	2	7	7	4	8
2011 Q2	4	2	6	5	3	7
2011 Q3	8	5	11	6	4	8
2011 Q4	8	5	11	7	6	9
2012 Q1	9	5	12	6	4	8
2012 Q2	10	5	14	7	6	9
2012 Q3	10	5	13	7	5	9
2012 Q4	11	7	14	7	6	9
2013 Q1	9	5	13	8	5	9
2013 Q2	7	3	12	7	5	8
2013 Q3	7	4	12	6	5	8
2013 Q4	8	4	11	6	4	8
2014 Q1	6	3	10	7	5	8
2014 Q2	6	3	9	5	4	6
2014 Q3	3	1	5	3	2	5
2014 Q4	2	1	4	3	2	4
2015 Q1	1	1	3	2	1	4
2015 Q2	1	1	3	2	2	3
2015 Q3	1	1	2	2	2	3
2015 Q4	1	1	1	3	1	3

Notes: The unit is one quarter.

Figure OA3: Prior and posterior distributions

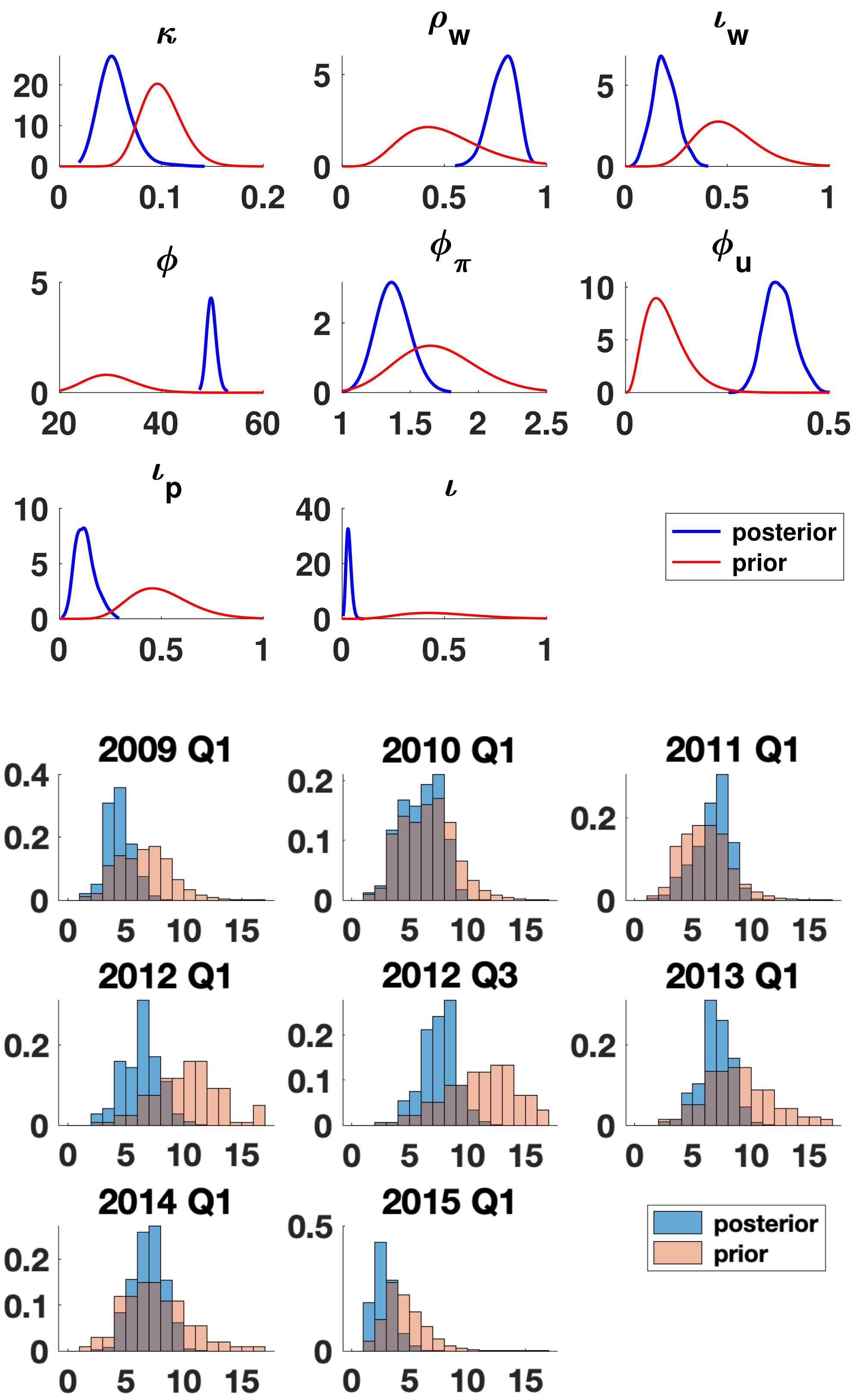
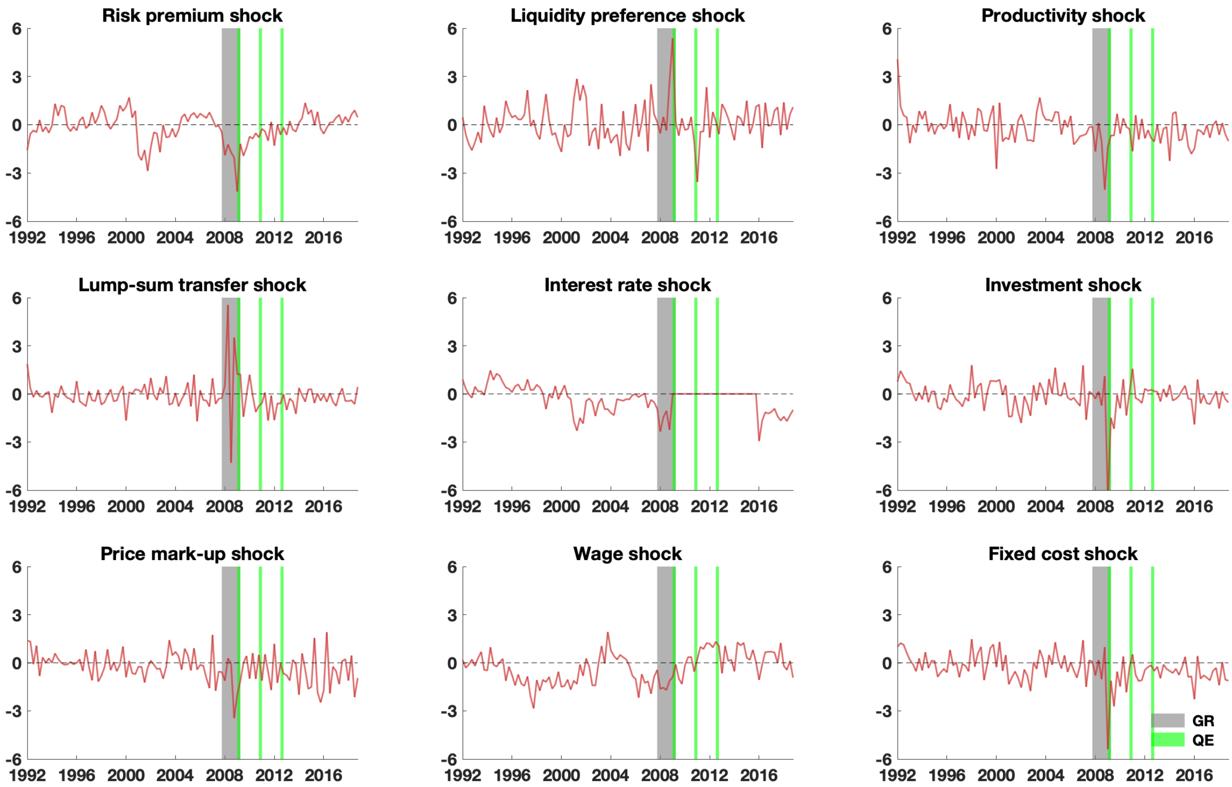


Figure OA4: Filtered shock series



Notes: The figure shows the time series of the filtered shocks during the sample periods as a ratio to its standard deviation. The shaded gray area represents the periods of the Great Recession. The transparent green bars represent the quarters in which QE 1, 2, and 3 are announced or implemented.

E Model Dynamics

A countercyclical response of profits to demand shocks is a common feature of New Keynesian models. Since the factor prices are relatively flexible while the price is assumed to be rigid, a markup of the price over marginal cost is countercyclical in New Keynesian models when demand shocks, such as monetary policy and government spending shocks, occur. Consequently, profits fall after an increase in aggregate demand.[68] Though this feature is not consistent with the existing empirical evidence, the literature has not paid much attention since, in representative agent New Keynesian models, the response of profits did not seem to matter for the model's implications on the aggregate dynamics of the economy.

However, recently, the literature started to challenge this feature of New Keynesian models. Broer et al. (2019) pointed out that a fall in profits is a key amplification channel through which an expansionary monetary policy shock leads to a strong output response. Specifically, a fall in profits induces households to increase their labor supply by generating a negative wealth effect. Alves et al. (2019) also demonstrate that the way profits are distributed affects the aggregate consequences of monetary policy shocks. In particular, when a larger share of profits is allocated to liquid assets, monetary policy shocks have greater amplification in

[68] A lower markup does not necessarily imply lower profits since, in principle, the response of the quantity sold can be large enough to offset the negative effect of markups on profits. However, in standard New Keynesian models, the effect of markup dominates as the quantity response is relatively moderate. As a result, profits decrease despite an increase in demand.

their model. These recent findings in the literature show the importance of profit responses in determining the aggregate dynamics of New Keynesian models.

In this paper, I emphasize the importance of profit dynamics for the distributional consequences of monetary policy. Since profits constitute a substantial portion of wealthy households' income, the way that profits respond to monetary policy determines their welfare gains/losses from the policy. In short, when profits respond strongly procyclically to monetary policy as in the data, wealthy households can enjoy a considerable amount of welfare gains from an expansionary monetary policy shock.

In the following subsections, I show the model's impulse responses, including a procyclical response of profits, to an expansionary monetary policy shock, and discuss how the model generates such a response.

E.1 Procyclical profits

Figure OA5 shows the responses of the model's aggregate variables to an expansionary monetary policy shock at the posterior mode of parameter values. The figure shows that, when a negative interest rate shock occurs, profits substantially increase in the model. This feature of the model contrasts starkly with existing New Keynesian models in which profits exhibit strong countercyclicality in response to monetary policy shocks. More importantly, such responses are consistent with empirical evidence; a monetary SVAR model presented in the appendix generates similar profits, wage, and unemployment rate responses in terms of the direction and the relative magnitudes.[69]

How does the model generate a procyclical profit response to changes in demand while existing models could not? First, wage rigidity and labor market frictions dampen the response of the real marginal cost. When the aggregate demand increases, firms expand their production by hiring more labor and capital services. In a standard New Keynesian model, such an increase in factor demand leads to an increase in the real marginal cost, or equivalently, a fall in markups. Thus, profits fall.[70] However, in the model, the real wage does not respond much because of wage rigidity. If labor supply adjusts only through intensive margin, little changes in the real wage imply little changes in the labor supply. Then, to increase output, firms need to utilize the capital more intensively, which results in a substantial increase in the capital rental rate or the variable depreciation. An extensive margin adjustment of labor supply via frictional labor markets allows firms to increase labor inputs without increasing the real wage and the capital rental rate much. Consequently, the real marginal cost does not respond strongly to an increase in demands in the model. [71]

Besides, based on a recent finding of Anderson et al. (2018), I assume that the fixed cost accounts for a significant proportion of the total production cost.[72] The presence of the fixed cost helps the model generate a procyclical profit response as well. What matters for firms'

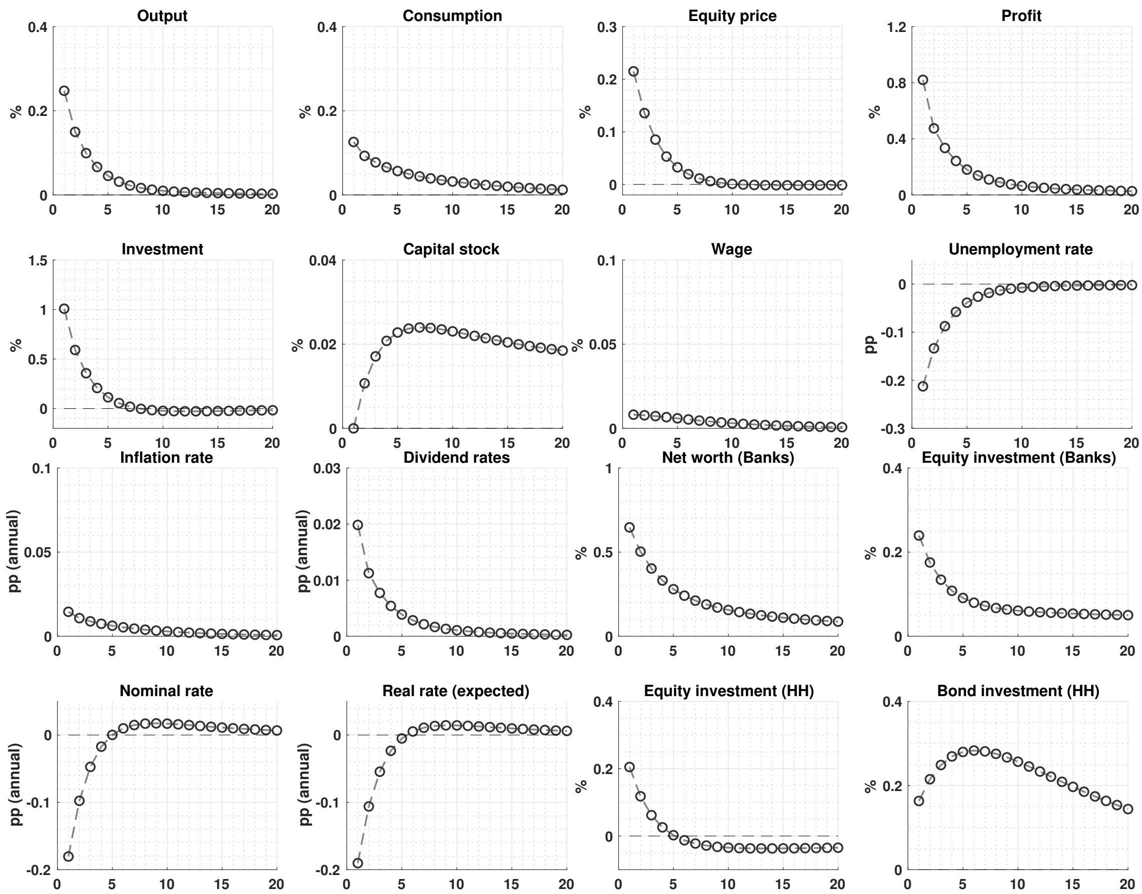
[69] A noticeable feature of the model, relative to the SVAR model, is the lack of the hump-shaped responses, which is a common feature of most of the existing HANK models. Since models do not feature internal delaying mechanisms, such as habits, the responses are immediate when there is an exogenous shock. Recently, Auclert et al. (2020b) develop a HANK model that incorporates sticky expectations and generate delayed responses of the aggregate variables to exogenous shocks in their model.

[70] In a standard New Keynesian model, the degree of price rigidity should be high for a monetary policy shock to have real effects. A high degree of price rigidity implies, in the absence of the factor price rigidity, a strong countercylicality of profits or markups, the latter of which has been often challenged in the literature.

[71] Note that the marginal cost for intermediate good firms is determined by the capital and labor rental rate, and I do not impose any rigidity on the labor rental rate. However, wage rigidity and labor supply via labor agencies effectively increase the elasticity of labor supply with respect to changes in the labor rental rate. Thus, to achieve the same amount of an increase in labor input, a smaller magnitude of the rental rate increase is required.

[72] Anderson et al. (2018) show that, using confidential retail sector transactions data, gross margin, which can be interpreted as markups in the model, is acyclical or mildly procyclical while net operating profits are highly

Figure OA5: Impulse responses to an expansionary monetary policy shock



Notes: The figure shows the model's impulse responses to a negative 25 basis points (annualized) interest rate shock. All variables are shown as the percentage deviations from their respective steady state values except for the nominal rate, the inflation rate, the dividend rates and the unemployment rate. The nominal rate, the inflation rate, and the dividend rates are expressed in terms of the annualized percentage point difference from the steady state values. The unemployment rate is shown as the percentage point difference from the steady state unemployment rate.

profit is not the marginal cost per se but the average production cost. When the fixed cost accounts for a substantial proportion of the total cost, the average cost can fall even though the marginal cost increases. Moreover, as the production sector is decentralized in the model, the sector-wide cost is lower than the cost of intermediate good firms.[73] Thus, as Figure OA6 shows, while the marginal cost of intermediate good firms mildly increases, the average cost of the entire non-financial sector decreases, which results in a substantial increase in non-financial firms' profits.

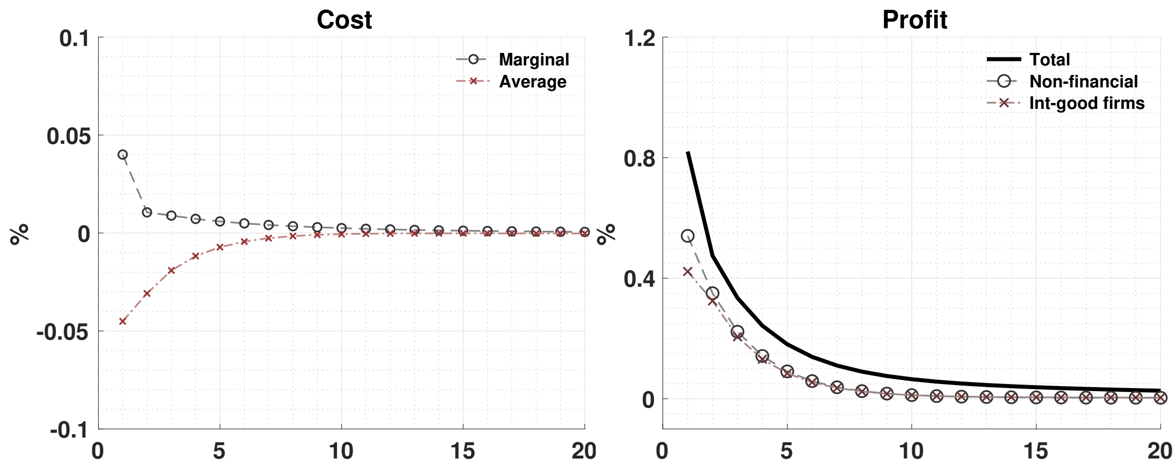
Finally, the presence of banks also helps the model generate a substantial increase in profits. First, an increase in banks' net-worth contributes to higher profits.[74] When the

procyclical. They interpret the latter result as suggesting the presence of fixed costs.

[73]Ignoring miscellaneous adjustment costs, the intermediate good firms' total cost can be expressed by $\Gamma_t Y_t + \Xi$, where Γ_t is the real marginal cost and Ξ is the fixed cost. In contrast, the total cost of the non-financial sector as a whole is $\delta(v_t)K_t + w_t L_t + \iota V_t + \Xi$. Because of accelerated depreciation and the wage rigidity, the latter is smaller than the former during an expansion unless ι is too high.

[74]The empirical evidence on the effects of monetary policy on banks' profitability is mixed and not conclusive. Borio et al. (2017) concluded that low interest rates and flat term structure erodes banks' profitability mainly through their negative impacts on banks' net interest income. However, they solely focused on the trend changes in the interest rate structure and, importantly, did not take into account any effects of monetary policy on the aggregate economy in their analysis. A more recent work by Altavilla et al. (2018) showed that an expansionary monetary policy shock does not reduce banks' profitability once they control for the endogeneity of the policy measures. Finally, Zimmerman (2019) showed, using the panel data of more than 100 countries for

Figure OA6: Responses of different types of costs and profits



Notes: The figure shows the response of different kinds of costs and profits to an expansionary monetary policy shock. The left panel shows the response of the intermediate good firms' marginal cost, which is shown with the gray dotted line with circles, and the average cost of the non-financial sector, which is shown with the red dotted line with crosses. The right panel shows the response of the total profits, total non-financial sector's profits, and the intermediate good firms' profits. The black solid line shows the response of the aggregate profits while the gray dotted line with circles and the red dotted line with crosses show the non-financial sector profits and the intermediate good firms' profits, respectively.

Table OA8: Posterior mode under the rigid and flexible wage assumption

	κ	ι_p	ρ_w	ι_w	ι	ρ_R	ϕ_π	ϕ_u
Rigid wage	0.0525	0.1219	0.7982	0.1835	0.0317	0.7927	1.3101	0.3748
Flexible wage	0.1114	0.0564	0	0	0.0929	0.8405	2.5354	0.1590

interest rate falls and investment increases, the equity price increases, and thus the gross return on banks' net-worth substantially increases on impact. The effects of an increased net worth propagate through a financial accelerator channel and persist for a long time, leading to higher aggregate profits.[75] In the process, banks also lead to strong investment responses. Thus, even though consumption response is relatively small due to a weak redistribution and the wage rigidity, the overall demand of goods can increase significantly because of banks' investment demand.

E.2 Comparison with a model with the flexible wage

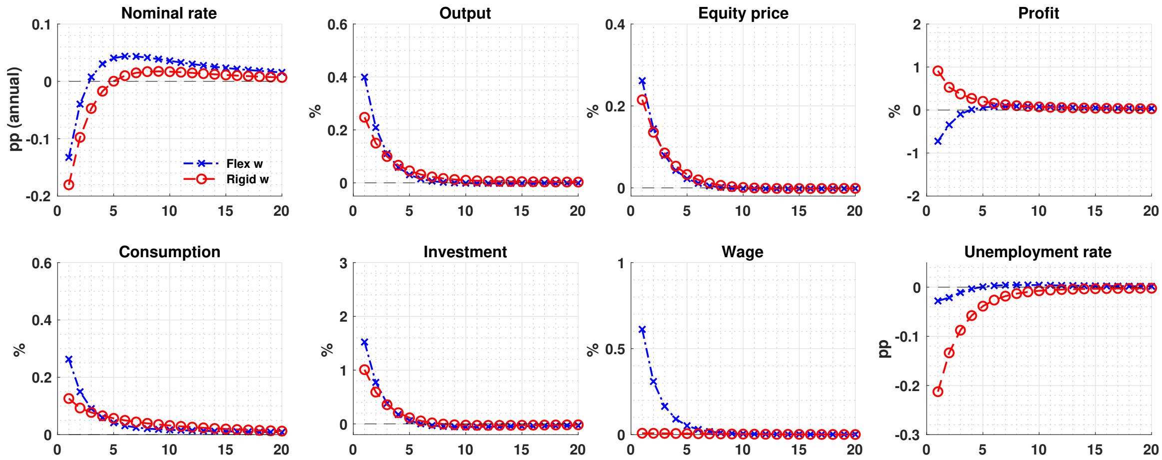
Figure OA7 shows the impulse responses of variables in the baseline model and the model with the flexible wage. For a fair comparison, I re-estimate the model by assuming that the wage is flexible, i.e., $\rho_w = 0$. Table OA8 shows the values of key parameters at the posterior mode.

Two things are noticeable in the figure. First, depending on the assumption of wage rigidity, the response of profits is entirely different. When the wage is assumed to be flexible, profits exhibit strong countercyclicality in response to monetary policy shocks. While profits fall substantially, the real wage soars after an increase in the aggregate demand. Due to a strong real wage response, the unemployment rate changes little in the model with the

more than 100 years, the importance of loan losses and credit growth for bank profits and shows that a monetary policy tightening leads to a fall in banks' profits in contrast with the previous findings.

[75]Due to the incentive problem characterized by Gertler and Karadi (2011), the total amount of deposits that a bank can take is limited to a certain fraction of the bank's net worth. Thus, an increase in the bank's net worth allows the bank to purchase more assets by taking more deposits, which leads to a further increase in its net worth.

Figure OA7: Wage rigidity and the IRFs to an expansionary monetary policy shock



Notes: The figure shows the impulse responses of variables in models with different assumptions on the wage rigidity. The blue dotted lines with crosses show IRFs from the model with flexible wage ($\rho_w = 0$), and the red dotted lines with circles show IRFs from the baseline model with wage rigidity. All parameters take on values at their respective posterior mode in each model. The unit for the nominal interest rate and the unemployment rate is percentage point. The unit for all other variables is the percentage deviation from the corresponding steady state value.

flexible wage. However, as I show in the appendix, these responses are not consistent with the empirical evidence.

The other result that is noticeable in the comparison is that, when the real wage is flexible, an expansionary monetary policy shock has stronger initial stimulus effects compared to a model with wage rigidity. For instance, an annualized 25 bp falls in the policy rate leads to 0.4% increase in output on impact when the wage is flexible. In contrast, the corresponding magnitude of the impact is only 0.25% in the baseline model. Given that the parameter values at the mode imply much smaller real effects of monetary policy shocks, i.e., a steeper Philips curve and stronger responsiveness of the policy rate to the inflation gap, the magnitude of the initial response under the flexible wage is substantial. Two channels are working behind this result. The first one is redistribution. When profits are strongly countercyclical, an expansionary monetary policy shock leads to a stronger redistribution from wealthy to working-class households. Since the latter has a higher marginal propensity to consume than the former, the aggregate consumption response from the monetary policy shock is larger when the wage is flexible. The other one is an amplification that arises from the complementarity between consumption and labor in GHH preference. When the real wage goes up, households supply more labor under the GHH preference. Then, they also demand more consumption since consumption and labor are complementary. Such an increase in demand for goods further stimulates the production and increases the real wage, creating a substantial amount of amplification. Auclert et al. (2020a) argue that, based on earlier findings of Monacelli and Perotti (2008) and Bilbiie (2009), such an amplification due to the complementary between consumption and labor results in unrealistically high fiscal multipliers in New Keynesian models with the flexible wage.

To recapitulate, the model with the flexible wage generates impulse responses of key aggregate variables that are not consistent with the data in terms of both direction and magnitude. Such results support the modeling approach adopted in this paper, which emphasize the role of wage rigidity and frictional labor markets.[76]

[76]The role of the wage rigidity recently regained attention in the literature. Broer et al. (2019) advocate focusing on the wage stickiness rather than the price stickiness because of its implications on the redistribution and the amplification in New Keynesian models. Nekarda and Ramey (2020) also do so based on their findings on the cyclicality of markups.

F Structural VAR analysis

In this section, I provide an empirical evidence on the effects of monetary policy on real wage, unemployment rates, and profits, which motivated a new HANK model that I develop in this paper. Specifically, I conduct a structural vector autoregression (VAR) analysis. The specification of the SVAR model is based on a standard monetary VAR model that apperar in Christiano et al. (1999) and Christiano et al. (2005). Specifically, I augment a 7 variable VAR model in Christiano et al. (1999) with the variables of interest in this paper, i.e., real wage, unemployment rates, and profits. In addition, to have a better understanding of the fiscal responses, I include the lump-sum transfer variable in the VAR model as well.

As is standard, it is assumed that the policy instrument, i.e., the Fed Funds rate, denoted by FF_t, is determined as follows.

$$\mathrm{FF}_t = f(\Omega_t) + \epsilon_{r,t} \tag{124}$$

where f is the feedback rule, Ω_t is the information set available to the central bank in period t, and $\epsilon_{r,t}$ is an exogenous shock to the policy decision. Let $\mathcal{Y}_t$ denote the vector of the variables included in the VAR model.

$$\mathcal{Y}_t = \begin{bmatrix} \log(\text{Output}_t) \\ \log(\text{Price index}_t) \\ \log(\text{Commodity price index}_t) \\ \log(\text{Real wage}_t) \\ \text{Unemployment rate}_t \\ \log(\text{Profits}_t) \\ \log(\text{Lump-sum transfer}_t) \\ \mathrm{FF}_t \\ \log(\text{Total reserves}_t) \\ \log(\text{Non-borrowed reserves}_t) \\ \log(\text{M2}_t) \end{bmatrix} \tag{125}$$

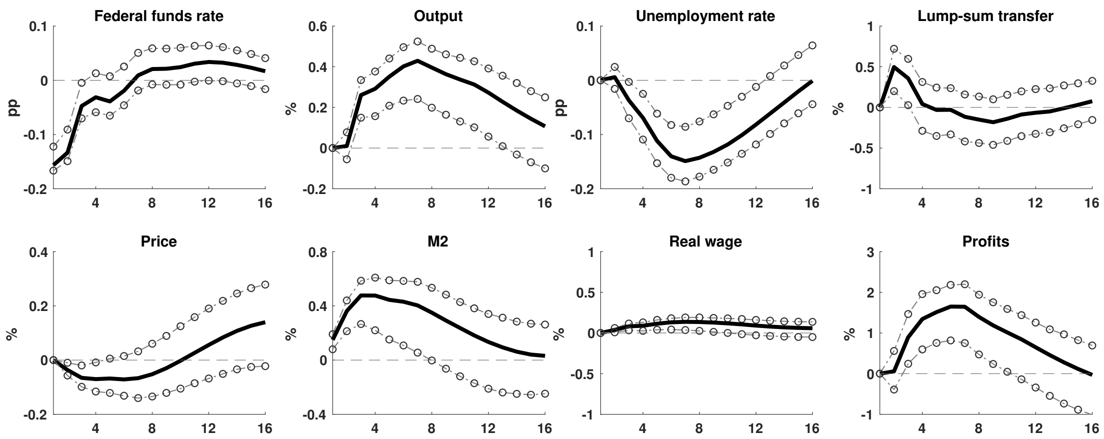
The information set available to the monetary authority includes the data on output, price index, commodity price, index, real wage, unemployment rate, profits, and lump-sum transfer. As in Christiano et al. (1999), I assume that the innovation $\epsilon_{r,t}$ is orthogonal to all variables in the central bank's information sect. Thus, the monetary policy shock is identified using a standard recursive identification strategy.

For the data, I use the same data that I used for the estimation of my model. The exceptions are commodity price index, total reserve, non-borrowed reserve, and M2, which are not included in the set of observaables for the estimation. For the commodity price index, I use the World Bank non-energy commodity price index, smoothing the quarterly change by taking a three quarter average.[77] For the number of lags, I use 4 lags, and the data period is from 1960 to 2007. For the robustness check, I also used 1) average hourly earnings of production and non-supervisory workers, and 2) profits before tax without investment valuation and capital consumption adjustment. Also, I compute impulse responses, using a short sample periods, i.e., from 1979 Q1 to 2007 Q4. Across different specifications, data, and sample periods, the results are similar.

Figure OA8 shows the impulse responses of variables to a 11 basis point expansionary monetary policy shock. As shown in the figure, in response to an expansionary monetary

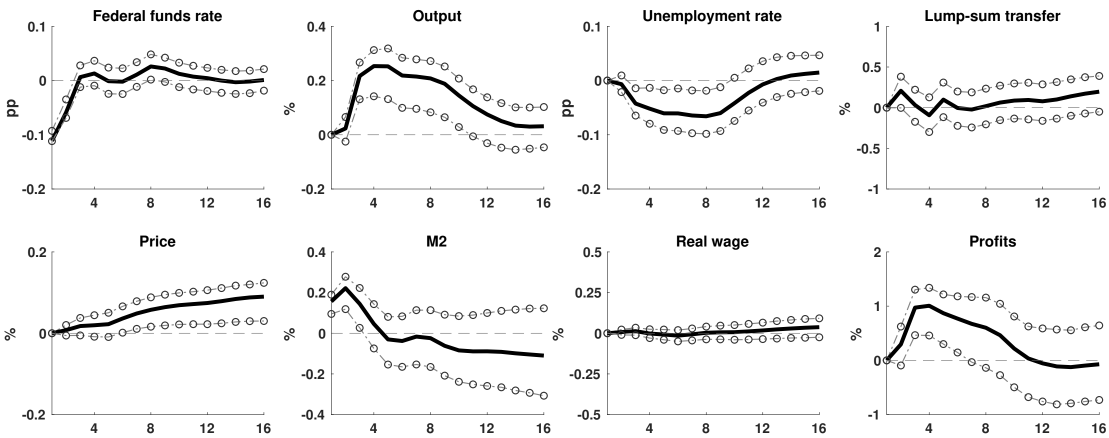
[77] The commodity price index is included to alleviate the 'price puzzle' phenomenon.

Figure OA8: Impulse responses to a shock to FFR: 1960 Q1 to 2007 Q4



Notes: The figure shows the impulse responses of variables to a negative one standard deviation fall in the Federal Funds rate in a SVAR model. The Federal funds rate and the unemployment rate are shown as the percentage point difference from the pre-shock levels. All other variables are shown as the percentage deviation from the pre-shock levels. The dotted lines with circles show 90% boot-strapped confidence intervals with 5,000 runs for each impulse response.

Figure OA9: Impulse responses to a shock to FFR: 1979 Q1 to 2007 Q4



Notes: The figure shows the impulse responses of variables to a negative one standard deviation fall in the Federal Funds rate in a SVAR model. The Federal funds rate and the unemployment rate are shown as the percentage point difference from the pre-shock levels. All other variables are shown as the percentage deviation from the pre-shock levels. The dotted lines with circles show 90% boot-strapped confidence intervals with 5,000 runs for each impulse response.

policy shock, the unemployment rate decreases substantially while the real wage responds little. The real wage responses are barely statistically significant. In contrast, profits rises significantly. The lump-sum transfer responds procyclically for the first few periods after the shock, but the responses are mostly statistically insignificant. The corresponding variables in the model exhibit similar dynamics except for the lack of hump-shaped responses, which is a common limitation of the most of existing HANK models in the literature.[78] Most of variables in the SVAR model peaks between 4th and 8th quarters after the shock. In contrast, in the model, responses are immediate.

[78]The only exception in the current literature is the model of Auclert et al. (2020b). They develops a HANK model with sticky expectations and generates hump-shaped responses of aggregate variables in a full-fledged HANK model.

G QE as a monetary policy rule

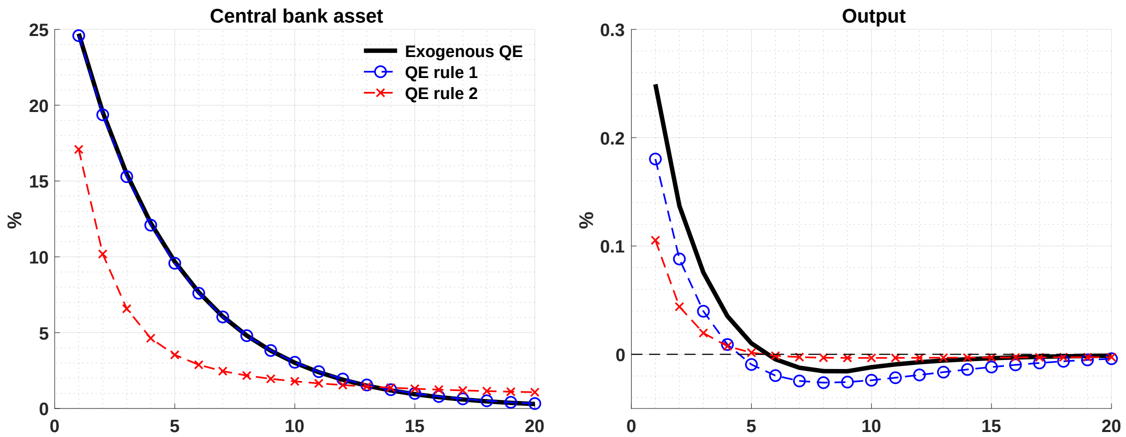
In this section, I consider an alternative specification in which QE is the main monetary policy instrument. QE follows a Taylor-type feedback rule in normal times, while the nominal policy rate is kept at its steady-state level:

$$A^{CB}_{t+1} = \Psi^{QE}_t A^{CB}, \qquad B^{CB}_{t+1} = q_t A^{CB}_{t+1}, \tag{126}$$

$$\Psi^{QE}_t = 1 + \rho_{QE}\left(\Psi^{QE}_{t-1} - 1\right) + (1-\rho_{QE})\left\{\phi^{QE}_{\pi}\log(\pi_t/\pi) + \phi^{QE}_u(u_t - u)\right\} + \epsilon^{QE}_t. \tag{127}$$

For comparability with the Taylor rule on the interest rate, I set $\rho_{QE} = \rho_R$. In the baseline specification (QE rule 1), $\phi^{QE}_{\pi} = -10\phi_{\pi}$ and $\phi^{QE}_u = -10\phi_u$. I also consider a stronger feedback case (QE rule 2) with $\phi^{QE}_{\pi} = -1000\phi_{\pi}$ and $\phi^{QE}_u = -1000\phi_u$, keeping the persistence ρ_{QE} equal to the baseline value.

Figure OA10: Comparison of QE policies: exogenous vs rules



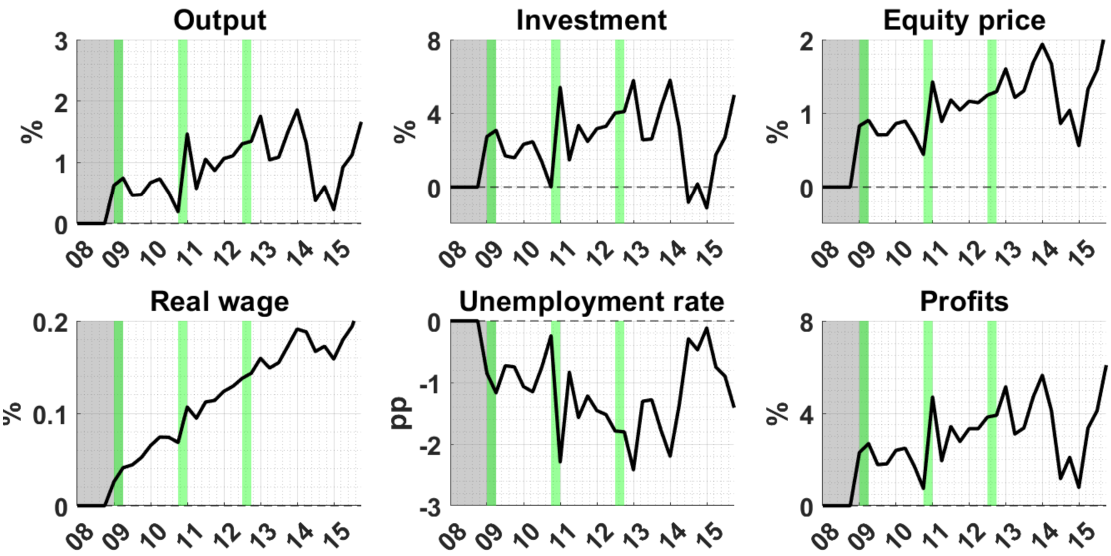
Notes: The figure compares impulse responses to a QE shock equivalent to 1.2 percent of steady state output under an exogenous QE process and under two feedback rules for QE. The left panel shows the path of central bank assets and the right panel shows the associated output responses. The black solid line corresponds to the exogenous QE process, the blue line with circles to QE rule 1, and the red line with crosses to QE rule 2. The persistence of the shock is the same across specifications.

As shown in Figure OA10, under a feedback rule with a modest degree of state dependence (QE rule 1) the path of the central bank balance sheet is very similar to that under the baseline exogenous QE process. The output response, however, is smaller on impact and remains below the baseline path over the horizon, indicating that the feedback rule dampens the expansionary effects of QE. Under the stronger feedback rule (QE rule 2), the impact response of output is smallest, while the subsequent reversal is more muted. Overall, the impulse responses under the exogenous process and the two feedback rules are qualitatively similar, with the main differences arising in the magnitude of the effects unless the feedback is very strong.[79]

Figure OA11 shows the aggregate effects of UMP when QE is implemented as a rule during the ELB episode. When the policy rate is constrained by the lower bound, the central bank conducts QE according to the feedback rules described above. With a modest degree of feedback (QE rule 1), the model implied aggregate effects of UMP are essentially identical to those in the baseline counterfactual experiments with an exogenous QE process. The

[79]Under QE rule 2, a 25 basis point fall in the annualized inflation rate from target leads the central bank to expand its balance sheet by almost 1 percent of steady state output.

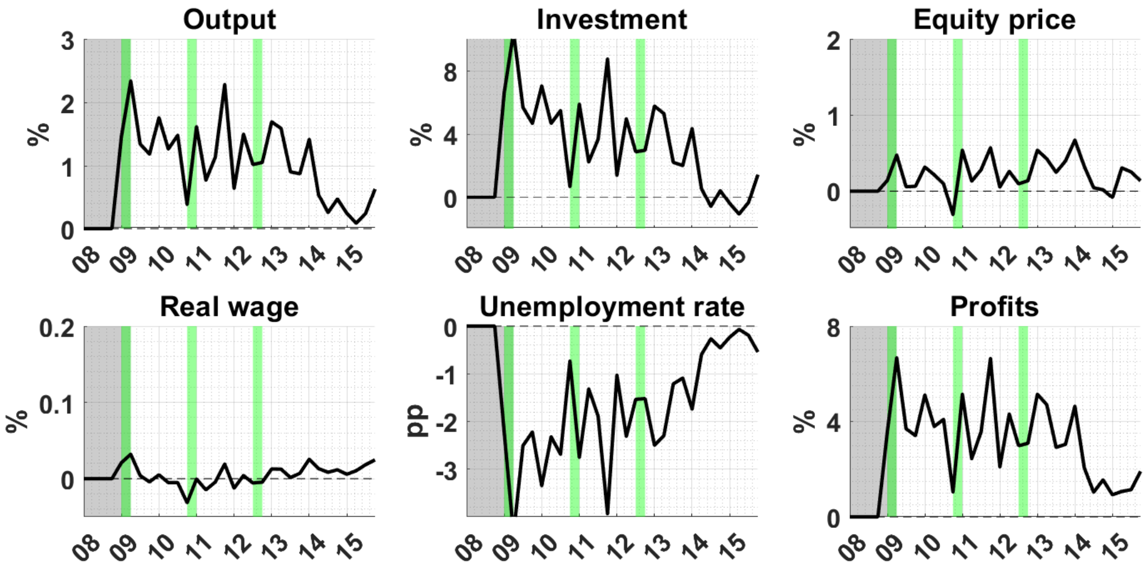
Figure OA11: Aggregate effects of UMP with QE rule 1



Notes: Figure shows the aggregate effects of UMP when QE follows a rule with $\rho_{\mathrm{QE}} = \rho_R$, $\phi_\pi^{\mathrm{QE}} = -10 \times \phi_\pi$, and $\phi_u^{\mathrm{QE}} = -10 \times \phi_u$. For level variables X, each panel reports $100 \times (X^{\mathrm{UMP}}/X^{\mathrm{NoUMP}} - 1)$ in percent. For the unemployment rate, panels report percentage point differences $u^{\mathrm{UMP}} - u^{\mathrm{NoUMP}}$.

filtered QE shocks differ in size across specifications, but the resulting paths of the main aggregates line up almost exactly. In contrast, when QE responds strongly to inflation and unemployment gaps (QE rule 2), the aggregate dynamics differ more visibly: QE delivers larger and more persistent effects on real activity while movements in real wages and asset prices are more muted. Thus modest feedback has little effect on the aggregate impact of UMP, whereas very strong feedback alters its transmission pattern.

Figure OA12: Aggregate effects of UMP with QE rule 2



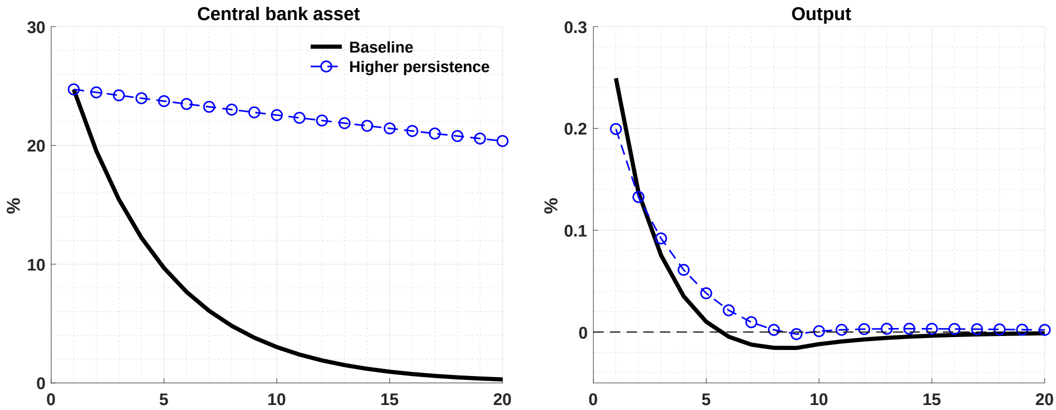
Notes: Figure shows the aggregate effects of UMP when QE follows a rule with $\rho_{\mathrm{QE}} = \rho_R$, $\phi_\pi^{\mathrm{QE}} = -1000 \times \phi_\pi$, and $\phi_u^{\mathrm{QE}} = -1000 \times \phi_u$. For level variables X, each panel reports $100 \times (X^{\mathrm{UMP}}/X^{\mathrm{NoUMP}} - 1)$ in percent. For the unemployment rate, panels report percentage point differences $u^{\mathrm{UMP}} - u^{\mathrm{NoUMP}}$.

H Additional results for monetary policy shocks

H.1 QE process with different degrees of persistence

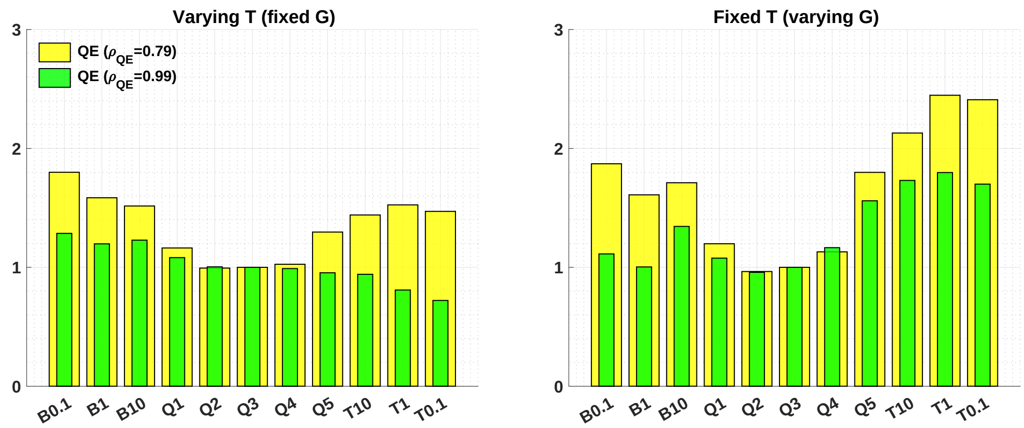
Figure OA14 compares the aggregate effects of QE for different degrees of persistence. When the QE process is much more persistent, the impact effect is somewhat smaller, reflecting a greater extent of crowding out of private investment. As the central bank then maintains its balance sheet at a higher level for longer, there is little subsequent contraction and the impulse responses become closer to those under conventional monetary policy. In the model, this indicates that for QE to generate aggregate dynamics similar to conventional policy, it needs to be substantially more persistent than the conventional policy rule.

Figure OA13: Comparison of QE processes: aggregate effects



Notes: The figure compares impulse responses to a QE shock equal to 1.2 percent of steady state output under exogenous QE processes with different degrees of persistence. The left panel shows the path of central bank assets and the right panel shows the associated output responses. The black solid line corresponds to the baseline QE process with $\rho_{\text{QE}} = 0.79$, while the blue line with circles shows the responses under the process with $\rho_{\text{QE}} = 0.99$.

Figure OA14: Comparison of QE processes: distributional effects



Notes: The figure shows relative welfare gains from a one-time QE shock with different degrees of persistence, expressed as ratios to the welfare gain of Q3 households, across the wealth distribution. The left panel reports welfare gains when the fiscal surplus induced by QE is returned to households as transfers (government purchases fixed), while the right panel reports welfare gains when the surplus is used for government purchases (transfers fixed). Yellow bars correspond to QE with persistence $\rho_{\text{QE}} = 0.79$, and green bars to QE with persistence ρ_{QE}

In terms of distributional effects, a more persistent QE process not only delivers larger overall welfare gains, but also yields a flatter distributional profile. That is, QE has more even effects across the wealth distribution when its persistence is higher. By sustaining the improvement in aggregate conditions, persistent QE disproportionately benefits poorer households, because a persistently higher job finding rate moves more unemployed households at the bottom into employment and generates lasting welfare gains. By contrast, the marginal value of additional profits and equity price increases declines for wealthy households as they become richer and earn more over time.

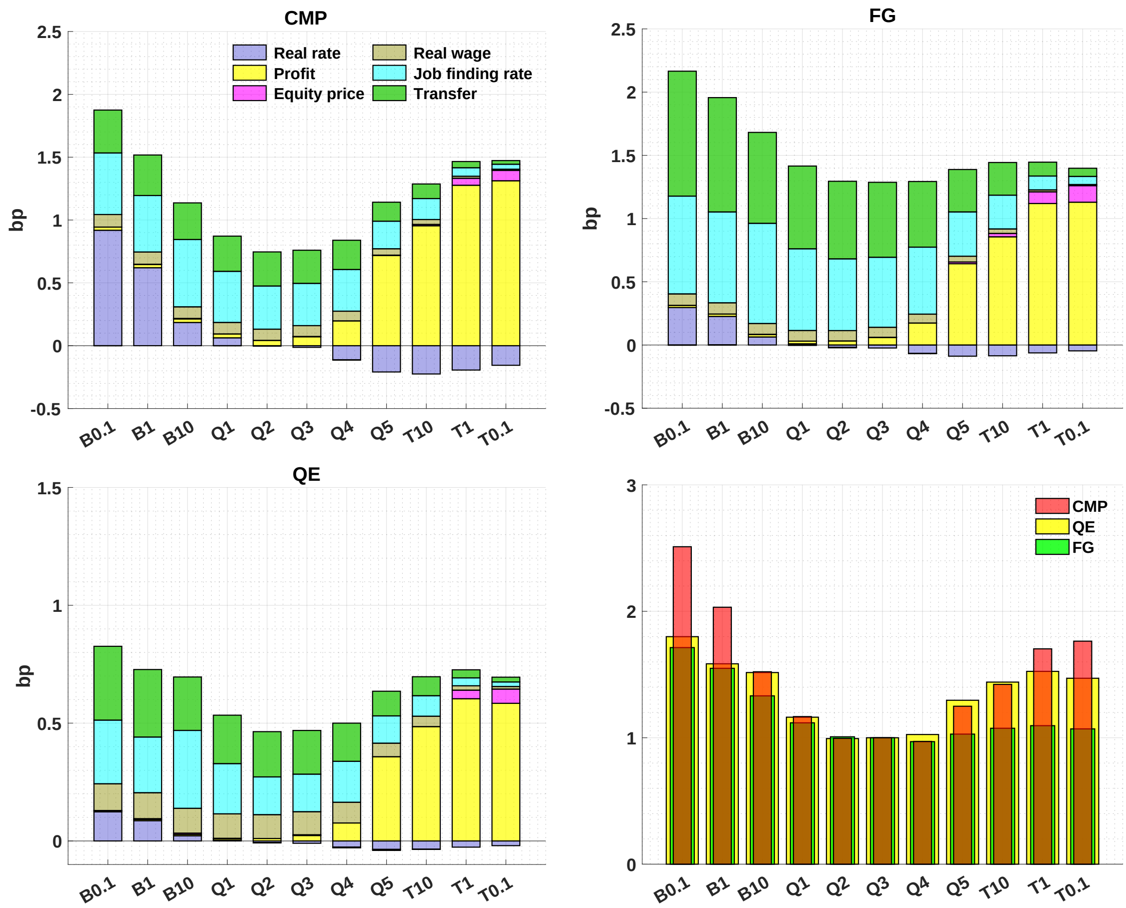
H.2 Welfare effects of one-time monetary policy shocks

In the main text, I focus on the short and medium run distributional effects of UMP shocks by reporting cumulative consumption responses over twenty quarters. In this section, I instead report long run welfare gains, measured in consumption equivalents, across the wealth distribution. Figure OA15 and Figure OA16 show the welfare gains when the fiscal surplus induced by UMP shocks translates into higher lump sum transfers or higher government purchases, respectively.

Two features are worth highlighting. First, in the baseline case where the larger fiscal surplus finances higher transfers, the contribution of lump sum transfers is sizable, especially for FG. As shown in Kaplan et al. (2018) and Lee (2019), fiscal adjustments have important implications for both the transmission and the distributional effects of monetary policy. Under FG, the MMMF behaves like a forward looking representative agent, so the induced transfer responses are large and highly persistent, and they have large effects on long run welfare gains. Because lump sum transfers matter disproportionately for poorer households, FG delivers relatively large welfare gains at the bottom, generating an almost monotonically decreasing profile across most of the distribution, even though households at the top of the wealth distribution still gain more than the middle. For CMP and QE, the U-shaped profile is preserved, with lump sum transfers contributing more to welfare gains than to the cumulative consumption responses in the short and medium run. The ranking of unevenness is the same as for the consumption based measure, CMP displaying the most pronounced U-shaped profile and FG the most even profile.

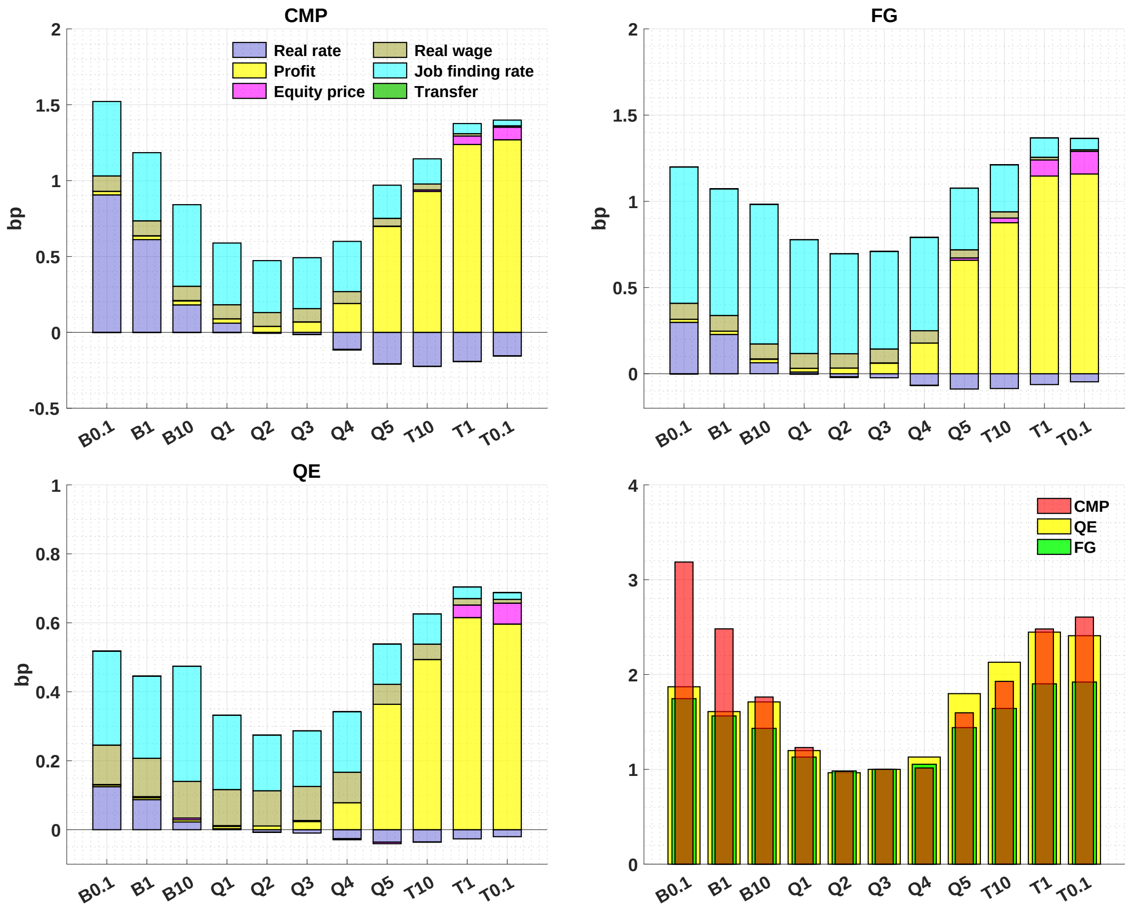
When the fiscal surplus instead shows up as higher government purchases, I obtain a distinct U-shaped distributional profile for all three types of monetary policy shocks. The relative unevenness across policies is unchanged compared with the baseline transfer case, with CMP again most U-shaped and FG the closest to a flat welfare profile.

Figure OA15: Welfare effects of CMP, QE, and FG shocks



Notes: The figure shows welfare gains from different types of monetary policy across the wealth distribution and their decomposition into contributions from real rates, profits, equity prices, real wages, job finding rates, and transfers. The top-left and top-right panels report gains from conventional monetary policy (CMP) and forward guidance (FG), respectively. The bottom-left panel reports gains from QE. The bottom-right panel shows, for each wealth group, welfare gains from CMP, QE, and FG expressed as ratios to the welfare gain of Q3 households.

Figure OA16: Welfare effects of CMP, QE, and FG shocks (Fix T)



Notes: The figure shows welfare gains from different types of monetary policy across the wealth distribution, when government purchase adjusts to balance the budget, and their decomposition into contributions from real rates, profits, equity prices, real wages, job finding rates, and transfers. The top-left and top-right panels report gains from conventional monetary policy (CMP) and forward guidance (FG), respectively. The bottom-left panel reports gains from QE. The bottom-right panel shows, for each wealth group, welfare gains from CMP, QE, and FG expressed as ratios to the welfare gain of Q3 households.

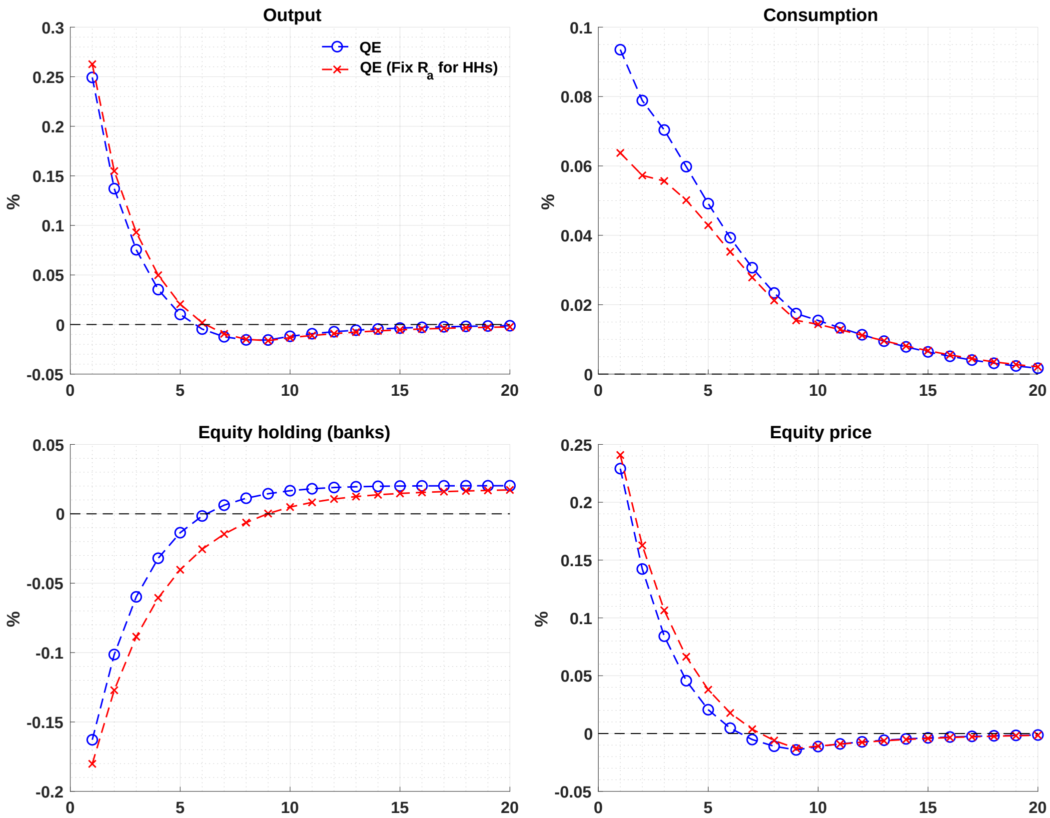
H.3 UMP shocks: shutting down household and bank channels

To better understand the role of households and banks in the transmission of UMP shocks, I examine model responses when shutting down either household or bank channels. Specifically, for a QE shock, I either fix R_a, the realized and expected gross rate of return on equity for households, or fix A_b, banks' equity holdings. For an FG shock, I either fix R, the realized and expected real rate for households, or fix banks' equity holdings. Figures in this section report the results.

In summary, the role of households is limited in the transmission of UMP shocks in the sense that shutting down household channels leads to relatively small differences in output responses. In the QE case, shutting down the household channel leads to slightly larger effects on output on impact, because households' equity investment is not crowded out by QE operations. In stark contrast, when banks' equity holdings are fixed, model dynamics are markedly different. Since banks' equity investment is not crowded out, the QE shock has very large effects on output, consumption, and equity prices on impact. However, because banks are forced to hold the same amount of equity, large increases in prices translate into overinvestment by banks, which leads to a subsequent contraction of the real economy.

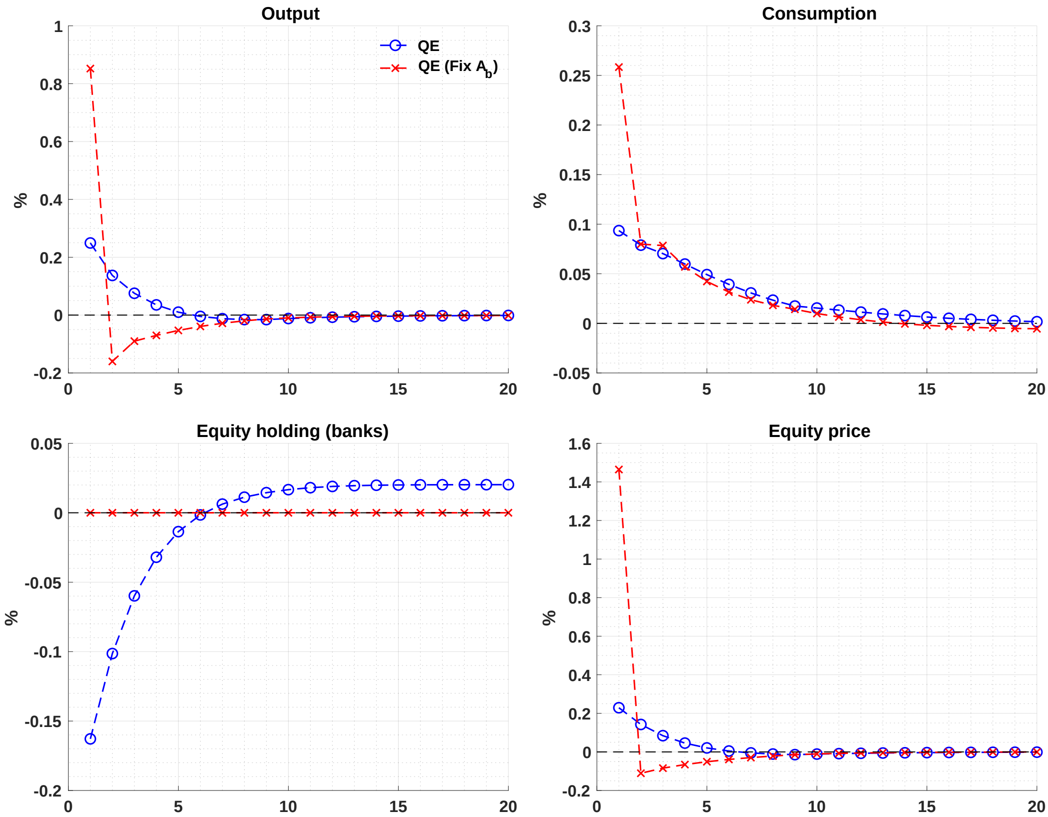
In the case of an FG shock, shutting down the household expectation channel leads to a slightly smaller effect on output on impact, since households' consumption response through intertemporal substitution is missing. Quantitative differences are small, implying that household responses mostly come from income effects, as argued in Kaplan et al. (2018). When banks' equity holdings are instead fixed during the interest rate peg periods, the model dynamics are markedly different. The effects on output and consumption are much smaller on impact and gradually increase until the actual rate cut occurs. In the absence of bank channels, forward guidance has much more muted effects because a significant fraction of households in the HANK model is hand-to-mouth.

Figure OA17: Aggregate effects of QE shocks (Fix R_a for HHs)



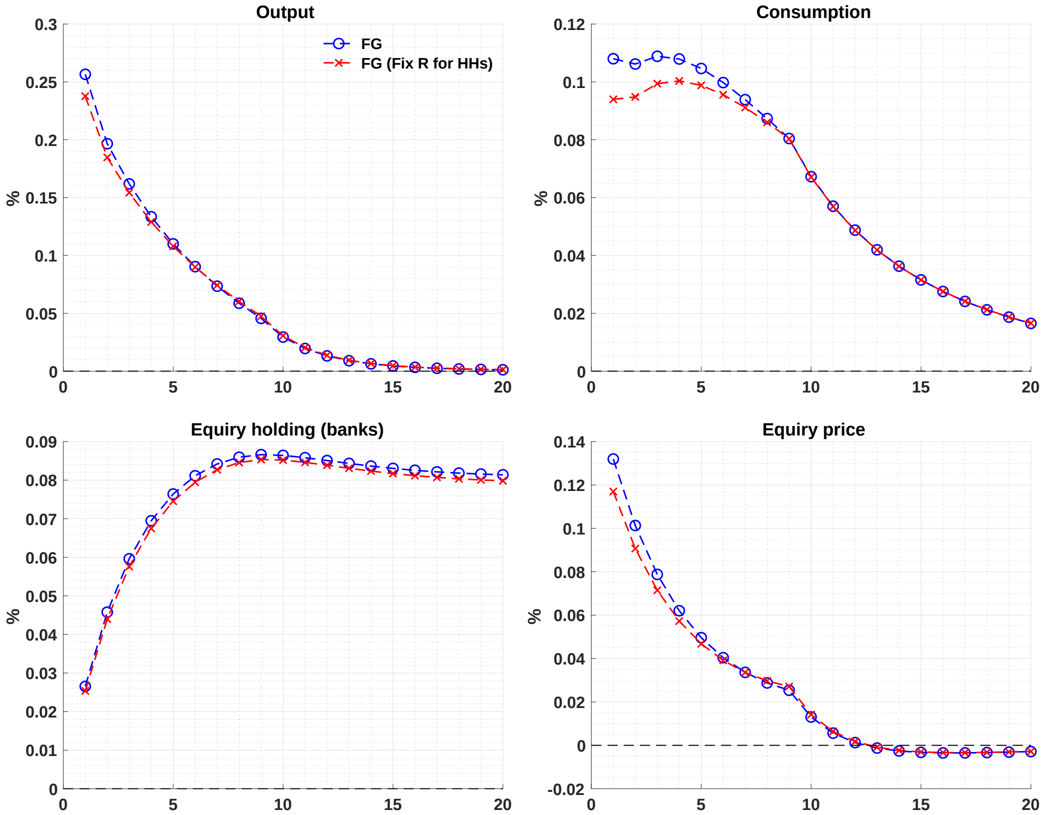
Notes: The figure shows impulse responses to a QE shock in the baseline model (blue lines) and in a variant where the gross equity return for households, R_a, is held fixed so that QE-induced movements in equity prices and returns do not directly affect households (red lines).

Figure OA18: Aggregate effects of QE shocks (Fix A_b)



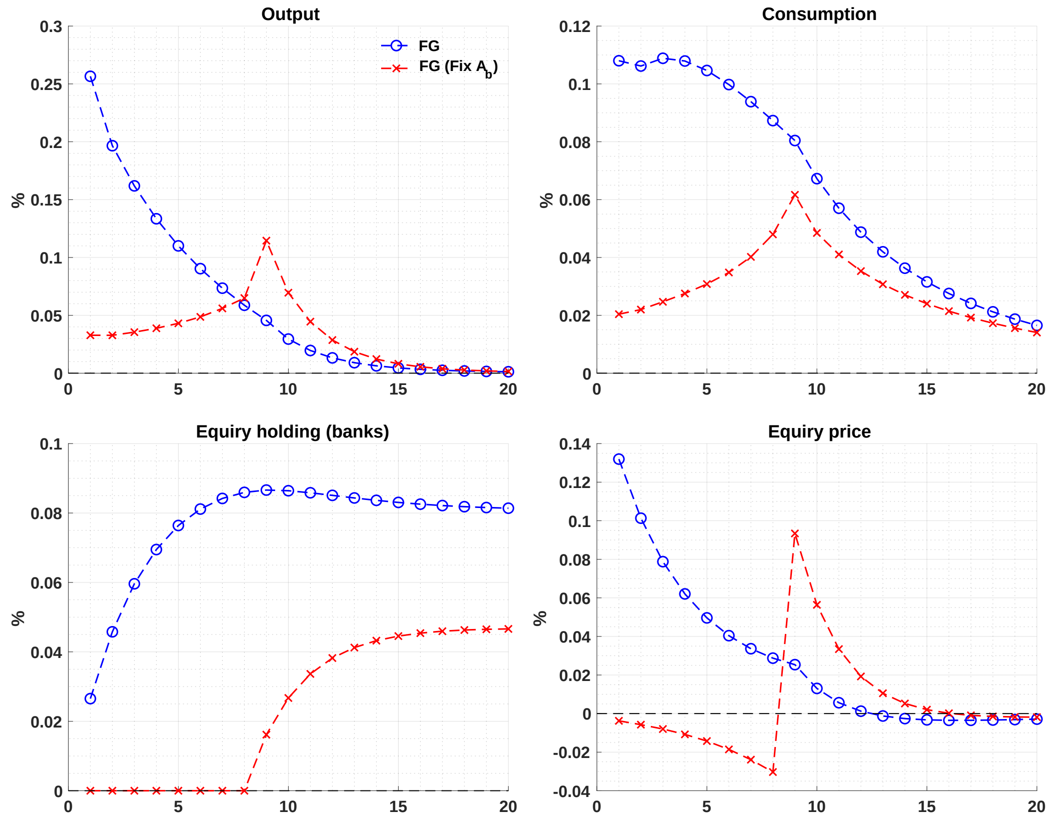
Notes: The figure shows impulse responses to a QE shock in the baseline model (blue lines) and in a variant where banks' equity holdings, A_t^b, are held fixed, so that QE-induced movements in equity prices and returns do not alter bank portfolios (red lines).

Figure OA19: Aggregate effects of FG shocks (Fix R for HHs)



Notes: The figure shows impulse responses to a forward guidance shock in the baseline model (blue lines) and in a variant where the realized and expected real interest rates faced by households are held fixed, so that anticipated future rate cuts have no direct effect on households (red lines).

Figure OA20: Aggregate effects of FG shocks (Fix A_b)

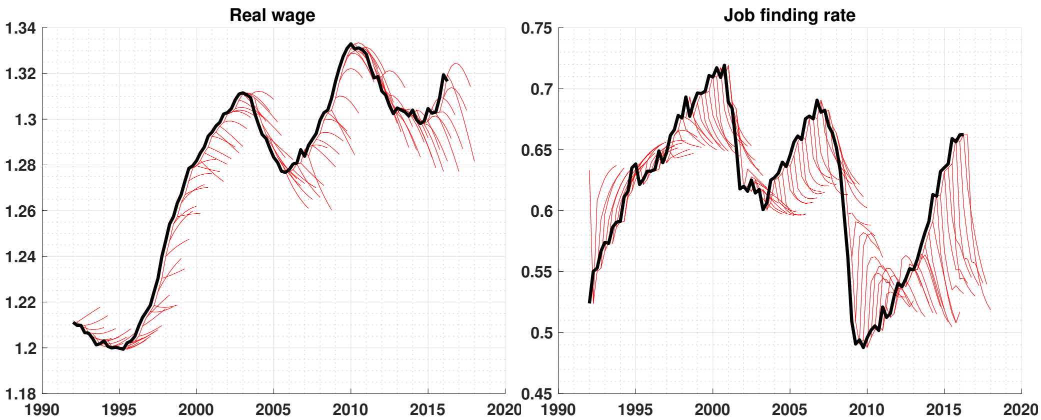


Notes: The figure shows impulse responses to a forward guidance shock in the baseline model (blue lines) and in a variant where banks' equity holdings are held fixed while the nominal interest rate is pegged. Once the peg ends, banks again respond to economic conditions and adjust their equity positions (red lines).

I Further details on the model simulations

I.1 The decomposition method

Figure OA21: Realized and expected paths of the real wage and the job-finding rate



Notes: The figure shows the realized values of the equity price in the sample along with its expected path in each period. The thick black line shows the realized path and the red 'hairs' are the expectations.

To evaluate the relative contribution of various channels, I compute no shock paths of the following variables each period in the estimation.

$$\{ w_{t,t+j}\ ,\ i_{t,t+j}\ ,\pi_{t,t+j}\ ,\ q_{t,t+j}\ ,\ r^a_{t,t+j},\ \Pi_{t,t+j}\ ,T_{t,t+j}\ ,f_{t,t+j}\}_{j=1}^{T} \tag{128}$$

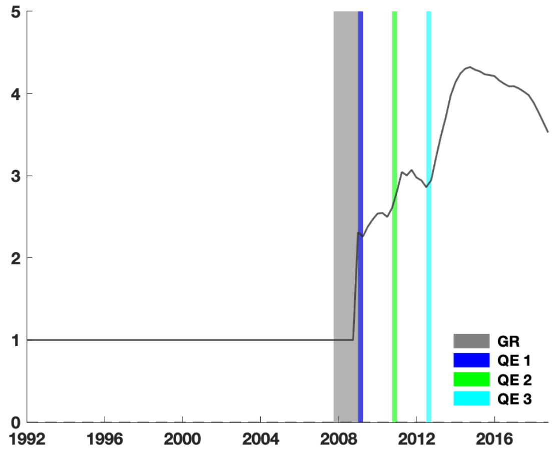
where $x_{t,t+j}$ is the expected value of x in period $t+j$ given the information in period t. T is a very large number that ensures that $x_{t,t+T}$ converges to its steady state value in T periods. The above eight variables, i.e., real wage, nominal rate, inflation rate, equity price, dividend rates, total lump-sum transfer, and the job-finding rate, are what determine the household's optimal decisions and welfare together with the expected future value (utility) of households' choices. Exploiting the fact that the expected future shocks are zero each period in the model, I compute the expected paths of the above variables both in the baseline and the alternative cases.[80] Using different combinations of these paths, I solve the household's problem from $t+T$ periods backwardly and compute households' optimal decisions and values (utility). For instance, in one path, I assume that only the job-finding rate follows the path in the baseline case, and all other variables follow the path in the alternative case. By computing households' optimal decisions and the associated utility in this path and comparing them with optimal decisions and values in the alternative case, I can compute the contribution of the job-finding rate on the behavior and expected welfare of households in a given period in the baseline case.[81] Figure OA21 shows the realized path of the real wage and the job-finding rate along with each period's household expectations on it.

[80] The number of the expected paths is equal to the number of periods in the sample multiplied by the number of variables. The starting value of each path, i.e., $x_{t,t}$, coincides with the realized value as it is observed, but all the future expectations are not necessarily correct because of unexpected shocks in the future.

[81] The decomposition uses eight counterfactual combinations in which one block of variables follows the baseline path and all others remain on their counterfactual paths: (1) none switched, (2) all switched, (3) profit and dividend rates, (4) nominal rate and inflation, (5) real wage, (6) job finding rate, (7) equity price, (8) total transfers.

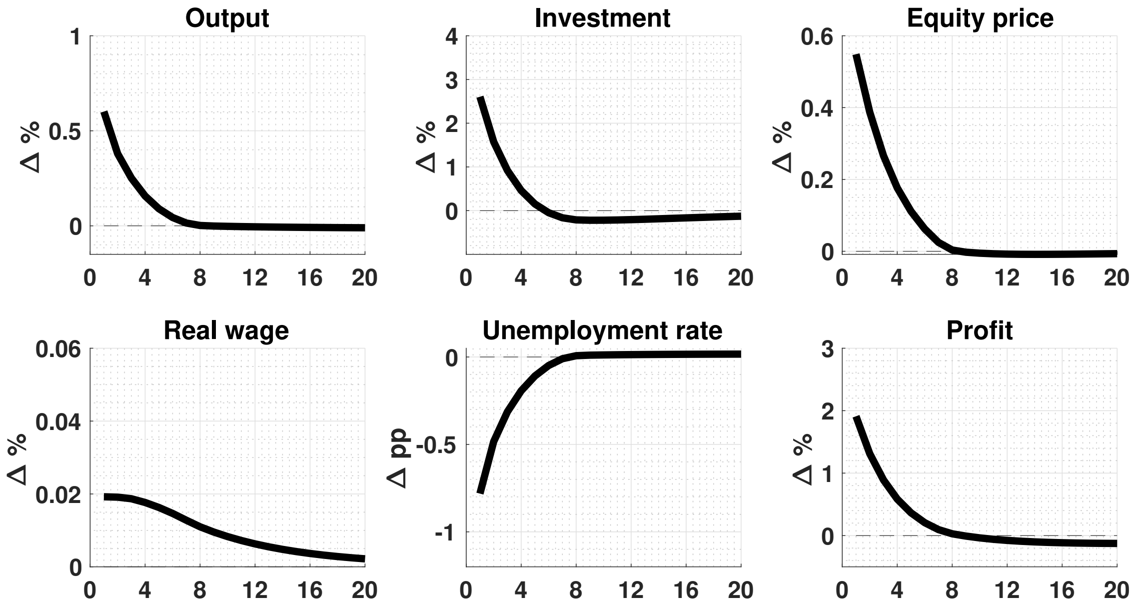
I.2 Additional results from model simulations

Figure OA22: The central bank's assets



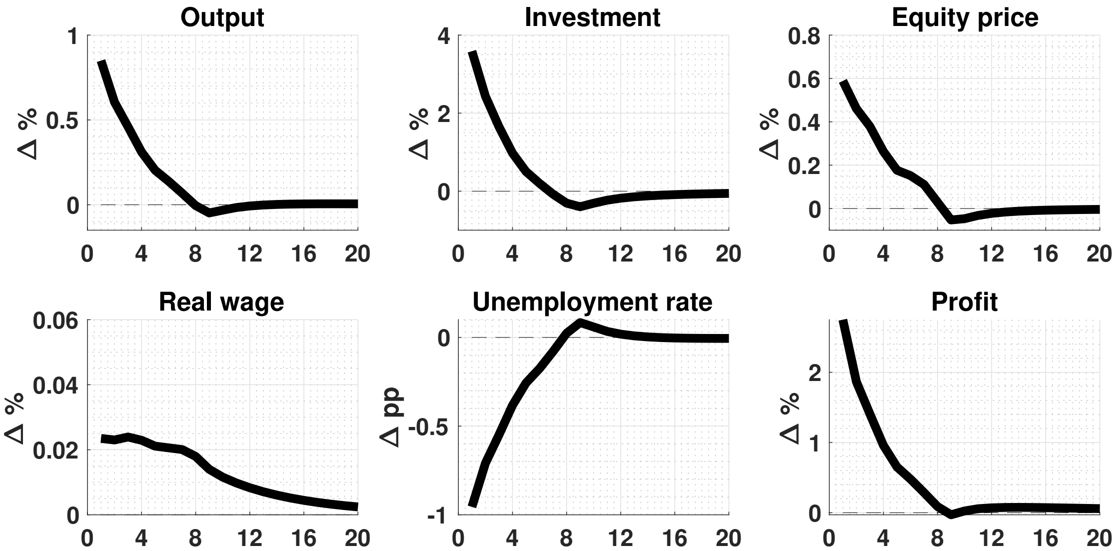
Notes: The figure shows the central bank's asset as the ratio to its end of 2007 level. Green, blue, green, and sky blue area depict the Great Recession periods, the period in which QE 1, 2, and 3 are announced.

Figure OA23: Average responses to QE at the ELB



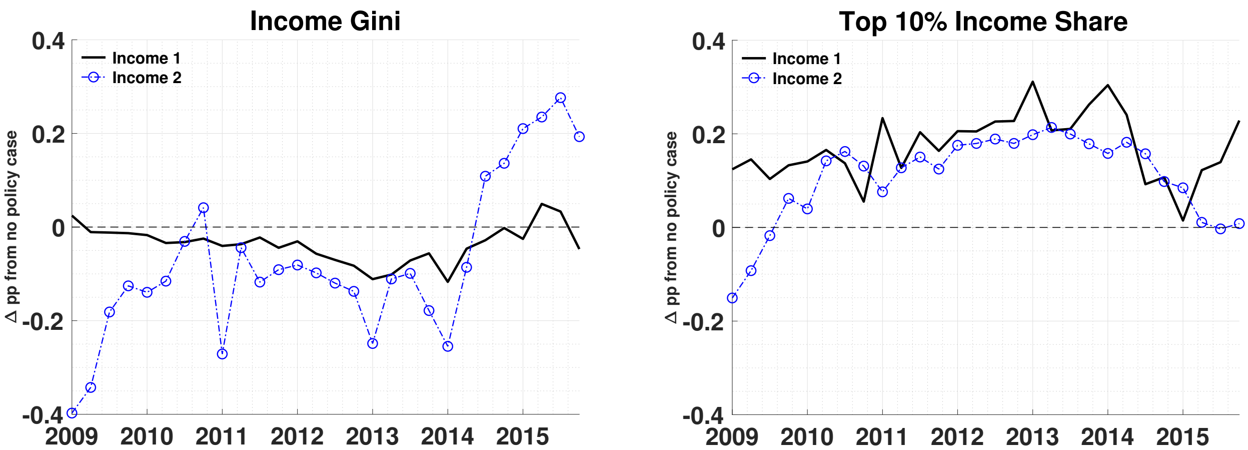
Notes: The figure shows average impulse responses to a QE shock scaled to 5 percent of steady state output. For each starting date between 2009Q1 and 2015Q4 in the ELB episode, the model is simulated for 20 quarters after an unanticipated QE shock with all other shocks set to zero. The black line plots, at each horizon, the average response across these 28 starting dates.

Figure OA24: Average responses to forward guidance at the ELB



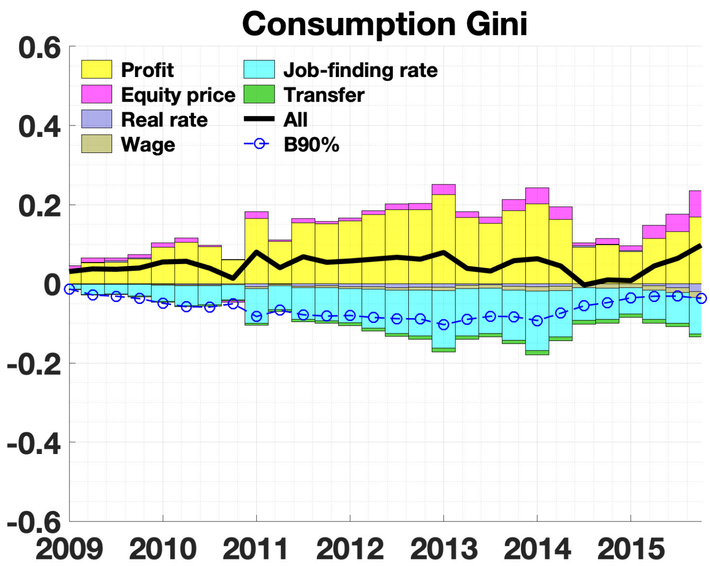
Notes: The figure shows average impulse responses to a forward guidance shock that lengthens the exogenous ELB duration by one quarter. For each starting date between 2009Q1 and 2015Q4 in the ELB episode, the model is simulated for 20 quarters after an unanticipated increase in the expected ELB duration, with all other shocks set to zero. The black line plots, at each horizon, the average response across these 28 starting dates.

Figure OA25: Effects of UMP on inequality measures: alternative income



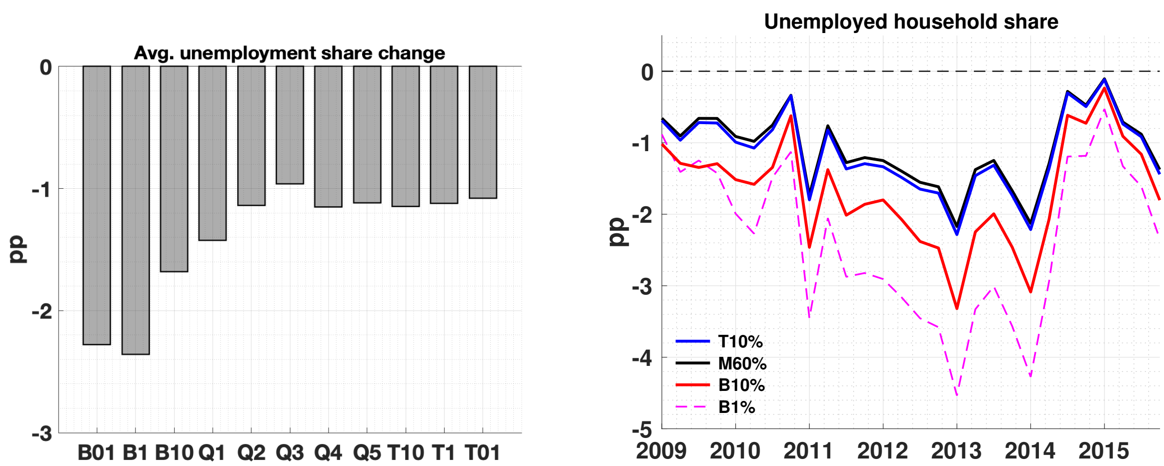
Notes: The left panel plots the difference between the baseline with UMP and the counterfactual without UMP for the income Gini, reported in index points relative to the no UMP case on a 0 to 100 scale. The right panel plots the corresponding difference for the top 10 percent income share, reported in percentage points relative to the no UMP case. In both panels, the black solid line shows the posterior mode path under the baseline income definition, which excludes capital gains. The blue dotted line shows the path under an alternative income definition that assumes all equity is purchased at the steady state price.

Figure OA26: Distributional effects of UMP: consumption Gini index



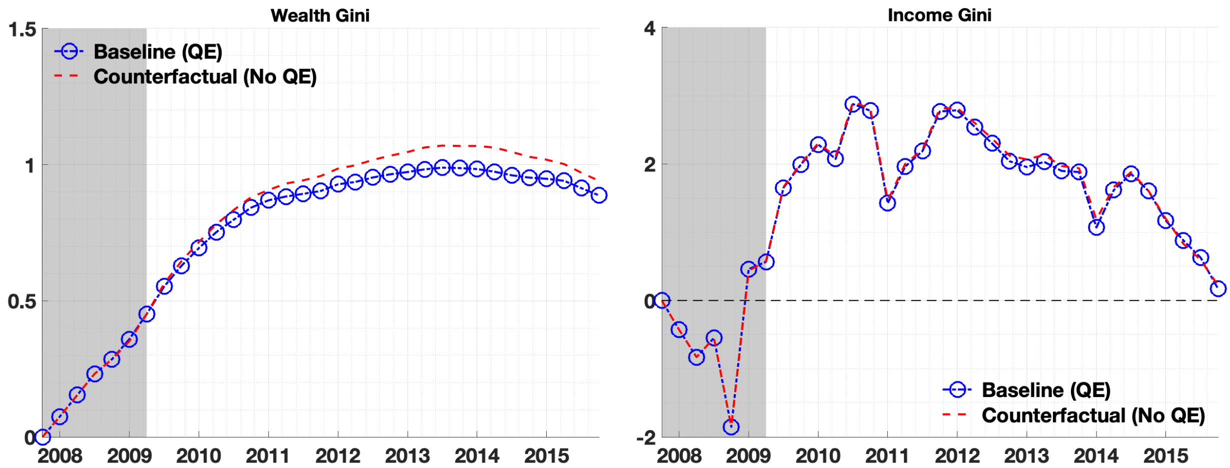
Notes: The figure shows relative degrees of inequality in the model during the ELB episode as differences in the Gini index between the baseline and the counterfactual case. The thick black line shows the overall effects of UMP, while each bar shows the contribution of each variable to the overall effects. The blue dotted line with circles shows the Gini index computed from households at the bottom 90% of the wealth distribution. The Y-axis unit is the difference in the Gini index, which is on a zero to 100 scale.

Figure OA27: Unemployed household shares across wealth groups



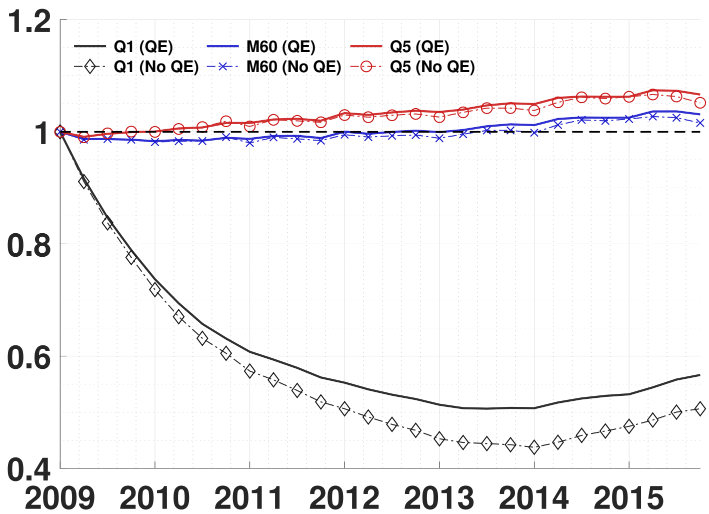
Notes: The left panel shows average changes in the share of unemployed households induced by UMP across wealth groups. The right panel shows the evolution of unemployed household shares during the ELB episode, as percentage point difference from the corresponding values in the counterfactual case with no unconventional policy interventions. The blue, black, red, and dashed pink lines show the share of unemployed households in the top 10%, the middle 60%, the bottom 10%, and the bottom 1% of the wealth distribution, respectively.

Figure OA28: Wealth and income inequality during the ELB period: Gini index



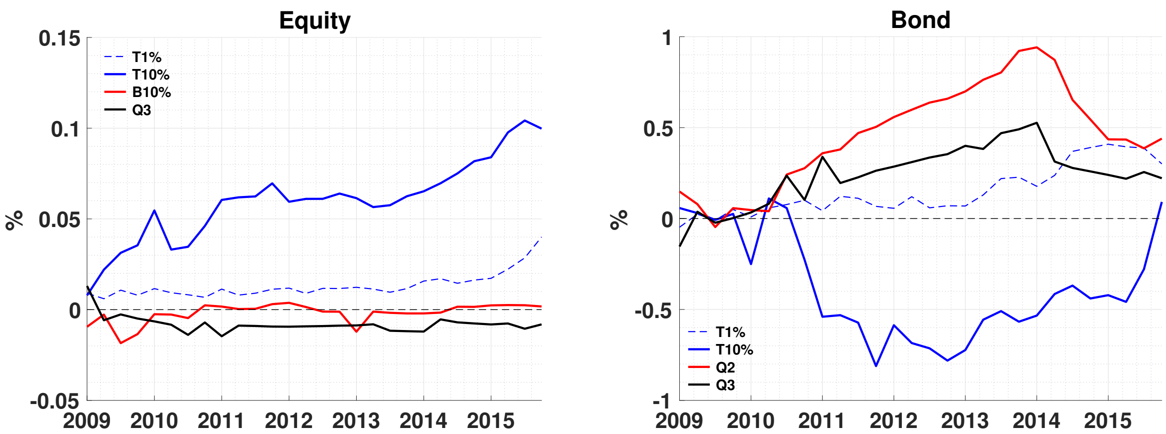
Notes: The figure shows the evolution of the wealth and income Gini indices during the ELB episode, as differnces of the index relative to its 2007 Q4 level. The blue lines with circles show the Gini indices in the baseline case. The dashed red lines show the Gini indices in the counterfactual case.

Figure OA29: Effects of UMP on households' wealth



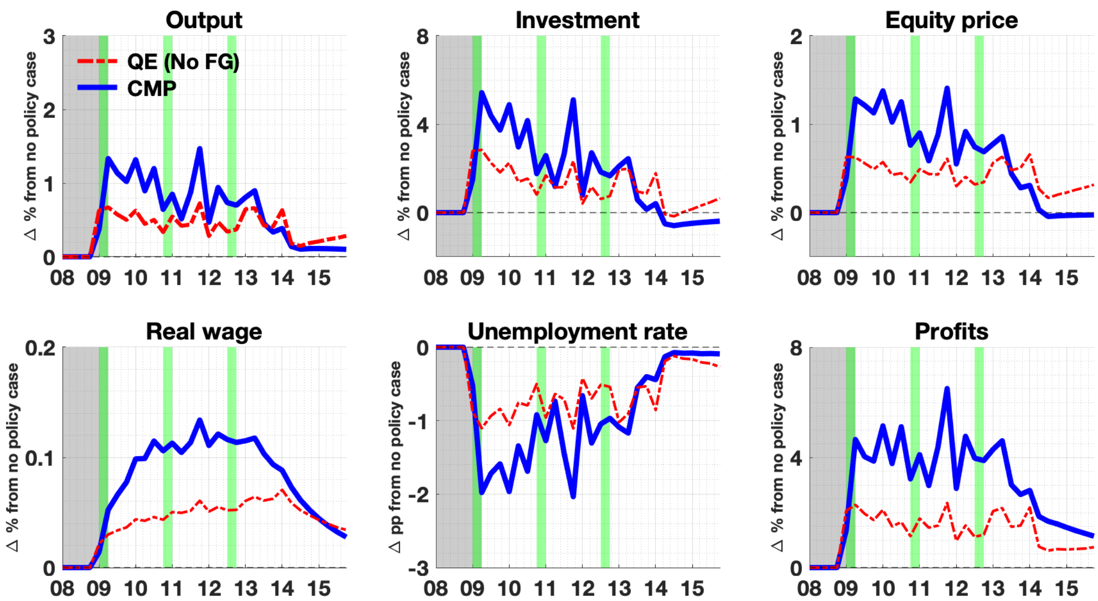
Notes: The black, blue, and red straight lines show the ratio of households' wealth during the ELB episode, relative to their 2009 Q1 level, in the baseline case with UMP. The black, blue, and red dashed lines with diamond, crosses, and circles show the ratio in the counterfactual case with no UMP.

Figure OA30: Effects of UMP on equity and bond shares



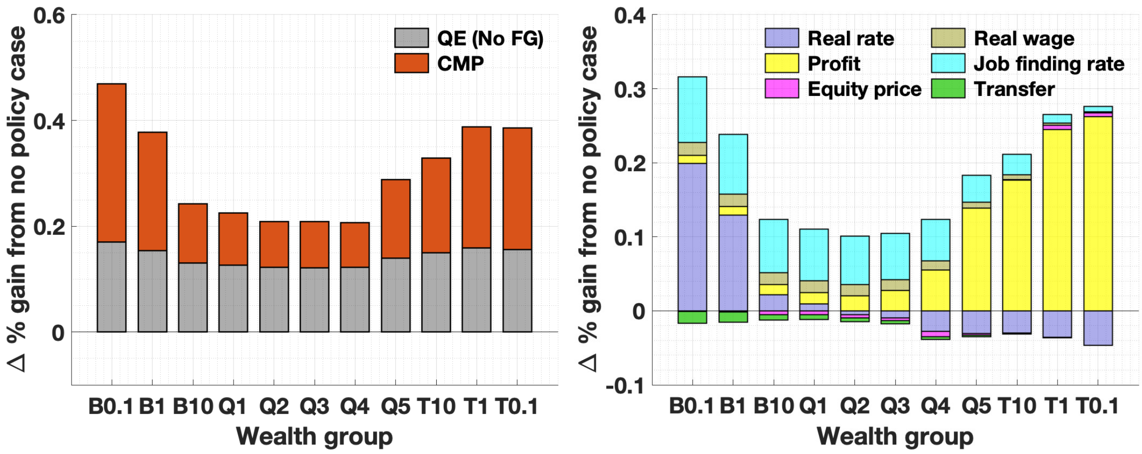
Notes: The figure shows different wealth groups' equity and bond shares during the ELB episode. T1%, T10%, B10%, Q2 and Q3 refers to the top 1% and 10%, the bottom 10%, the second and the middle quintile, respectively. The unit is the difference in the share of equity and bond between the baseline and the counterfactual case.

Figure OA31: Aggregate effects: QE based UMP vs CMP



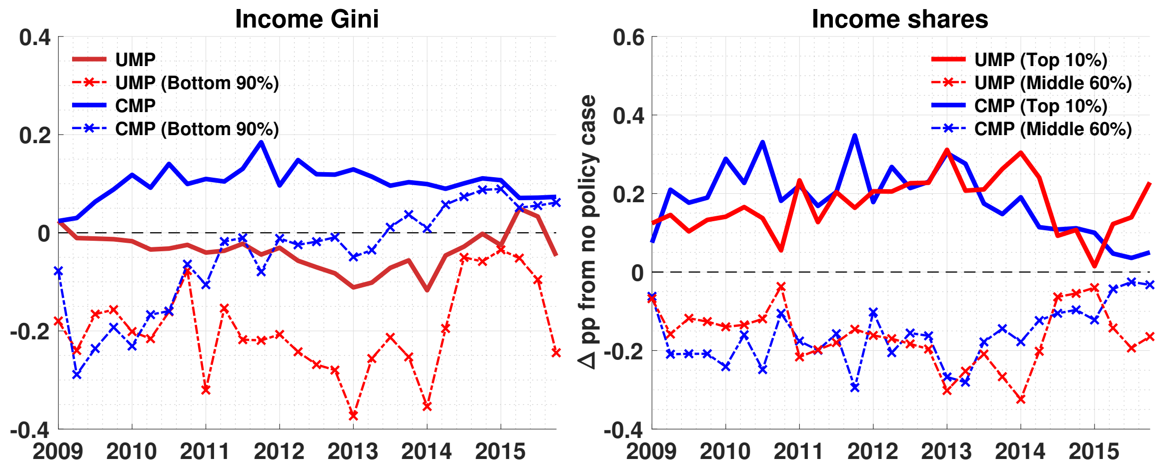
Notes: Red and blue lines denote QE and CMP, respectively. All variables except the unemployment rate are percentage differences from the no policy counterfactual. The unemployment rate is in percentage point differences.

Figure OA32: Welfare gains: QE vs CMP



Notes: Left panel reports consumption equivalent gains across wealth groups for QE (gray) and the additional gains under CMP (dark orange). Right panel decomposes the additional CMP gains.

Figure OA33: Distributional effects on inequality measures: UMP vs CMP



Notes: The left panel shows the effects of QE and CMP on the income Gini index, with the blue and red solid lines representing the overall index, while the blue and red dashed lines indicate the effects on the Gini index for the bottom 90%. The right panel illustrates the effects on income shares across different household groups, with the red lines representing the effects of QE alone and the blue lines showing the effects of CMP.

J A RANK version

In this section, I build a RANK version of the HANK model used in the paper by incorporating a representative family assumption. The family consists of members who may work, be unemployed, or own businesses. Individual members' incomes are pooled by the family head, who chooses consumption and saving for the family. The family's problem is

$$V(B_t, A_t) = \max_{X_t, B_{t+1}, A_{t+1}} \frac{X_t^{1-\sigma}}{1-\sigma} + \beta\Delta^h \mathbb{E}[V(B_{t+1}, A_{t+1})] \tag{129}$$

s.t.

$$X_t + \frac{B_{t+1}}{1-\zeta} + \frac{q_t A_{t+1}}{1-\zeta} + \Phi^h(A_{t+1}, A_t) = Y_t^h + \frac{1+i_t}{\pi_t} B_t + (q_t + \widetilde{r}_{a,t})A_t + \frac{\iota_b}{\pi} B^- + T_t \tag{130}$$

$$Y_t^h = (1-\tau_w)\left\{(1-u_t)(1-\mu^e)\bar{s}_e w_t n_t \times \frac{\xi}{\xi+1} + \nu u_t (1-\mu^e)\bar{s}_u w n + \mu^e \Pi_t\right\}, \tag{131}$$

where μ^e denotes the mass of business owners, Y_t^h is after tax family income, B^- is steady state aggregate borrowing from the baseline HANK economy, and X_t is the consumption and hours composite

$$X_t \equiv C_t - (1-u_t)(1-\mu^e)\, \bar{s}_e\, \psi\, \frac{n_t^{1+1/\xi}}{1+1/\xi}. \tag{132}$$

The equity adjustment cost is given by

$$\Phi^h(A_{t+1}, A_t) \;=\; \phi_1\,(A_{t+1} - A) \;+\; \frac{\phi_2}{2}\,(A_{t+1} - A)^2. \tag{133}$$

The formulation reflects HANK specific features, such as borrowing, stochastic death, and annuity arrangements, to ensure that the representative family chooses the same aggregate resource allocations as in the HANK model. In particular, the effective purchase prices are $1/(1-\zeta)$ for bonds and $q_t/(1-\zeta)$ for equity, and the effective dividend rate on equity is $\widetilde{r}_{a,t+1} = r^a_{t+1} + q_t \frac{\zeta}{1-\zeta}$. There also exists a fixed wedge Δ^h to ensure that the steady state liquid asset optimality condition holds with β.[82]

First-order conditions Let λ_t denote the Lagrange multiplier on the budget constraint. The first order condition with respect to X_t is

$$\lambda_t \;=\; \frac{\partial}{\partial X_t}\left(\frac{X_t^{1-\sigma}}{1-\sigma}\right) \;=\; X_t^{-\sigma}. \tag{134}$$

Define the stochastic discount factor

$$M_{t,t+1} \;\equiv\; \beta\Delta^h \frac{\lambda_{t+1}}{\lambda_t} \;=\; \beta\Delta^h \left(\frac{X_{t+1}}{X_t}\right)^{-\sigma}. \tag{135}$$

(Deposit holding Euler equation) Differentiating with respect to B_{t+1} yields

$$1 \;=\; (1-\zeta)\,\mathbb{E}_t\left[M_{t,t+1}\,\frac{1+i_{t+1}}{\pi_{t+1}}\right]. \tag{136}$$

[82] In the HANK version, $\beta(1-\zeta) < \pi/(1+i)$ due to idiosyncratic income risk and the associated precautionary motive.

(Equity holding Euler with adjustment costs) With $\Phi^h(A_{t+1}, A) = \phi_1(A_{t+1} - A) + \frac{\phi_2}{2}(A_{t+1} - A)^2$,

$$\Phi_{1,t} \equiv \frac{\partial \Phi^h(A_{t+1}, A)}{\partial A_{t+1}} = \phi_1 + \phi_2 (A_{t+1} - A), \qquad \frac{\partial \Phi^h(A_{t+2}, A)}{\partial A_{t+1}} = 0. \tag{137}$$

The first order condition for A_{t+1} is therefore

$$\frac{q_t}{1-\zeta} + \phi_1 + \phi_2 (A_{t+1} - A) = \mathbb{E}_t \left[M_{t,t+1} \left(q_{t+1} + \widetilde{r}_{a,t+1} \right) \right]. \tag{138}$$

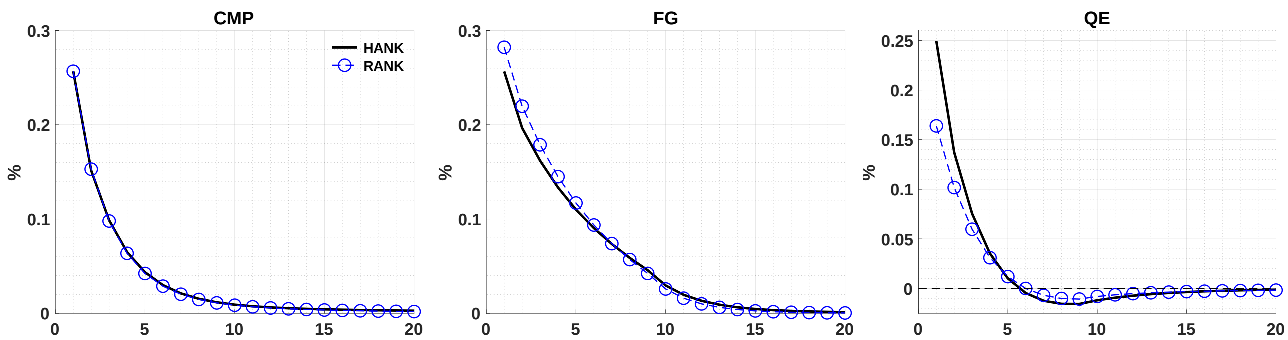
Steady state At a nonstochastic steady state with $A_{t+1} = A$ and $X_{t+1} = X_t$, we have $M_{t,t+1} = \beta\Delta^h$. Then

$$\text{Bonds:} \quad 1 = (1-\zeta)\,\beta\Delta^h \frac{1+i}{\pi}, \tag{139}$$

$$\text{Equity:} \quad \frac{q}{1-\zeta} + \phi_1 = \beta\Delta^h \left(q + \widetilde{r}_a \right), \tag{140}$$

the above steady state conditions pin down the wedge Δ^h and the coefficient ϕ_1 on the linear part of the equity adjustment cost. I have freedom to adjust ϕ_2 to control household equity saving dynamics out of the steady state. For comparison with the baseline HANK model, I set ϕ_2 to 0.024 so that the RANK version of the model delivers an impulse response of output to a 25 bp expansionary monetary policy shock that is closest to the IRF from the HANK version.

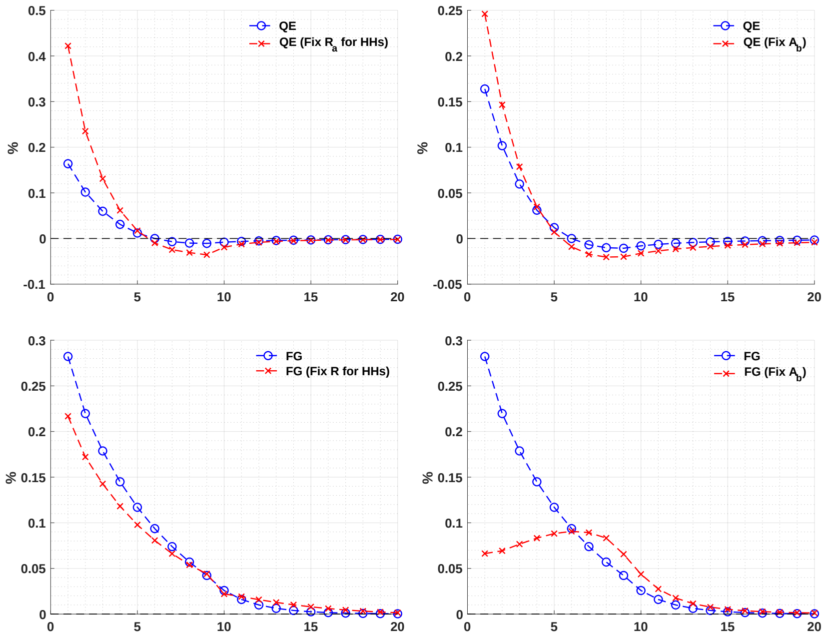
Figure OA34: UMP shocks in HANK and RANK



Notes: Figure shows IRFs of output to different kinds of monetary policy shocks in the baseline HANK and the correspdnding RANK version of the model. The black solid lines are IRFs in the HANK, while blue lines with circles are IRFs in the RANK version.

Figure OA34 compares the output responses in the HANK and RANK versions of the model to different types of monetary policy shocks. By construction, the equity adjustment process in the RANK model is chosen so that the impulse response of output to a CMP shock matches that in the HANK model, so the CMP responses coincide. For UMP, however, the responses differ. In the RANK version, FG is more effective while QE is less effective. The difference reflects how many households are exposed to asset return and expectation channels. FG works by lowering expected real interest rates and inducing households to shift consumption toward the present, whereas QE reduces the expected gross return on equity and crowds out private equity investment. In the RANK model, all households are exposed to these intertemporal channels, so FG has a larger effect and QE experiences stronger crowding out. In the HANK model, equity is held by a small fraction of wealthy households and there is a sizable share of hand-to-mouth households, so expectation and asset return channels operate on a narrower part of the population. As a result, QE produces stronger net stimulus in the HANK model, because fewer households reduce equity holdings, while FG has smaller effects because intertemporal substitution in consumption is weaker.

Figure OA35: UMP shocks in RANK: shutting down household and bank channels



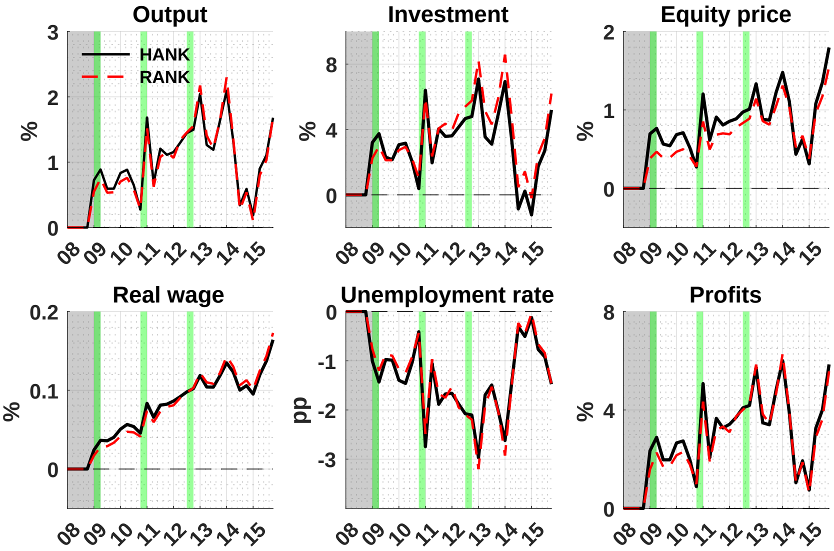
Notes: The figure shows output responses to QE and FG shocks in the RANK model under the baseline specification (blue lines) and counterfactuals that shut down household or bank channels (red lines). The top panels plot responses to QE shocks: the top left panel fixes the equity return faced by households, and the top right panel holds banks' equity holdings fixed. The bottom panels plot responses to forward guidance shocks: the bottom left panel fixes realized and expected real rates for households, and the bottom right panel holds banks' equity holdings fixed while the nominal interest rate is pegged for the first eight quarters.

Experiments that shut down asset return and expectation channels for households and banks confirm that these channels matter more for households in the RANK model than in the HANK model. As shown in Figure OA35, when the asset channels are switched off for households, the model dynamics differ much more from the baseline than in the analogous experiments for the HANK version. For example, when the asset return channel does not operate for households, QE generates a much stronger stimulus, with the impact increase in output more than twice as large as in the baseline. Likewise, when forward guidance no longer induces intertemporal substitution in consumption, the impact response of output is about 0.1 percent lower.

Consistent with these differences in the effects of UMP between the HANK and RANK versions, the aggregate effects of UMP are smaller in the RANK model early in the ELB episode, when QE plays the dominant role. Later in the episode, when FG becomes more important, the model-implied effects are larger in the RANK version, as shown in Figure OA36. Because there are differences in model dynamics, I also estimate the RANK model, and Table OA9 reports the results.[83] At the posterior mode of the RANK estimation, the aggregate effects of UMP are substantially larger, especially in later periods when forward guidance plays a predominant role in boosting the economy.

[83]Expected ELB durations are not re-estimated and are instead fixed at the posterior mode values from the estimation with the HANK model to speed up the estimation.

Figure OA36: Aggregate effects of UMP in the HANK and the RANK model



Notes: The figure shows the aggregate effects of UMP in the HANK (solid black) and the corresponding RANK version (dotted red). For level variables X, each panel reports $100 \times (X^{\mathrm{UMP}}/X^{\mathrm{NoUMP}} - 1)$ in percent. For the unemployment rate, panels report percentage point differences $u^{\mathrm{UMP}} - u^{\mathrm{NoUMP}}$.

Figure OA37: Aggregate effects of UMP in the HANK and the RANK model

Notes: The figure shows the aggregate effects of UMP in the HANK (solid black) and the corresponding RANK version with a newly estimated posterior mode (dotted red). For level variables X, each panel reports $100 \times (X^{\mathrm{UMP}}/X^{\mathrm{NoUMP}} - 1)$ in percent. For the unemployment rate, panels report percentage point differences $u^{\mathrm{UMP}} - u^{\mathrm{NoUMP}}$.

Table OA9: Prior and posterior distributions of structural parameters

Symbol	Description	Prior			Posterior		
		Prior Density	Mean	Std	Mode	10%	90%
		Frictions					
κ	Slope of Phillips curve	Gamma	0.10	0.02	0.0431	0.0341	0.0639
ι_p	Price indexation	Gamma	0.50	0.15	0.1448	0.0858	0.2089
ρ_w	Wage autocorrelation	Beta	0.50	0.20	0.8111	0.7003	0.8656
ι_w	Wage indexation	Beta	0.50	0.15	0.1282	0.0871	0.2190
ϕ	Capital adjustment cost	Normal	30.00	5.00	46.976	45.778	48.796
ι	Vacancy posting cost	Gamma	0.05	0.02	0.0321	0.0207	0.0534
		Government policy					
ρ_B	Bond issuance rule	Beta	0.50	0.20	0.5931	0.0426	0.7092
ρ_g	Lump sum transfer shock AR	Beta	0.50	0.20	0.9348	0.9271	0.9903
σ_G	Lump sum transfer shock std dev	Inverse-Gamma	0.10	2.00	0.0685	0.0635	0.0766
ρ_R	Interest rate smoothing	Beta	0.50	0.20	0.7680	0.7249	0.7863
σ_R	Interest rate shock std dev	Inverse-Gamma	0.10	2.00	0.1786	0.1621	0.2116
ϕ_π	Taylor rule inflation gap response	Normal	1.70	0.30	1.4721	1.3745	1.5938
ϕ_y	Taylor rule unemployment gap response	Gamma	0.10	0.05	0.3751	0.3431	0.4232
		Structural Shocks					
ρ_l	Liquidity preference shock AR	Beta	0.50	0.20	0.9988	0.9995	0.9999
σ_l	Liquidity preference shock std dev	Inverse-Gamma	0.10	2.00	0.0498	0.0447	0.0567
ρ_z	TFP shock AR	Beta	0.50	0.20	0.9927	0.9906	0.9947
σ_z	TFP shock std dev	Inverse-Gamma	0.10	2.00	0.5690	0.5300	0.6284
ρ_p	Price markup shock AR	Beta	0.50	0.20	0.9604	0.9462	0.9672
σ_p	Price markup shock std dev	Inverse-Gamma	0.10	2.00	1.7355	1.4484	2.1744
ρ_k	Investment shock AR	Beta	0.50	0.20	0.9807	0.9524	0.9877
σ_k	Investment shock std dev	Inverse-Gamma	0.10	2.00	0.0666	0.0608	0.0723
ρ_b	Risk premium shock AR	Beta	0.50	0.20	0.9433	0.9346	0.9530
σ_b	Risk premium shock std dev	Inverse-Gamma	0.10	2.00	0.2156	0.1907	0.2444
ρ_Ξ	Fixed cost shock AR	Beta	0.50	0.20	0.9126	0.8946	0.9330
σ_Ξ	Fixed cost shock std dev	Inverse-Gamma	0.10	2.00	0.9783	0.8965	1.1160
σ_w	Wage shock std dev	Inverse-Gamma	0.10	2.00	0.7259	0.4770	1.3226

Notes The values for the standard deviations and the measurement error are multiplied by 100.

K time-varying copula for the household distribution

For the model solution, I represent the joint distribution of household states (b, a, s) by separating marginal distributions and the dependence structure. Let $m_{b,t}(b)$, $m_{a,t}(a)$, and $m_{s,t}(s)$ denote the period t marginals over bond holdings, illiquid assets, and idiosyncratic earnings states, and let $U_{b,t}(b)$, $U_{a,t}(a)$, and $U_{s,t}(s)$ be the associated marginal CDFs. By Sklar's theorem, the joint density can be written as

$$\mu_t(b, a, s) \;=\; c_t\big(U_{b,t}(b), U_{a,t}(a), U_{s,t}(s)\big)\, m_{b,t}(b)\, m_{a,t}(a)\, m_{s,t}(s), \tag{141}$$

where c_t is a copula density on $[0, 1]^3$ that captures the dependence across the three state variables. In the fixed copula specification I set $c_t(u) \equiv c^{\mathrm{SS}}(u)$ equal to its steady state value, so that only the marginals move over time. This preserves the steady state dependence structure while allowing the cross section to evolve through changes in the marginal distributions.[84]

As a robustness check, I also consider a time-varying copula following Bayer and Luetticke (2020). On the unit cube $u = (u_b, u_a, u_s) \in [0, 1]^3$ I approximate the copula density by a deviation around the steady state benchmark using an orthogonal cosine basis,

$$c_t(u) \;\approx\; c^{\mathrm{SS}}(u) \;+\; \sum_{k=1}^{N_{\mathrm{cop}}} \theta_{k,t}\, \psi_k(u), \tag{142}$$

where $\{\psi_k\}$ are three dimensional discrete cosine basis functions defined on a finite copula grid in (u_b, u_a, u_s) and only a low frequency triangular subset is retained. The coefficients $\{\theta_{k,t}\}_{k=1}^{N_{\mathrm{cop}}}$ summarize time variation in cross sectional dependence in a low dimensional way and are stacked together with the marginals in the reduced state vector,

$$S_t \;=\; \big(m_{b,t}, m_{a,t}, m_{s,t}, \theta_{1,t}, \ldots, \theta_{N_{\mathrm{cop}},t}\big).$$

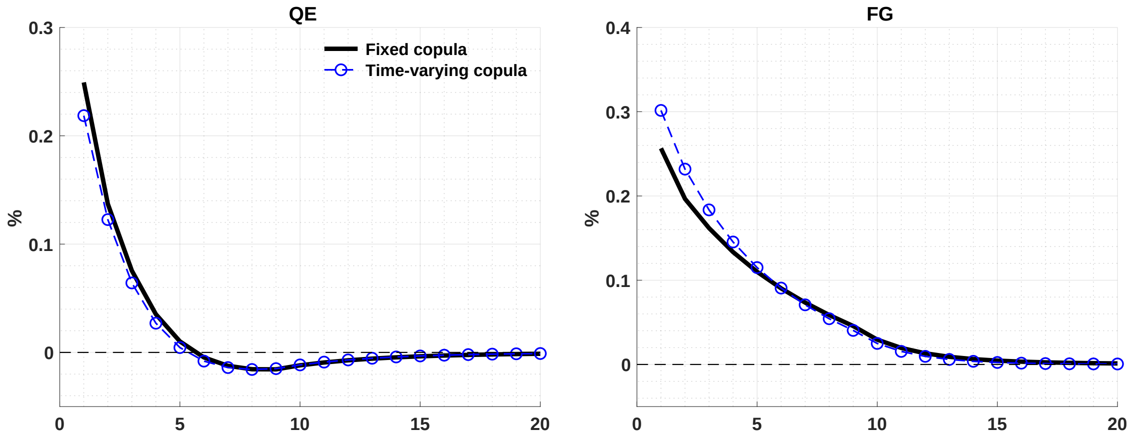
In the code, I follow Bayer and Luetticke (2020) and update S_t by mapping the full joint distribution forward through the transition matrix, computing the implied joint CDF on the copula grid, taking the deviation from the steady state copula CDF, and projecting the resulting density deviation back onto the discrete cosine basis. This procedure keeps the effective state dimension small while allowing the dependence structure to move over time. A time-varying copula can capture cross sectional dynamics more accurately when dependence across marginals varies over time in a meaningful way, while if this dependence is close to constant the fixed and time-varying copula specifications deliver very similar model dynamics.[85]

Figure OA38 compares the effects of UMP shocks on output under solutions with a fixed copula and a time-varying copula. Model dynamics differ somewhat across the two solutions, indicating that the dependence structure between marginals and the joint distribution evolves in a nontrivial way in the model. For QE shocks, output responses are very similar,

[84] In the numerical implementation I follow Bayer and Luetticke (2020) and store the joint distribution on a discrete grid as a function of the marginal CDFs rather than storing the pure copula density in (141). The steady state joint CDF is computed by cumulative sums over the full (b, a, s) grid and reinterpreted on the axes given by the marginal CDFs. A coarse grid in (u_b, u_a, u_s) is constructed from these steady state marginals, and all copula objects are represented on this copula grid.

[85] Bayer and Luetticke (2020) show that in their two asset HANK application a fixed copula already produces aggregate dynamics and cross sectional distributions that are very close to those implied by the full Reiter solution, and that allowing for a small number of time-varying copula coefficients mainly matters in experiments where shocks move the joint distribution of assets and income strongly, for example with shocks to idiosyncratic income risk.

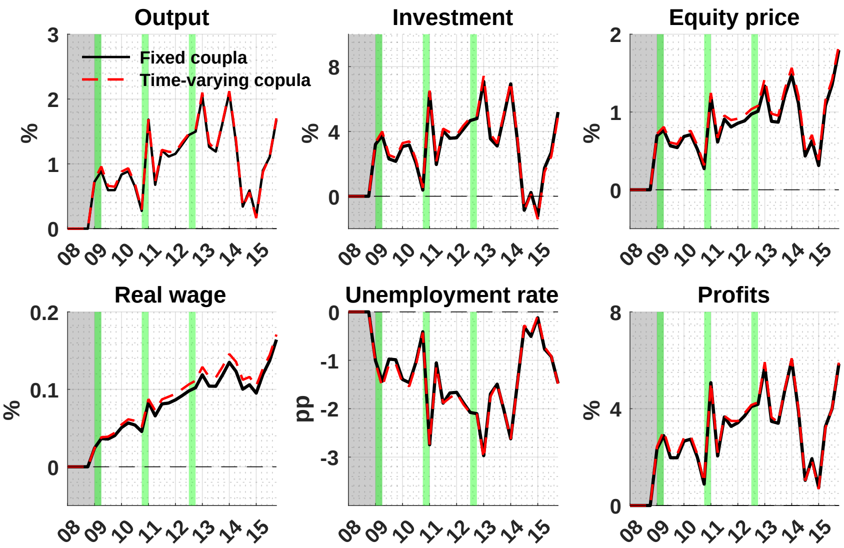
Figure OA38: Effects of UMP shocks: fixed vs time-varying copula



Notes: The figure shows impulse responses of output to QE and FG shocks in the HANK model under two solution methods. The solid line plots the responses under the baseline solution with a fixed copula, and the blue line with circles plots the responses under the solution with a time-varying copula.

with the time-varying copula delivering slightly smaller effects over the horizon. For FG shocks, by contrast, the time-varying copula implies noticeably larger effects on output on impact. The larger differences for FG suggest that allowing the copula to vary over time helps capture more accurately the dynamics of liquid asset holders, who are particularly sensitive to changes in the expected path of interest rates.

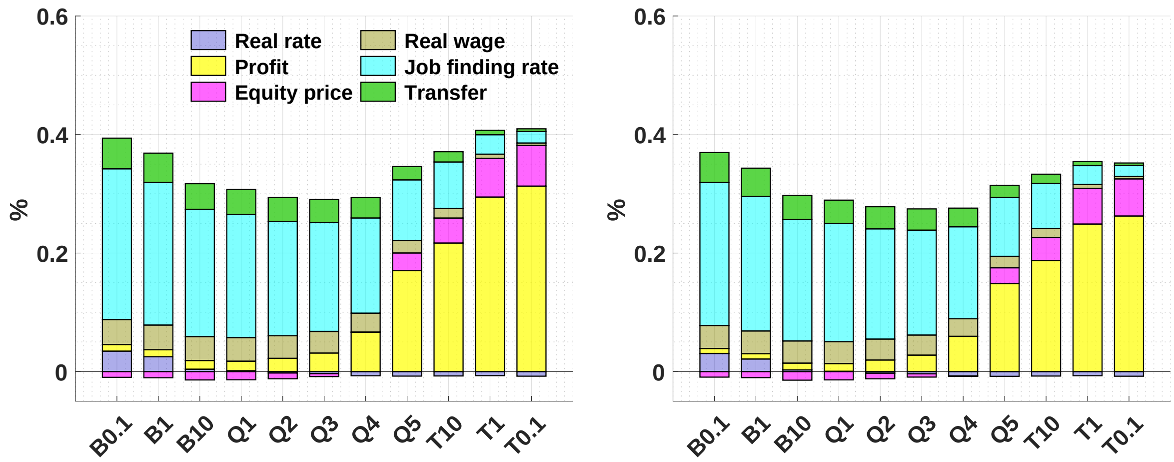
Figure OA39: Aggregate effects of UMP in the HANK model with a time-varying copula



Notes: The figure shows the aggregate effects of UMP in the HANK version of the model with a fixed copula (solid black) and a time-varying copula (dotted red). For level variables X, each panel reports $100 \times (X^{\text{UMP}}/X^{\text{NoUMP}} - 1)$ in percent. For the unemployment rate, panels report percentage point differences $u^{\text{UMP}} - u^{\text{NoUMP}}$.

When I use the posterior mode from the estimation with the fixed copula and simulate the model with a time-varying copula, the implied effects of UMP during the ELB episode are slightly larger, as shown in Figure OA39, although the quantitative differences are small. These differences in aggregate effects carry over to welfare, with somewhat larger overall gains and more pronounced U-shaped distributional effects in the time-varying copula case, while the qualitative conclusions remain the same. Because model dynamics depend on the copula specification, I also re-estimate the model using the time-varying copula solution.[86] Table OA10 reports the estimation results. Relative to the fixed copula estimation, the time-varying copula solution implies a slightly flatter Phillips curve, higher wage rigidity, and a lower vacancy posting cost. Using the posterior mode from the time-varying copula estimation, I then examine aggregate and welfare effects of UMP. Qualitatively, the results are unchanged relative to the fixed copula solution, as shown in Figure OA41 and Figure OA42.

Figure OA40: Welfare effects of UMP: fixed vs time-varying copula



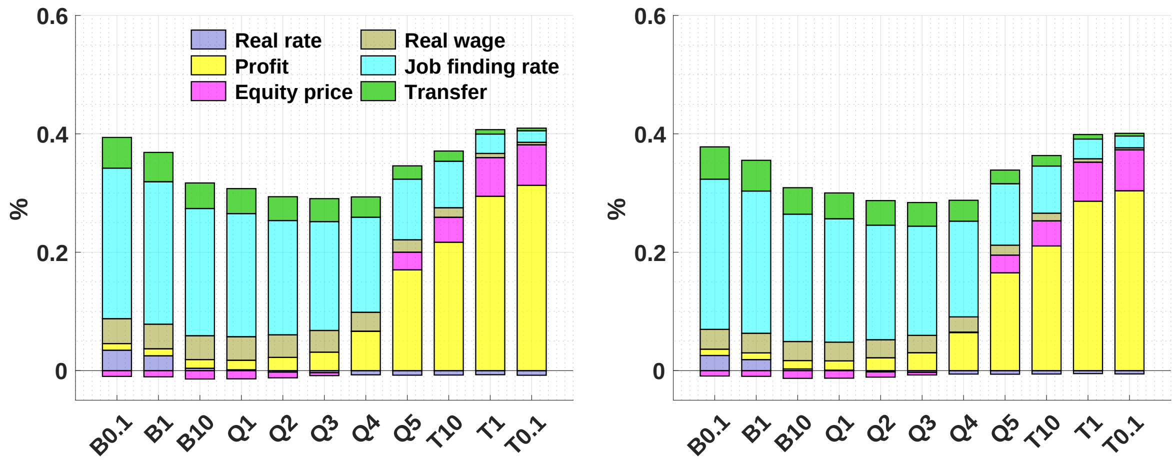
Notes: The figure shows welfare gains from UMP, measured in consumption-equivalent units at 2009Q1. The left panel reports results for the time-varying copula solution, and the right panel reports results for the fixed copula solution. Units are percent, so positive values indicate that the baseline with UMP is preferred to the counterfactual without UMP.

Figure OA41: Aggregate effects of UMP in the HANK model with a time-varying copula

Notes: The figure shows the aggregate effects of UMP in the HANK version of the model with a fixed copula (solid black) and a time-varying copula (dotted red). For the time-varying copula, the posterior mode from the estimation with the time-varying copula solution is used. For level variables X, each panel reports $100 \times (X^{\mathrm{UMP}}/X^{\mathrm{NoUMP}} - 1)$ in percent. For the unemployment rate, panels report percentage point differences $u^{\mathrm{UMP}} - u^{\mathrm{NoUMP}}$.

[86] For estimation speed, I fix expected ELB durations at the posterior mode from the estimation with the fixed copula.

Figure OA42: Welfare effects of UMP: time-varying copula with different modes



Notes: The figure shows welfare gains from UMP, measured in consumption equivalent units at 2009Q1. The left panel reports results from the posterior mode of the estimation with a fixed copula, and the right panel reports results from the posterior mode of the estimation with a time-varying copula. Units are percent, so positive values indicate that the baseline with UMP is preferred to the counterfactual without UMP.

Table OA10: Prior and posterior distributions of structural parameters

Symbol	Description	Prior Density	Prior Mean	Prior Std	Posterior Mode	Posterior 10%	Posterior 90%
		Frictions					
κ	Slope of Phillips curve	Gamma	0.10	0.02	0.0474	0.0341	0.0647
ι_p	Price indexation	Gamma	0.50	0.15	0.1779	1.0237	0.2655
ρ_w	Wage autocorrelation	Beta	0.50	0.20	0.8131	0.6827	0.8864
ι_w	Wage indexation	Beta	0.50	0.15	0.1894	0.1306	0.2846
ϕ	Capital adjustment cost	Normal	30.00	5.00	48.873	47.572	50.196
ι	Vacancy posting cost	Gamma	0.05	0.02	0.0251	0.0163	0.0460
		Government policy					
ρ_B	Bond issuance rule	Beta	0.50	0.20	0.4881	0.3487	0.6304
ρ_g	Lump sum transfer shock AR	Beta	0.50	0.20	0.9593	0.9384	0.9879
σ_G	Lump sum transfer shock std dev	Inverse-Gamma	0.10	2.00	0.0696	0.0631	0.0755
ρ_R	Interest rate smoothing	Beta	0.50	0.20	0.7893	0.7560	0.8179
σ_R	Interest rate shock std dev	Inverse-Gamma	0.10	2.00	0.1645	0.1492	0.1945
ϕ_π	Taylor rule inflation gap response	Normal	1.70	0.30	1.3062	1.1769	1.4543
ϕ_y	Taylor rule unemployment gap response	Gamma	0.10	0.05	0.3790	0.3222	0.4106
		Structural Shocks					
ρ_l	Liquidity preference shock AR	Beta	0.50	0.20	0.9990	0.9993	0.9999
σ_l	Liquidity preference shock std dev	Inverse-Gamma	0.10	2.00	0.0587	0.0537	0.0673
ρ_z	TFP shock AR	Beta	0.50	0.20	0.9948	0.9919	0.9959
σ_z	TFP shock std dev	Inverse-Gamma	0.10	2.00	0.5683	0.5281	0.6260
ρ_p	Price markup shock AR	Beta	0.50	0.20	0.9529	0.9358	0.9615
σ_p	Price markup shock std dev	Inverse-Gamma	0.10	2.00	1.7970	1.5169	2.2913
ρ_k	Investment shock AR	Beta	0.50	0.20	0.9839	0.9660	0.9915
σ_k	Investment shock std dev	Inverse-Gamma	0.10	2.00	0.0709	0.0656	0.0782
ρ_b	Risk premium shock AR	Beta	0.50	0.20	0.9932	0.9814	0.9962
σ_b	Risk premium shock std dev	Inverse-Gamma	0.10	2.00	0.1386	0.1288	0.1639
ρ_Ξ	Fixed cost shock AR	Beta	0.50	0.20	0.9614	0.9437	0.9718
σ_Ξ	Fixed cost shock std dev	Inverse-Gamma	0.10	2.00	0.8756	0.8277	0.9896
σ_w	Wage shock std dev	Inverse-Gamma	0.10	2.00	0.9562	0.4504	1.6640

Notes The values for the standard deviations and the measurement error are multiplied by 100.