STAFF REPORTS

NO. 1108
JULY 2024

REVISED
OCTOBER 2024

Unconventional Monetary Policies and Inequality

Donggyu Lee

FEDERAL RESERVE BANK of NEW YORK

Unconventional Monetary Policies and Inequality
Donggyu Lee
Federal Reserve Bank of New York Staff Reports, no. 1108
July 2024; revised October 2024
https://doi.org/10.59576/sr.1108

Abstract

This paper examines the effects of unconventional monetary policies on household welfare across the wealth distribution following the Great Recession. Using a heterogeneous agent New Keynesian model, estimated with Bayesian methods, I analyze how forward guidance and quantitative easing affected inequality during this period. The findings show that while these policies boosted economic activity and benefited all households, they had non-linear distributional effects. Unconventional monetary policies reduced inequality within the bottom 90 percent by lowering unemployment but widened the income gap between the top 10 percent and the rest by raising profits and equity prices. Additionally, I find that forward guidance amplified both the aggregate and distributional effects of asset purchase programs. In comparison to conventional monetary policy, interest rate policy would have had more adverse distributional effects than quantitative easing if the policy rate had not been constrained by the zero lower bound.

JEL classification: E12, E30, E52, E58
Key words: unconventional monetary policy, inequality, Heterogeneous Agent New Keynesian model, quantitative easing, Bayesian estimation, zero lower bound

---

Lee: Federal Reserve Bank of New York (email: donggyu.lee@ny.frb.org). This paper was previously circulated under the title "Quantitative Easing and Inequality" and was part of the author's doctoral dissertation at the University of Maryland, College Park. The author thanks his advisors Borağan Aruoba, Luminita Stevens, Thomas Drechsel, and John Shea for their invaluable guidance and support. He also thanks Pierre De Leo, Guido Kuersteiner, Seth Murray, Gonzalo Garcia-Trujillo, Jun Hee Kwak, and Eugene Oue for helpful comments.

1 Introduction

In recent decades, income and wealth inequality have been increasing in the United States, motivating the use of heterogeneous-agent macroeconomic models, in which the propagation of aggregate fluctuations and the effectiveness of policy interventions are evaluated within frameworks that capture the large degree of household inequality present in the data. A particularly lively debate has centered on the distributional consequences of monetary policy.[1] While much of the literature focuses on conventional monetary policy, much less has been established about how unconventional monetary policies (UMP hereinafter), such as forward guidance and quantitative easing (QE hereinafter), affects welfare across the wealth distribution. Though effective at stimulating aggregate economic activity, UMP, especially QE, launched by the Federal Reserve in the aftermath of the Great Recession have often been criticized for exacerbating already wide disparities in income and wealth among U.S. households.[2] Yet, whether or how UMP raise inequality remains an open question and a topic of heated debate.

Gauging the distributional effects of UMP is a challenging task, as various forces compete in determining its net effects on inequality. First, unconventional monetary policies, such as QE, can exacerbate income and wealth inequality by raising profits and asset prices. Since stocks and equity, i.e., claims for profits, are mainly held by the top of the wealth distribution, QE might disproportionately benefit that part of the distribution. Conversely, QE can reduce inequality by lowering the unemployment rate, which mainly benefits the bottom of the wealth and income distributions, or by stimulating wage growth, boosting income shares in the middle of the distribution.[3] Finally, higher inflation induced by, for instance, forward guidance re-distributes wealth from savers to debtors by lowering real rates. A proper evaluation of the net effect of UMP on inequality needs to take into account these channels comprehensively.[4]

This paper provides a structural evaluation of the aggregate and distributional effects of UMP using a medium-scale HANK model designed to capture and quantify the dynamics of the channels mentioned above. For a model to effectively perform this task, it must meet two key requirements. First, it should properly capture heterogeneity in household wealth and income composition. Second, it must produce empirically plausible responses for vari-

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

[2] See, for instance, Schwartz (2013) and Cohan (2014).

[3] Heathcote et al. (2010) show that earnings at the bottom of the income distribution are mainly affected by changes in the unemployment rate and hours worked while earnings at the top of the income distribution are mainly affected by changes in hourly wage. Thus, as long as monetary policy has stronger effects on unemployment rates than on real wages, it can reduce income inequality.

[4] For a more detailed discussion on the relevant channels, see Coibion et al. (2017) and Amaral (2017).

ables that impact household wealth and income, such as profits, asset prices, wages, and unemployment, to monetary policy shocks. The interplay between these two factors, i.e., wealth/income components and their responses to monetary policy, will determine the distributional effects of monetary policy. Existing models in the literature are often inadequate for studying distributional effects, as they frequently omit crucial channels, such as portfolio heterogeneity or the unemployment margin, and fail to generate convincing responses of profits to monetary policy.[5] The model developed in this paper addresses these shortcomings by meeting both of the aforementioned requirements.

To meet the first requirement, the model incorporates portfolio choice and endogenous unemployment. Households can hold two types of assets (deposits and equity), and their employment status (employed or unemployed) varies endogenously over time. Following Bayer et al. (2019), I introduce an additional working status where households receive a fraction of profits as income without supplying labor. These features result in a heterogeneous composition of wealth (deposits and equity) and income (labor, assets, and business income) among households, as well as varied exposures to unemployment risk. In the steady state, the top 10% of wealthy households hold about 70% of total wealth, primarily in the form of equity, with business and asset income accounting for about 50% of their total income, consistent with U.S. data. In contrast, households in the lower 80% of the wealth distribution rely predominantly on labor income, with a larger proportion of those at the bottom of the distribution being unemployed and thus more vulnerable to unemployment risk.[6]

Regarding the second requirement, I first address a well-known problem of New Keynesian models, namely the counter-cyclical response of profits to monetary policy shocks, which is inconsistent with the empirical evidence.[7] Fixing this problem is crucial, given the importance of profits for wealthy households. For this purpose, I assume a substantial share of fixed cost in production, inspired by Anderson et al. (2018), in addition to wage rigidity and an extensive margin of labor supply.[8] These features help the model generate a

[5]See Broer et al. (2019) for a discussion on counter-cyclical responses of profits to monetary policy shocks in New Keynesian models.

[6]According to the 2007 SCF data, labor income (wages and salaries) constitutes about 80% of the total income for the bottom 80% of the wealth distribution, with the remainder largely consisting of transfer income. In stark contrast, for the top 0.1% of wealthy households, labor income makes up only 16%, and transfer income accounts for less than 1% of their total income. The remaining 85% primarily comes from profit-related sources, such as business income and dividends. For the top 10% of households, labor income comprises about 50%.

[7]In the appendix, I provide empirical evidence on the responses of profits, wages, and unempoyment rates to monetary policy shocks, using a structural VAR model. See also Christiano et al. (2005), Coibion et al. (2017), and Lenza and Slacalek (2018) for further evidence.

[8]Anderson et al. (2018) shows, using microdata on the retail sector, that net operating profit margins are strongly procyclical while gross margins, which are proportional to the inverse of the real marginal cost, are mildly procyclical or acyclical over the business cycle. They interpret that their results suggest the presence of sizeable fixed costs.

procyclical profit response to monetary policy shocks.

I then estimate the model using Bayesian methods to identify the shock processes that pushed the economy to the effective lower bound (ELB hereinafter) and discipline the parameter values that determine the model's dynamic responses to monetary policy. Importantly, I explicitly take into account the binding ELB constraint and Fed's UMP between 2009 and 2015 in the estimation. In the model, QE and forward guidance take the form of the central bank's direct private asset purchase, as in Gertler and Karadi (2011), and the extended periods of expected zero policy rates. By transforming the demand for a non-productive asset (government bonds or deposits) into a productive asset (capital) and creating inflationary pressures, UMP increased aggregate demand and offset the contractionary effects of the binding ELB, which mimic those of a series of contractionary interest rate shocks. At the posterior mode, the model suggests that, between 2009 and 2015, UMP on average led to 3.3% higher profits, 0.9% higher equity prices, a 1.5% lower unemployment rate, but only a 0.1% increase in real wages, compared to a counterfactual of no policy intervention.

Together with heterogeneity in households' wealth/income composition and exposure to unemployment risk, these aggregate effects generated non-linear distributional effects. The top decile's income shares increased by 0.17 percentage point during the ELB episode, mainly because of higher profits and equity prices. However, at the same time, QE reduced the income Gini indices by 0.04 percentage point on average by lowering the unemployment rate. As to welfare gains, stimulative effects of QE improved welfare for all households, and the average welfare gain was equivalent to 0.27 percent of lifetime consumption. However, welfare gains were U-shaped. QE benefited households at both ends of the wealth distribution more than the middle class. The bottom and the top decile (1%) enjoyed gains of about 0.3% (0.33%). Conversely, the welfare gain of the middle 60% was about 0.26% in terms of consumption equivalents.

Comparing different types of unconventional and conventional monetary policies, I find that QE had less adverse effects on inequality than conventional monetary policy. If the Federal Reserve had been able to conduct conventional monetary policy according to an estimated Taylor rule, only the bottom 1% and the top 10% would have enjoyed higher welfare gains. Lower real rates primarily benefit low-wealth debtors while hurting savers. However, the financial sector also benefits from lower financing costs, boosting the net worth of banks and indirectly increasing profits for the wealthy households. Consequently, under conventional monetary policy, households in the top decile experience higher welfare gains despite the negative impact of lower real rates. The model suggests that lower real rates have a stronger positive effect on banks compared to QE, which partly crowds out bank investment and reduces profitability. Lastly, the decomposition exercise shows that expansionary forward guidance and extended periods of zero interest rates amplified both the aggregate and distributional effects of QE. The model estimates that the Fed maintained rates at zero longer than fundamentals alone would suggest, contributing to 55% of the total stimulus from unconventional policies. However, forward guidance also intensified QE's adverse effects on inequality, increasing the top 10%'s income share by an additional 0.09 percentage points.

Related Literature

This paper contributes to the empirical literature on the distributional consequences of both unconventional and conventional monetary policies, where findings are mixed. For conventional policy, Coibion et al. (2017), Mumtaz and Theophilopoulou (2017), and Furceri et al. (2018), among others, find that tightening increases income inequality, while other work, such as Inui et al. (2017), Davtyan (2017), and Hafemann et al. (2018), reports neutral or even opposite effects. Regarding unconventional policies like QE, findings on the Federal Reserve's asset purchase program also conflict. For instance, Bivens (2015) argues that QE reduced inequality by boosting employment, while Montecino and Epstein (2015) finds it worsened inequality through higher equity returns. In Europe, Casiraghi et al. (2018) and Lenza and Slacalek (2018) report that ECB QE programs reduced inequality, while Bank of England (2012) and Domanski et al. (2016) claim QE increased wealth inequality. My paper adds to this debate by offering a unified framework that evaluates the aggregate and distributional effects of both forward guidance and QE, and compares them with conventional monetary policy.

Theoretically, this paper builds on the so-called HANK literature, with a particular focus on the distributional consequences of monetary policy. While much of the existing work, including Kaplan et al. (2018), Luetticke (2020), and Auclert et al. (2020b), focuses on how heterogeneity or inequality influences the aggregate dynamics of macroeconomic models or the transmission mechanisms of monetary policy, this paper emphasizes the distributional impacts. To provide a fair and comprehensive evaluation, the model developed here incorporates both a two-asset structure and frictional labor markets, making it more suitable for studying distributional effects than existing models. For example, the models in Kaplan et al. (2018) and Bayer et al. (2020) include a two-asset structure but cannot capture the effects of monetary policy on labor market adjustments. On the other hand, while Gornemann et al. (2016) introduces labor market frictions to study the distributional consequences of conventional monetary policy, their model features only one asset, making it unable to capture the positive effects of expansionary monetary policy on wealthy households through higher equity prices and returns. In contrast, the model in this paper effectively captures the impact of monetary policy on both employment and asset markets, providing insights into its effects across the wealth distribution.

Importantly, my model addresses the issue of counter-cyclical profit movements in New Keynesian models both qualitatively and quantitatively. Previous studies, like Broer et al. (2019) and Auclert et al. (2023), have acknowledged this problem, often proposing wage rigidity as a solution. While this can induce pro-cyclical profit responses, the magnitude is typically smaller than observed in the data.[9] In contrast, my model includes features such as search-and-matching labor market frictions, wage rigidity, fixed costs, and a financial sector, all of which result in a strong pro-cyclical profit response to monetary policy shocks, aligning with real-world data. This allows the model to accurately capture the benefits of expansionary monetary policy for wealthy households, whose income depends heavily on profits.

Lastly, my work also contributes to the literature on estimating HANK models. Bayer et al. (2020) extend the work of Smets and Wouters (2007) and Justiniano et al. (2011) and estimate a HANK model using Bayesian techniques to study the drivers of inequality in the U.S. during the post-war period. Auclert et al. (2020b) also estimate their HANK model to discipline the parameter governing the degree of sticky expectation in their model using the aggregate data of the U.S. economy. My contribution lies in estimating a HANK model that incorporates an occasionally binding constraint alongside UMP. To the best of my knowledge, this paper is the first to estimate a HANK model with an occasionally binding ELB constraint.

The remainder of the paper proceeds as follows. Section 2 describes the model. Section 3 explains the parametrization and estimation strategy, and presents the estimation results. Section 4 conducts counterfactual experiments to examine the aggregate and distributional effects of UMP during the ELB episode. Section 5 compares QE and forward guidance. Section 6 compares QE and conventional monetary policy. Section 7 concludes.

2 Model

The model introduces financial intermediaries, the ELB, and QE in the form of central bank asset purchases, into a medium-scale DSGE model with heterogeneous households, uninsurable income risk, aggregate uncertainty, and a two-asset structure. The household block mostly follows the HANK models of Kaplan et al. (2018) and Bayer et al. (2019), while the modeling of financial intermediaries and QE draws on the work of Gertler and Karadi (2011). On the supply side, I incorporate frictional labor markets, a fixed cost of production,

[9] See Lee (2019) for more on addressing counter-cyclical profit issues in New Keynesian models.

and wage and price rigidities.

2.1 Household

There is a unit mass of households who are ex-ante identical but ex-post heterogeneous due to the evolution of their idiosyncratic productivity s, holdings of illiquid and liquid assets, a and b, and employment status e. In each period, households can be employed, unemployed, or business owners, with transitions occurring either exogenously or endogenously, as I will explain below.

Households derive utility from consumption and disutility from labor, and die with exogenous probability ζ each period.[10] Surviving households maximize their lifetime utility with a discount rate $\beta \in (0, 1)$:

$$\max_{\{a_{it+1}, b_{it+1}, c_{it}, n_{it}\}} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t (1-\zeta)^t \left\{ u(c_{it}, n_{it} | s_{it}, e_{it}) - \chi_{it} \mathbb{1}_{\{a_{it+1} \neq a_{it}\}} \right\} \right] \ , \tag{1}$$

where c_{it} is consumption, n_{it} is labor supply, a_{it+1} is illiquid asset holding, and b_{it+1} is liquid asset holding; s_{it} and e_{it} are idiosyncratic productivity and employment status, respectively; and χ_{it} is the stochastic disutility incurred when the household adjusts its portfolio of illiquid holdings. The period utility function has the specification of Greenwood et al. (1988),

$$u(c_{it}, n_{it}) = \frac{\left[c_{it} - \psi s_{it} \frac{n_{it}^{1+\xi}}{1+\xi} \right]^{1-\sigma}}{1-\sigma} \ , \tag{2}$$

where σ is the inverse of the elasticity of inter-temporal substitution (IES), ξ is the inverse of the Frisch elasticity, and ψ is a scale parameter. As in Bayer et al. (2019), this specification implies uniform hours worked among employed households, simplifying computation.

Households optimally choose consumption, hours worked, and portfolio composition subject to the following budget constraint and borrowing limits:

$$c_{it} + q_t a_{it+1} + b_{it+1} = (1-\tau) y_{it} + (q_t + r_t^a) a_{it} + (1 + r_t^b) b_{it} + T_t \ , \tag{3}$$

$$a_{it+1} \geq 0 \ , \quad b_{it+1} \geq \underline{\mathrm{b}} \ , \tag{4}$$

where q_t is the price of illiquid assets, r_t^a is its dividend rate, and r_t^b is the net real rate of

[10] As in Kaplan et al. (2018), I assume that upon death, a household is replaced by a new household with zero wealth and the wealth of the deceased is redistributed to surviving households in proportion to their asset holdings. Stochastic death helps the model generate a substantial mass of households with zero assets at the steady state.

return on liquid assets. T_t is the lump-sum transfer from the government and the money market mutual fund. The tax rate on households' income is denoted by τ. The period income y_{it} depends on the household's working status,

$$y_{it} = \begin{cases} w_t s_{it} n_{it} & \text{for employed} \qquad (e_{it} = 1) \\ w v \min\{s_{it}, s\} & \text{for unemployed} \qquad (e_{it} = 2) \\ \nu \Pi_t & \text{for business owners} \qquad (e_{it} = 3) \end{cases} \tag{5}$$

where w_t is the real wage per efficiency unit, and w its steady state value. Employed households earn wage income that is proportional to their productivity, which evolves according to a first-order Markov process. If unemployed, households receive unemployment benefits equal to a fraction of their steady state labor income, based on their current productivity but capped by the average productivity s, with the replacement ratio v. If households become business owners, they receive a fraction of profits as income.

Working status evolves as follows: at the beginning of the period, an employed household becomes unemployed with an exogenous separation rate λ, while business owners lose their ownership state with an exogenous probability $\tilde{P}^e$ and also become unemployed. The newly unemployed households search for jobs along with previously unemployed households. The job finding rate f_t is determined endogenously, based on the aggregate state of the economy. At the end of the period, a fraction P_e of non-business owners become business owners.[11]

Lastly, households transfer wealth inter-temporally via two assets. Liquid assets b_t are subject to an exogenous borrowing limit $\underline{\text{b}}$, and pay a real rate r_t^b that depends on whether the household borrows or saves:

$$r_t^b = \begin{cases} \frac{1+i_t}{\pi_t} & \text{if } b_{it} \geq 0 \\ \frac{1+i_t+i}{\pi_t} & \text{if } b_{it} < 0 \end{cases}, \tag{6}$$

where i_t is the nominal interest rate, π_t is the gross inflation rate, and i is the nominal borrowing premium. Illiquid assets a_t earn a return of r_t^a, which is proportional to profits in the economy. Adjusting illiquid asset holdings incurs a stochastic utility adjustment cost.[12] Following Bayer et al. (2019), the independently and identically distributed adjustment costs

[11] The introduction of business owners helps the model match the overall wealth inequality in the data as they are the highest income groups in the model.

[12] A stochastic adjustment cost preserves the global concavity of the household's value function, facilitating the computation of households' optimal policies.

are drawn from a logistic distribution, with cumulative probability

$$F(\chi_t) = \frac{1}{1 + \exp\left\{-\frac{\chi_t - \mu_\chi}{\sigma_\chi}\right\}} \quad , \tag{7}$$

where μ_x and σ_x are the location and the scale parameter of the logistic distribution.

2.2 Final good firm

The final good is a standard CES aggregator,

$$Y_t = \left[\int Y_{jt}^{\frac{\eta_t - 1}{\eta_t}} \, dj\right]^{\frac{\eta_t}{\eta_t - 1}} \quad , \tag{8}$$

where Y_{jt} is firm j's intermediate good, and η_t is the time-varying elasticity of substitution. Profit maximization yields individual demand and the associated aggregate price index,

$$Y_{jt} = \left(\frac{P_{jt}}{P_t}\right)^{-\eta_t} Y_t \tag{9}$$

$$P_t = \int P_{jt}^{1-\eta_t} dj \tag{10}$$

where P_{jt} is good j's price.

2.3 Intermediate goods firms

There is a continuum of intermediate good firms that produce differentiated products using labor and capital rental services, according to the following production function:

$$Y_{j,t} = Z_t K_{j,t}^{\theta} L_{j,t}^{1-\theta} \quad , \tag{11}$$

where $K_{j,t}$ and $L_{j,t}$ are capital and labor rental services, respectively, Z_t is total factor productivity, and θ is the share of capital in production.

Each firm maximizes the following expected present discounted value of future profits subject to its demand (9) and the production function (11).

$$\max_{\{P_{jt}, L_{jt}, K_{jt}\}} \sum_{t=0}^{\infty} \mathbb{E}_0\left[\Lambda_{0,t} \Pi_t^I\right] \quad , \quad \Pi_{jt}^I = P_{jt} Y_{jt} / P_t - r_t^l L_{jt} - r_t^k K_{jt} - \Phi^P(P_{jt}, P_{jt-1}) - \Psi_t^F Y \quad , \tag{12}$$

where r_t^l is the labor rental rate, r_t^k is the capital rental rate, and $\Psi_t^F Y$ is the fixed cost of operation. The fixed cost, following an AR(1) process, is a random proportion of steady-state output and help the model generate procyclical profit responses to demand shocks. The firm's discount factor $\Lambda_{t,t+1}$ is the average marginal rate of subsitution of business owners. Regarding price setting, firms are subject to price rigidity a lá Rotemberg (1982). Specifically, price adjustment cost is given by

$$\Phi^P(P_{jt}, P_{jt-1}) = \frac{\eta_t}{2\kappa}\left(\log \frac{P_{jt}}{P_{jt-1}} - \log \pi_{t-1}^{\iota_p} \pi^{1-\iota_p}\right)^2 Y_t \quad , \tag{13}$$

where κ is the slope of the Phillips curve and ι_p is the degree of backward looking price-setting behavior in an equivalent Calvo price-setting setup.[13]

Under symmetric equilibrium assumption, the firm sector can be summarized by the following Phillips curve.

$$\log \pi_t - \log \pi_{t-1}^{\iota_p} \pi^{1-\iota_p} = \mathbb{E}_t\left[\Lambda_{t,t+1}\left(\log \pi_{t+1} - \log \pi_t^{\iota_p} \pi^{1-\iota_p}\right)\right] + \kappa\left(\mathrm{MC}_t - \frac{1}{\Psi_t^p}\right), \tag{14}$$

where $\mathrm{MC}_t = \frac{(r_t^k)^\theta (r_t^l)^{1-\theta}}{Z_t}\left(\frac{1}{\theta}\right)^\theta \left(\frac{1}{1-\theta}\right)^{1-\theta}$ is the real marginal cost and $\Psi_t^p = \frac{\eta_t}{\eta_t - 1}$ is the price mark-up shock.

2.4 Labor agencies

Labor agencies work as an intermediary between households and intermediate good firms. They post vacancies to hire households and provide labor services to firms. A household can supply labor only via a labor agency.

A labor agency that is matched to a household i earns the margin between the labor rental rate that the intermediate good firms pay, and the wage paid to the household.

$$(r_t^l - w_t - \Xi^L) s_{it} n_{it} \tag{15}$$

where Ξ^L is the cost for maintaining a match.[14]

For the determination of the real wage w_t, I follow Gornemann et al. (2016) and assume

[13] In the equivalent Calvo pricing model, ι_p denotes the degree of indexation to the previous inflation rate when firms are not allowed to adjust their price.

[14] The cost Ξ^L is introduced only to enable the estimation of the vacancy posting cost. I adjust Ξ^L to make sure that the expected value of a matching equals the vacancy posting cost in the estimatil.

a function of the form.[15]

$$\frac{w_t}{w} = \left\{\epsilon_{w,t}\left(\frac{r_t^l}{r^l}\right)\right\}^{(1-\rho_w)} \times \left\{\frac{w_{t-1}}{w} \times \left(\frac{\pi_{t-1}}{\pi_t}\right)^{\iota_w} \times \left(\frac{\pi}{\pi_t}\right)^{1-\iota_w}\right\}^{\rho_w} \quad , \quad 0 < \rho_w \ , \ \iota_w < 1 \tag{16}$$

Equation (16) implies a wage determination mechanism that is similar to Calvo wage setting. First, a fraction ρ_w of the wage is subject to nominal wage rigidity.[16] Specifically, the fraction ι_w of this part of the wage adjusts based on the previous inflation rate π_{t-1} while the fraction $1-\iota_w$ adjusts based on the steady state inflation rate.[17] The remaining fraction $1-\rho_w$ varies with the labor rental rate r_t^l. The responsiveness of the real wage to its rental rate can change due to an exogenous shock $\epsilon_{w,t}$ that follows an i.i.d. process.

In a given period, a match between a household and a labor agency ends in the following three cases: (i) if a matched household dies (probability ζ), (ii) if the match is exogenously dissolved (probability λ) or (iii) if a matched household becomes a business owner (probability P^e). Given the termination probability, a labor agency's value is given by

$$J^L(s_{it}) = (r_t^l - w_t - \Xi^L)s_{it}n_{it} + \mathbb{E}\Big[\Lambda_{t,t+1}(1-\zeta)(1-\lambda)(1-P_e)J^L(s_{it+1})\Big] \quad , \tag{17}$$

where $\Lambda_{t,t+1}$ is the same discount factor used by intermediate goods firms, i.e., the average MRS of business owners.

The total number of vacancies is determined by the free-entry condition,

$$\iota = \frac{M_t}{V_t}\int J^L(s_t)d\mu_t(s_{it}) \quad , \tag{18}$$

where ι is the vacancy posting cost, V_t is the total number of vacancies, and $\mu_t(s_{it})$ is the household distribution over idiosyncratic productivity.

Finally, to determine the number of matches, I follow den Haan et al. (2000) and use the

[15] In principle, one would need to solve a bargaining problem to find the equilibrium wage that applies to a match between an agency and a household. However, since each household's outside option depends not only on their idiosyncratic productivity but also on their asset holding and the level of adjustment costs, the equilibrium wage can differ at each point in the idiosyncratic state space. This feature of the model makes computing wages as a solution to a bargaining problem computationally demanding. However, there exists a set of wages that support an equilibrium, and a given wage function can support an equilibrium as long as the wage given by the function belongs to such a set. Under the parameterizations and the simulations examined in this paper, the wages implied by the wage function always remain in the bargaining set. Thus, maintaining a match is always beneficial for both labor agencies and households.

[16] A difference is that in Calvo setting a wage setter expects the possibility that the wage cannot be adjusted in the future. That is, Calvo wage setting is forward-looking. The wage function used in this paper does not feature forward looking behavior.

[17] One can interpret ι_w as the degree of indexation to the previous inflation rate in a Calvo sticky wage model.

following matching function

$$M_t = \frac{\left(U_t + \lambda N_t\right)V_t}{\left\{\left(U_t + \lambda N_t\right)^{\alpha} + V_t^{\alpha}\right\}^{\frac{1}{\alpha}}} \quad , \quad \alpha > 0 \quad , \tag{19}$$

where U_t is the mass of unemployed households at the beginning of period t, N_t is the mass of employed households at the beginning of period t, and V_t is the total number of vacancies.[18] The parameter α determines the efficiency of matching process in the model. The job-finding rate is determined by $\frac{M_t}{U_t+\lambda N_t}$.

2.5 Capital firm

A representative capital firm determines the capital utilization rate and accumulates capital as demanded by investors, i.e., households and banks.[19] For a given capital stock K_t, the capital firm earns the following profit

$$r_t^k v_t K_t - \delta(v_t) K_t \quad , \tag{20}$$

where v_t and $\delta(\cdot)$ are the variable utilization rate and the variable depreciation rate. The first-order condition associated with capital utilization implies that the capital rental rate is equal to the marginal increase in the variable depreciation rate. That is,

$$r_t^k = \delta'(v_t) \tag{21}$$

For variable depreciation, I use a standard functional form used in Greenwood et al. (1988),

$$\delta(v_t) = \delta_0 v_t^{\delta_1} \quad , \quad \delta_1 > 1 \ , \tag{22}$$

where δ_0 is the depreciation rate under full utilization and δ_1 governs the degree of acceleration of depreciation.

Regarding capital accumulation, I assume that the capital firm purchases new capital from its investment department on behalf of investors.[20] The investment department has

[18] Note that, since a certain fraction of households belong to the business owners group, the sum of the masses of unemployed and employed household is not equal to 1.

[19] To simplify price determination, I assume that the capital accumulation is determined entirely by the demand side. This assumption implies that the capital firm does not solve the dynamic problem associated with capital accumulation.

[20] For simplicity, I assume that these two entities operate independently from each other. Thus, one does

a technology that can convert a unit of the final good to a unit of new capital subject to capital adjustment costs. Specifically, it makes profits as follows.

$$q_t K_{t+1} - \Psi_t^k \left\{ K_{t+1} + \frac{\phi}{2} \left(\log \frac{K_{t+1}}{K_t} \right)^2 K_{t+1} \right\}, \tag{23}$$

where Ψ_t^k is a shock to the efficiency of the capital production. In the sense that the shock affects the price of capital and the efficiency of capital transformation technology, it resembles an investment specific technology shock or the marginal efficiency of investment (MEI) shock in Justiniano et al. (2011).

With the assumption that the investment department discounts the future profits with the average MRS of business owners, one can derive the price of new capital as follows.

$$q_t = \Psi_t^k \left\{ 1 + \phi \log \frac{K_{t+1}}{K_t} + \frac{\phi}{2} \left(\log \frac{K_{t+1}}{K_t} \right)^2 \right\} - \mathbb{E}_t \left[\Lambda_{t,t+1} \Psi_{t+1}^k \phi \left(\log \frac{K_{t+2}}{K_{t+1}} \right) \frac{K_{t+2}}{K_{t+1}} \right] \tag{24}$$

Finally, investment expenditure is defined as

$$\tilde{I}_t = \Psi_t^k K_{t+1} \left\{ 1 + \frac{\phi}{2} \left(\log \frac{K_{t+1}}{K_t} \right)^2 \right\} - \{1 - \delta(v_t)\} K_t. \tag{25}$$

2.6 Equity mutual fund

There exists a hypothetical mutual fund that owns all firms in the model. To distinguish it from the other type of mutual fund that I will introduce below, I call it the equity mutual fund. The roles of the equity mutual fund include collecting profits from firms, paying out dividends to shareholders, and issuing new equity for capital accumulation. I assume that the fund operates in a perfectly competitive environment. Thus, there are no retained earnings, and the fund pays out all profits as dividends. The funds acquired by issuing equity are transferred to the capital firm for the purchase of capital. The period cash-flow constraint of the equity mutual fund is as follows:

$$(1 - \tau)(1 - \nu)\Pi_t - q_t(K_{t+1} - K_t) + q_t(A_{t+1} - A_t) = r_t^a A_t \quad , \tag{26}$$

where Π_t is the sum of all firms' profits and ν is the share of profits that is given to business owners.[21] The tax rate on firms' profits is denoted by τ. Given that the amount of aggregate

not take into account the effects of its own decision on the other.

[21] I assume that the fund itself is owned by business owners, and thus a fraction of profits is distributed to them regardless of their equity holding.

capital is equal to the amount of equity in the model, the price of equity is equal to the price of new capital, and the dividend rate is

$$r_t^a = (1-\tau)(1-\nu)\Pi_t/K_t \ , \tag{27}$$

namely, the dividend rate is profits net of tax payments and net of the amount given to business owners, divided by total equity.

2.7 Banks

The model features a financial sector as in Gertler and Karadi (2011), where a continuum of banks, indexed by $j \in (0,1)$, take deposits and purchase equity. The balance sheet for bank j is given by

$$q_t A_{jt+1}^b = N_{jt} + B_{jt+1}^b \quad , \tag{28}$$

where A_{jt+1}^b and B_{jt+1}^b are the bank's equity holding and deposits at the end of period t, respectively. The bank's net worth at the beginning of period t is denoted by N_{jt}, which evolves as follows.

$$N_{jt+1} = R_{t+1}^a q_t A_{jt+1}^b - R_{t+1} B_{jt+1}^b \quad , \tag{29}$$

where $R_{t+1}^a = (q_{t+1} + r_{t+1}^a)/q_t$ and $R_{t+1} = 1 + r_{t+1}^b$ are the gross real rate of return on illiquid and liquid assets, respectively.

Each period, only a θ_b fraction of banks continue to operate, while the rest exiting the market. Let $J^b(N_{jt})$ denote the value of a surviving bank. Under the environment described so far, the value of bank j is given by

$$J^b(N_{jt}) = \max_{\{A_{jt+1}^b, B_{jt+1}^b, N_{jt+1}\}} \mathbb{E}_t \Big[\Psi_t^b \Lambda_{t,t+1} \Big\{ (1-\theta_b) N_{jt+1} + \theta_b J^b(N_{jt+1}) \Big\} \Big] \tag{30}$$

s.t.

$$q_t A_{jt+1}^b = N_{jt} + B_{jt+1}^b \quad , \quad N_{jt+1} = R_{t+1}^a q_t A_{jt+1}^b - R_{t+1} B_{jt+1}^b \tag{31}$$

$$J^b(N_{jt}) \geq \Delta q_t A_{jt+1}^b \tag{32}$$

where Ψ_t^b denotes the aggregate risk premium shock, which follows an AR(1) process. As shown in (32), banks face an incentive compatibility constraint, meaning they can purchase equity only up to the point where the bank's value exceeds a specific fraction of its asset value.[22] In the model, the constraint always binds, leading to

$$q_t A_{t+1}^b = \Theta_t N_t \tag{33}$$

where A_{t+1}^b and N_t are the financial sector's equity holding and net-worth, respectively. The leverave ration Θ_t is given by

$$\Theta_t = \frac{\vartheta_t^n}{\Delta - \vartheta_t^a}, \tag{34}$$

where ϑ_t^n and ϑ_t^a are the marginal value of net-worth and equity, respectively. Since banks are owned by the equity mutual fund, their profits are distributed to equity holders and business owners as part of aggregate profits.

2.8 Money market mutual fund

The model includes a hypothetical mutual fund, referred to as the money market mutual fund, which primarily serves to provide liquidity to the financial sector.[23] The fund receives contributions from the government and invests in liquid assets. With these contributions, along with proceeds from its investments, the fund makes lump-sum transfers to households. I also assume that the fund smooths the flow of lump-sum transfers.

$$\max_{\{T_t^m, B_{t+1}^m\}} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \Psi_t^l \beta_m^t \frac{(T_t^m)^{1-\sigma}}{1-\sigma} \right] \tag{35}$$

subject to

$$T_t^m + B_{t+1}^m = C_t^g + (1 + r_t^b) B_t^m \ , \tag{36}$$

where T_t^m and B_{t+1}^m are the fund's lump-sum transfer and liquid asset holding, respectively. The MMMF's IES is denoted by σ.[24] The contribution that the fund receives from the

[22] See the appendix for a detailed description on the associated moral hazard problem.

[23] Since banks are leveraged investors, there must be an equivalent amount of liquid assets corresponding to their illiquid asset holdings in the model. If households were the sole providers of funds to banks, the share of liquid assets in household portfolios would need to be high. However, data from the SCF shows that households hold only about 10% of their total assets as liquid assets. Additionally, according to Financial Accounts data (formerly the Flow of Funds), the share of household liquid assets (e.g., checkable and time deposits, and corporate bonds) in the domestic financial sector's liabilities—which include deposits, bonds, open market paper, loans, and other liabilities—has remained around 25% since 2000. Based on these facts, I assume there is a significant non-household liquidity provider, represented by the money market mutual fund.

[24] In principle, the MMMF's IES does not need to equal to the household's IES. However, to save the notation, I assume that MMMF and households have the same IES.

government is denoted by C_t^g. Unlike any other entities in the model, I assume that the fund discounts future lump-sum transfer flows with its own discount factor β_m.[25] Finally, the MMMF is subject to an AR(1) liquidity preference shock Ψ_t^l.

2.9 Monetary authority

The monetary authority sets its policy rate according to a Taylor rule with interest rate smoothing.

$$i_{t+1} = \min\{0, \hat{i}_{t+1}\} \quad \text{with} \quad \frac{1+\hat{i}_{t+1}}{1+\hat{i}} = \left(\frac{1+\hat{i}_t}{1+\hat{i}}\right)^{\rho_R} \left[\left(\frac{\pi_t}{\pi}\right)^{\phi_\pi} \exp\{-\phi_u(u_t - u)\}\right]^{1-\rho_R} \exp(\epsilon_{R,t}) \ , \tag{37}$$

where $\epsilon_{R,t} \sim N(0, \sigma_R^2)$ is a monetary policy shock, $0 < \rho_R < 1$ is the degree of interest rate smoothing, and $\hat{i}_{t+1}$ and i_{t+1} are the shadow and actual policy rates, respectively. The responsiveness of the nominal rate to inflation and the unemployment gap are denoted by ϕ_π and ϕ_u. Note that the actual policy rate is constrained by the ELB. Thus, if the shadow rate becomes negative, the policy rate can no longer respond to inflation and the unemployment gap.[26] However, in the model, the central bank can affect the economy even when the ELB binds, through asset purchases and forward guidance.

Regarding asset purchase programs, I follow Gertler and Karadi (2011) and assume that the central bank directly purchases illiquid assets from the private sector. Specifically, the central bank issues bonds and sell them to the private sector and, with proceedings, purchases equity.

$$A_{t+1}^{\mathrm{CB}} = \Psi_t^{\mathrm{QE}} A^{\mathrm{CB}} \quad , \quad B_{t+1}^{\mathrm{CB}} = q_t A_{t+1}^{\mathrm{CB}} \ , \tag{38}$$

where A_{t+1}^{CB} is the central bank's illiquid asset holding at the end of period t and Ψ_t^{QE} is an AR(1) QE shock that determines the amount of asset purchases as a fraction of the central bank's steady state asset holding. B_{t+1}^{CB} denotes bonds issued by monetary authority in period t. From its asset holdings, the central bank earns cash flows, i.e., dividend income net of interest payments. The central bank remits all of its proceedings to the fiscal authority.

In addition to QE, the central bank can implement forward guidance in the form of

[25] The steady state optimality condition of the MMMF requires the MMMF's discount factor to be the inverse of the steady state real interest rate. However, because of idosyncratic income risks, the average of business owners' MRS is not equal to the inverse of the real rate at the steady state.

[26] The ELB for the policy rate does not need to be zero. In practice, several countries' central banks set negative policy rates. However, in the U.S., the Federal Reserve never set negative policy rates. In this paper, I assume that the ELB is zero.

exogenous expected ELB durations if the economy is at the ELB.[27] That is, by assumption, the central bank can determine households' and firms' expectations regarding the number of periods during which the central bank would maintain the policy rate at zero. If the exogenous expected ELB duration is longer than the endogenous ELB duration (the number of periods during which the ELB constraint is expected to bind based on the aggregate state of the economy), it is equivalent to agents expecting future negative (expansionary) interest shocks. Hence, via inter-temporal substitution, forward guidance also can stimulate economic activity.

2.10 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 the following simple autoregressive rule, as in Woodford (1995).

$$\frac{B^g_{t+1}}{B^g} = \left(\frac{R_t/\pi_t \times B^g_t}{R/\pi \times B^g}\right)^{\rho_B} \quad , \; 0 \leq \rho_B < 1 \quad , \tag{39}$$

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

Since economic agents in the model form rational expectations, the government should meet the following inter-temporal budget constraint.

$$B^g_t = \sum_{l \geq t}^{\infty} \left\{ \prod_{i=t}^{l} \left(\frac{\pi_i}{R_i}\right) \right\} \left\{ \mathbb{T}_l - (G_l + T^g_l + D_l + T^{CB}_l + C^g_l) \right\} \quad , \tag{40}$$

where $\mathbb{T}$, G, T^g, D, and C^g are tax revenues, government purchases, lump-sum transfers (or taxes) to households, unemployment benefits, and contributions to the MMMF, respectively. $T^{\mathrm{CB}}_t = q_t A^{\mathrm{CB}}_{t+1} - (q_t + r^a_t) A^{\mathrm{CB}}_t + R_t B^{\mathrm{CB}}_t - B^{\mathrm{CB}}_{t+1}$ is the transfer from (or to) the monetary authority.

Equation (40) implies that in each period, the debt level must be equal to the present discounted value of all future government surpluses. When the real value of government debt changes, at least one fiscal instrument must adjust to meet the solvency condition. In this paper, I assume that the fiscal authority adjusts its contribution to the MMMF to balance the budget, while government purchases are fixed and lump-sum transfer to households varies

[27] The model's solution at the ELB is computed backward using the method of Kulish et al. (2014) and Jones (2017), which requires an expected duration of the binding ELB constraint as a part of the solution. The methodology is similar to the OccBin method of Guerrieri and Iacoviello (2015), but allows the duration of the temporary regime to be exogenous.

according to the following stochastic process.[28]

$$T_t^g = \left(1 - \frac{1}{\Psi_t^g}\right) Y \ , \tag{41}$$

where Ψ_t^g is a lump-sum transfer shock and Y is the steady state output.

2.11 Market clearing conditions

To close the model, I state the market clearing conditions for each market. The equity market clearing condition is

$$\underbrace{\underbrace{A_{t+1}^h}_{\text{households}} + \underbrace{A_{t+1}^b}_{\text{banks}} + \underbrace{A_{t+1}^{CB}}_{\text{central bank}}}_{\text{equity demand}} = \underbrace{K_{t+1}}_{\text{equity supply}} \ , \tag{42}$$

where $A_{t+1}^h = \int a_{t+1} d\mu_t$ is the aggregate equity demand of households. As shown above, three entities invest in equity: households, banks, and the central bank. The sum of their asset demands should equal the total equity supply, i.e., aggregate capital.

The market-clearing condition for liquid assets, i.e., bonds and deposits, is given as follows.[29]

$$\underbrace{\underbrace{B_{t+1}^h}_{\text{households}} + \underbrace{B_{t+1}^m}_{MMMF}}_{\text{liquid asset demand}} = \underbrace{\underbrace{B_{t+1}^b}_{\text{bank deposits}} + \underbrace{B_{t+1}^g}_{\text{gov. bond}} + \underbrace{B_{t+1}^{QE}}_{\text{bond for QE}}}_{\text{liquid asset supply}} \ , \tag{43}$$

where $B_{t+1}^h = \int b_{t+1} d\mu_t$ is the aggregate liquid asset demand of households. Note that, as households and the money market mutual fund do not distinguish between bonds and deposits, the composition of bank deposits and government bonds in the liquid asset market is determined by the supply side.

Market clearing for capital services implies that the capital stock utilized in the current

[28] Because markets are incomplete and households value liquidity, the model is non-Ricardian. Thus, the fiscal responses matter, especially for the distributional effects of monetary policy. Given that there is only short-term government debt, the effects of these fiscal responses can be particularly strong, as shown in Lee (2019). However, the assumption that I adopt in this paper dampens the effect of the fiscal response. An increase in contributions to the MMMF will increase lump-sum transfers from it, but the responses are modest since I assume that the MMMF smoothes out lump-sum transfer flows.

[29] In the model, agents do not distinguish between bonds and deposits.

period must equal the capital services demanded by the intermediate goods producers:

$$v_t K_t = \int_0^1 K_{j,t} dj = K^I \tag{44}$$

Similarly, the labor supplied by households (via labor agencies) must equal the labor services demanded by the intermediate good firms,

$$\int \mathbb{1}_{\{e_t=1\}} s_t n_t d\mu_t = \int_0^1 L_{j,t} dj\ . \tag{45}$$

If the above-mentioned markets clear, by Walras' law, the goods market also clears.

2.12 Solution method

I solve the model using the perturbation method of Reiter (2009), extended by Winberry (2018) and Bayer and Luetticke (2020). First, I compute the steady state using the endogenous grid method of Carroll (2006), then linearize the model around the steady-state and apply perturbation. Given the high dimensionality due to idiosyncratic states (e.g., asset holdings, skill level, work status), I reduce the state space following Bayer et al. (2019) and Bayer and Luetticke (2020).

To address the occasionally binding constraint on the policy rate, I follow Kulish et al. (2014) and Jones (2017), treating the model with a binding ELB as a temporary alternative regime with exogenous durations. For comparison, I also apply Guerrieri and Iacoviello (2015) for models with endogenous ELB durations. Further numerical details are in the appendix.

3 Parametrization

I adopt a two-stage approach to parameterize the model. First, I set a subset of parameters to ensure that the model matches moments of households' wealth distribution and income composition in the microdata at the steady state. I then estimate the remaining parameters using full-information Bayesian methods, leveraging time-series data on aggregate macro variables to inform the parameters governing the model's dynamics. Importantly, the estimation explicitly accounts for the occurrence of the binding ELB constraint and the implementation of UMP.

3.1 Calibration

For the calibration, data from the Survey of Consumer Finances (SCF) is used, as it provides detailed information on households' wealth and income composition, particularly among wealthy households. Among the different waves, the 2007 data is used as it is the last survey conducted before the Great Recession. In the data, liquid assets are defined as the sum of various types of deposits, bonds, and credit card balances, while illiquid assets include financial assets, net housing wealth (with 40% of housing wealth included, following Kaplan et al. (2018)), business interests, and installment loans. Consumer durables are excluded from illiquid assets, and households with negative illiquid assets are also omitted. Income is categorized into labor, capital, and transfer income. Labor income consists of wages, capital income includes business and asset income, and transfer income covers social security benefits and pensions.

For calibration targets, moments from the household wealth distribution, such as the shares of borrowers and wealthy hand-to-mouth households, as well as households' income composition across wealth groups, are used. As Kaplan et al. (2018) have shown, household wealth distribution matters for the responsiveness of a HANK model to monetary policy shocks. Additionally, households' income composition and the responses of each income component will determine individual households' gains and losses from monetary policy.

Table 1 shows the calibrated parameters. Given the model's many parameters, I will discuss the calibration of a few key ones here and refer readers to the appendix for a more detailed description. Each period in the model corresponds to a quarter. While most household parameters are calibrated exogenously, the discount factor and the scale parameter of labor disutility are calibrated internally. This internal calibration aims to match the proportion of wealthy hand-to-mouth households in the data and ensure that each household's steady-state labor supply is 1. Additionally, the parameters related to illiquid asset adjustment costs are calibrated internally to achieve an average adjustment probability of 6.5% and to ensure that the top 10% of wealthy households in the model hold about 70% of the total illiquid assets.

The income process consists of a standard component, which follows an AR(1) process with the autocorrelation and the standard deviation of 0.98 and 0.02, and two boundary states (super low-skilled and high-skilled) are added to match wealth inequality in the data. Additionally, transition probabilities associated with the business owner state are calibrated to match the share of liquid assets held by the top 10% of wealthy households. The values for idiosyncratic productivity and the state transition matrix for workers and business owners. The values for idiosyncratic productivity and the state transition matrix for workers and business owners are shown in Tables 2 and 3.

Table 1: Calibrated parameters

Parameter	Value	Description	Reference or targets
Households			
σ	1.5	Relative risk aversion	Standard value
β	0.9932	Household's discount factor	Mass of wealthy hand-to-mouth households
ξ	3	Inverse Frisch elasticity	Chetty et al. (2011)
ψ	0.8476	Disutility of labor	SS labor supply of 1
ζ	1/180	Probability of death	Average life span of 45 years
μ_χ	9.0490	Mean of χ dist	SS adj. prob. of 6.5%
σ_χ	3.4205	Scale parameter for χ dist	Top 10% illiquid asset share
P_e	0.05%	Prob. of becoming business owner	Bayer et al. (2019)
$\tilde{P}_e$	20.6%	Prob. of losing business	Top 10% liquid asset share
Labor Market			
λ	0.1	Job separation rate	den Haan et al. (2000)
$\bar{w}$	1.2112	SS real wage	SS labor share to output ratio of 60%
α	1.7127	Matching efficiency	SS vacancy filling rate of 70%
Ξ^L	0.0076	Cost of maintaining a match	SS unemployment rate of 5.5%
Goods producers			
η	3	Elasticity of substitution	Gornemann et al. (2016)
θ	0.27	Exponent of capital in the production function	SS capital share to output ratio of 40%
Ξ/Y	0.2012	Ratio of the fixed cost to output	Capital to output ratio of 3.03
Capital firm			
δ_0	0.0150	SS depreciation rate	SS depreciation rate 6% (annual)
δ_1	1.0025	Elasticity of dep w.r.t. utilization	SS utilization rate of 1
Financial sector			
$\tilde{\Lambda}$	$(1+i)/\bar{\pi}$	MMMF's discount factor	SS optimality condition
τ_m	0.0533	MMMF contribution share to tax revenue	SS lump-sum transfer to output ratio 0.1
Δ	0.3410	Degree of limited enforcement	SS leverage ratio of 3
θ_b	0.97	Bank's survival rate	Gertler and Karadi (2011)
ω	0.0076	Initial net worth of new banks	Banks' equity share of 55%
ν	0.2380	Fraction of profits given to business owners	Gini Net worth
Government			
τ	0.30	Tax rate	Data
υ	0.4	Replacement ratio	Standard value
$\underline{\mathrm{i}}$	0.0253	Borrowing premium	Mass of households with zero assets
$\underline{\mathrm{b}}$	1.3006	Borrowing limit	Mass of households with debt
Central bank			
π	1.0050	Inflation target	Fed's target
$1+i$	1.0100	SS nominal rate	Households' liquid to illiquid asset ratio
A^{CB}/Y	0.05	SS CB's assets to output ratio	Data
ρ^{QE}	0.99	Autocorrelation of QE shocks	See the main text

Table 2: 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 3: 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

Labor market parameters, such as the steady-state real wage and matching efficiency, are calibrated to ensure a steady-state labor share of output of 60%, a vacancy filling rate of 70%, and an unemployment rate of 5.5%. For goods producers, the steady-state elasticity of substitution is set to 3, following Gornemann et al. (2016). A relatively low elasticity of substitution implies a high steady-state markup, allowing for a substantial share of fixed costs in production. The steady-state fixed cost is set to match a capital-to-output ratio of 3.03. Parameters governing capital utilization is set to ensure steady utilization of 1 and depreciation of 6%. The financial sector is parameterized following Gertler and Karadi (2011).

For the government sector, standard values are used for most parameters, including a replacement ratio of 0.4 and a 2% (annualized) inflation target. The steady-state policy rate is internally calibrated to match the households' illiquid asset ratio in the data. Additionally, I assume that the central bank's assets equal 5% of output at the steady state, based on the historical average before the implementation of QE. The auto-correlation of the central bank's assets is set to 0.99.

Table 4 and 5 show the model's performance in matching key moments from the data, with targetd moments highlighted in blue and bold text. As shown in Table 4, the model successfully aligns with data on the mass of households with zero liquid wealth, overall saving, portfolio composition, wealth inequality, and indebtedness, though it overestimates the share of households with zero illiquid assets likely due to the absence of housing. Despite targeting only the top 10% share of each asset type, Table 5 demonstrates that the model does a reasonable job of approximating the overall asset distribution, which is known to be very challenging. Regarding income composition, Figure 1 shows that the model closely matches the data, reflecting trends in capital and labor income across wealth groups. For the bottom 60% of the wealth distribution, labor income makes up about 80% of total income. For the top 0.1% of wealth households, labor income is 16% in both the model and data, while capital income shares are 81% in the model and 83% in the data.

Table 4: Targeted moments and model fit 1

	Data	Model
Capital to output ratio	3.03	3.02
Liquid to illiquid asset ratio	0.10	0.10
Gini net worth	0.82	0.83
Fraction with b < 0	0.14	0.15
Fraction with b = 0 and a > 0	0.20	0.20
Fraction with b = 0 and a =0	0.11	0.10
Fraction with b = 0	0.31	0.30
Fraction with a = 0	0.14	0.26

Data : SCF 2007, NIPA

Table 5: Targeted moments and model fit 2

	Liquid Assets		Illiquid Assets	
Moments	Data	Model	Data	Model
Top 0.1 percent share	19	10	15	3
Top 1 percent share	45	39	38	19
Top 10 percent share	84	84	74	73
Bottom 50 percent share	-4	-3	3	1
Bottom 25 percent share	-5	-3	0.2	0
Gini Coefficient	0.97	0.95	0.82	0.85

Data : SCF 2007, Notes : The blue color indicates targeted moments.,

3.2 Estimation

To discipline the model dynamics with the data, I estimate the remaining model parameters using Bayesian methods. One of the biggest challenge in estimating a HANK model is that it takes a significant amount of time to update the solution for each new set of parameters, due to the large number of idiosyncratic states. To address this, I follow Bayer and Luetticke (2020) and estimate only a subset of parameters that do not affect the steady state objects, such as households' value functions and the time-invariant distribution. This approach allows me to update only a relatively small number of elements in the Jacobian of the linearized system during each iteration, making the estimation process feasible.

Another challenge in this paper's estimation relates to the occasionally binding constraint on the policy rate, specifically the ELB. Since the model's solution depends on the expected duration of the ELB while the economy is at the bound, it is necessary to determine the expected ELB durations during periods when the policy rate is constrained. However, finding ELB durations and filtering out shocks can be challenging and time consuming. To address this challenge, I adopt the approach of Kulish et al. (2014), Jones (2017) and Jones et al.

Figure 1: Income composition in the data and the model

Notes: Labor income in the data is the sum of wage and salary. Capital income is the sum of business income and asset income, which is the sum of interest and dividend income and capital gains. Transfer income includes unemployment benefits, social benefits, e.g., food stamps, and other miscellaneous transfers.

(2018), assuming a sequence of exogenous expected ELB durations during the estimation process.[30] I estimate these durations along with other structural parameters of the model by employing the randomized blocking scheme developed by Chib and Ramamurthy (2010), where structural parameters and expected ELB durations are updated separately.[31]

3.2.1 Observables

For the observables, I use the following set of variables spanning from Q1 1992 to Q4 2018.[32]

$$\left[\Delta\log Y_t, \Delta\log C_t, \Delta\log\tilde{I}_t, \log\pi_t, \log(1+i_t), \log u_t, \Delta\log w_t, \log T_t^g, \log\Pi_t, \log A_t^{\mathrm{CB}}\right], \quad (46)$$

where Y_t, C_t, $\tilde{I}_t$, π_t, $1+i_t$, u_t, w_t, T_t^g, Π_t, and A_t^{CB} correspond to 1) output, 2) consumption, 3) investment, 4) the inflation rate, 5) the nominal interest rate, 6), the real wage, 7) the unemployment rate, 8) lump-sum transfers, 9) corporate profits, and 10) the central bank's assets, respectively. Figure 2 shows the evolution of these observables over the sample period.

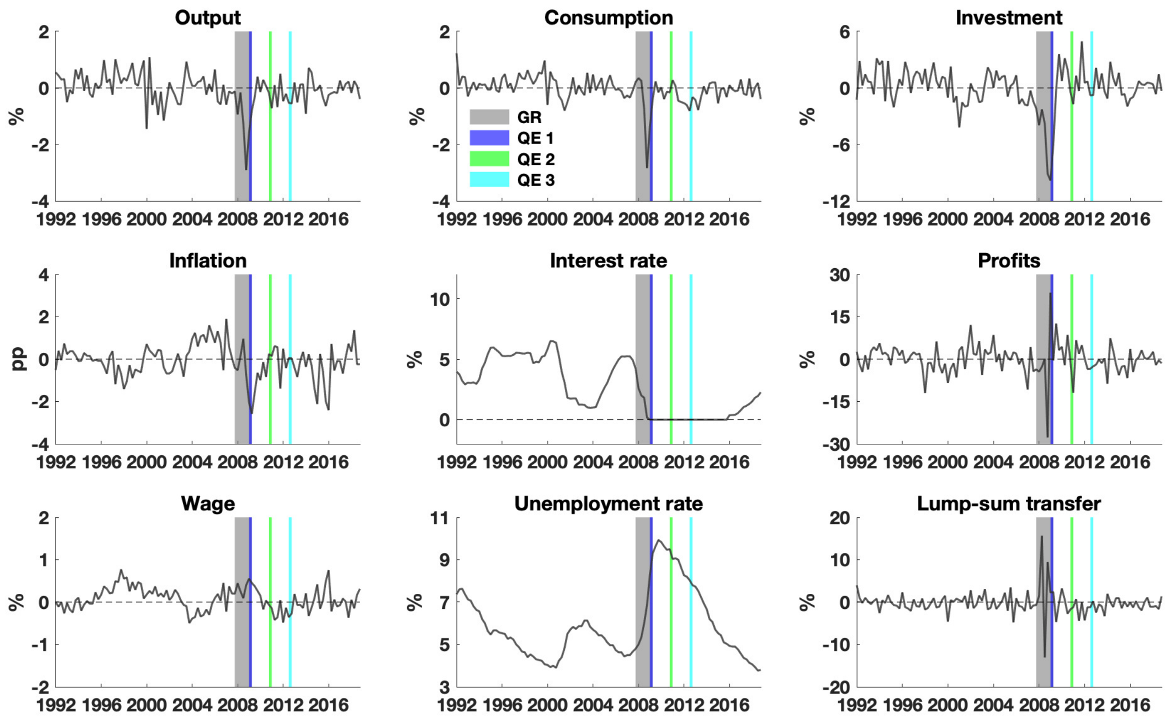
For structural shocks, I assume the following processes: 1) the MMMF's liquidity pref-

[30] Alternatively, one can find the endogenous expected ELB duration in each period during the ELB episode in the estimation as in Guerrieri and Iacoviello (2017), Atkinson et al. (2019), and Cuba-Borda et al. (2019). However, this method is computationally burdensome, as it requires a repeated computation of the inverse of very large matrices.

[31] See the appendix for further details on the estimation procedure.

[32] I use a relatively short sample period to avoid periods with high interest rates.

Figure 2: 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.

erence shock Ψ_t^l, 2) the total factor productivity shock Z_t, 3) the price-mark up shock Ψ_t^p, 4) the wage shock $\epsilon_{w,t}$, 5) the investment technology shock Ψ_t^k, 6) the banks' risk premium shock Ψ_t^b, 7) lump-sum transfer shock Ψ_t^g, 8) monetary policy shock $\epsilon_{R,t}$ 9) the fixed cost shock Ψ_t^F, and 10) the QE shock Ψ_t^{QE}, i.e., the shock to the central bank's asset holding.[33]

[33]For a detailed description of the data, including mnemonic, and a summary of the shock processes, see the appendix.

Table 6: 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 preferences shock stt 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 mark-up shock AR	Beta	0.50	0.20	0.9608	0.9457	0.9720
σ_p	Price mark-up 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.

3.2.2 Prior and posterior distributions

For structural parameters, I follow the literature and use standard priors. For the slope of the Phillips curve, I assume a gamma prior with a mean of 0.1 and a standard deviation of 0.02, corresponding to an average price duration of one year under a Calvo price contract. For inflation indexation in price and wage setting, I use gamma priors with means of 0.50 and standard deviations of 0.15, based on Smets and Wouters (2007) and Justiniano et al. (2011). For capital adjustment frictions, I assume a normal prior with a mean of 30 and standard deviation of 5. The autocorrelation of the real wage follows a beta distribution with a mean of 0.5 and standard deviation of 0.2.

For the policy parameters, I also use fairly standard distributions as priors. For the inflation gap response in the Taylor rule, I assume a normal prior distribution with mean 1.7 and standard deviation 0.2. For the unemployment gap response, I use a gamma prior with mean 0.1 and standard deviation 0.05. The priors for the bond autocorrelation, lump-sum transfer autocorrelation, and interest rate smoothing are set to beta distributions with mean 0.5 and standard deviation 0.2, which is a standard prior used in the literature.

For the shock processes, I use a beta distribution with mean 0.5 and standard deviation 0.2 for the autocorrelations and an inverse-gamma distribution with mean 0.001 and standard deviation 0.02 for the standard deviation of the shock, following Smets and Wouters (2007).

The estimated parameters indicate high wage and price rigidity, low vacancy posting costs, and significant capital adjustment frictions.[34] The posterior distribution for the Phillips curve slope is centered at a low value, suggesting high price rigidity, with an average price duration of six quarters. Wage rigidity is also substantial, with only a fifth of the real wage adjusting to labor productivity changes. In contrast, vacancy posting costs are estimated to be low. Capital adjustment costs are particularly high due to the presence of banks in the model.[35] As shown in the appendix, these parameters generate strong, procyclical responses in profits, equity prices, and unemployment rates, with an almost acyclical real wage response to monetary policy shocks.[36]

[34] The model's high wage rigidity reflects stable real wages and their weak correlation with output, while profits are volatile and positively correlated with output. The model's procyclical profits require low vacancy posting costs.

[35] Following Gertler and Karadi (2011), a financial accelerator applies to banks, leading to high volatility in investment due to changes in equity prices and expected returns. Significant capital adjustment frictions are needed to match this volatility in the data.

[36] See the appendix for impulse response functions.

Table 7: 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.

Table 7 shows the prior and posterior distributions of expected ELB durations. To construct the prior distribution, I use the periodic primary dealer survey of the Federal Reserve Bank of New York, following Jones (2017).[37] In the prior distributions, the mode of the expected ELB duration is notably higher in 2012 and 2013. However, the estimated posterior distributions for expected ELB durations show only slight increases during these periods. Overall, the modes of the estimated expected ELB durations range from 6 to 8 quarters until 2014, falling to one or two quarters in the final year of the ELB episode due to tight priors.

4 UMP during the ELB episode

Now, I turn to the central question of this paper: did UMP conducted by the Federal Reserve raise inequality in the U.S. during the aftermath of the Great Recession? To answer this question, I conduct a counterfactual analysis that compares the economy's actual outcomes (the baseline) to an alternative scenario without UMP, such as QE and forward guidance. As a result of the estimation, the aggregate variables that correspond to the observables exactly follow the data counterparts in the baseline case. In the counterfactual case, the economy still experiences the same shock realizations, but the central bank does not conduct UMP during the ELB episode. Instead, it maintains its asset holdings at their pre-crisis level, and it gives no forward guidance so that the expected ELB duration in each period is endogenously determined solely as a function of the aggregate state of the economy. Furthermore, the central bank adheres to its interest rate rule as soon as fundamentals warrant nominal interest rate liftoff. By comparing these two cases, I gauge the effects of UMP relative to a scenario of a more passive central bank.

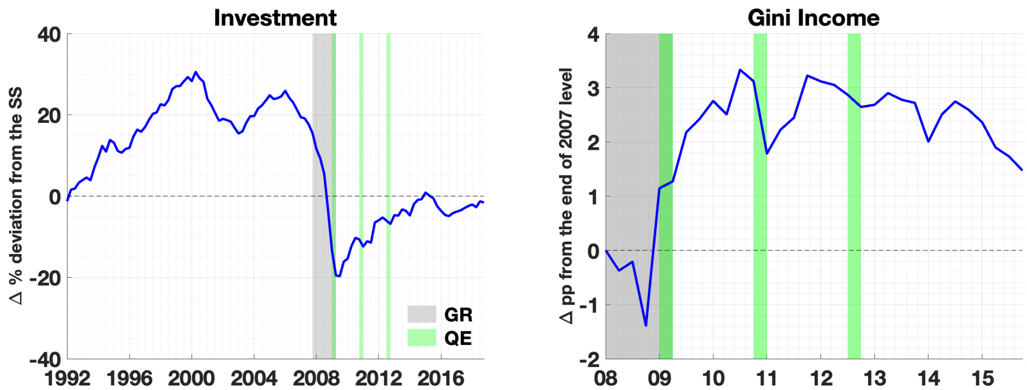
4.1 Great Recession and its aftermath

In the model, the Great Recession was primarily driven by a series of significantly negative risk premium shocks around 2008 and 2009, which led financial institutions to drastically reduce their investments. As a result, the economy experienced a sharp downturn in economic activity, with investment falling by more than 20% relative to its pre-crisis level, as shown in the left panel of Figure 3.[38] In the aftermath, the economy gradually recovered while

[37] In this survey, the New York Fed asks the primary dealers about their expectations for when the Federal Reserve will raise the policy rate above zero. Since survey results prior to January 2011 are unavailable, I use the January 2011 survey for prior distributions covering the period from Q1 2009 to Q4 2010. Some survey results indicate a positive probability of ELB durations exceeding four years. However, for numerical stability, I truncate the prior distributions at a maximum of four years.

[38] Since investment is one of the observables, its dynamics exactly match the data.

Figure 3: Great Recession and the evolution of inequality



Notes: The left panel illustrates the evolution of investment as a percentage deviation from its steady-state level, while the right panel depicts the evolution of the income Gini index at the posterior mode of the parameter distribution as a percentage point deviation from its 2007 year-end level. The grey bar represents the Great Recession, and the green bars represent the periods in which different rounds of QE is implemented.

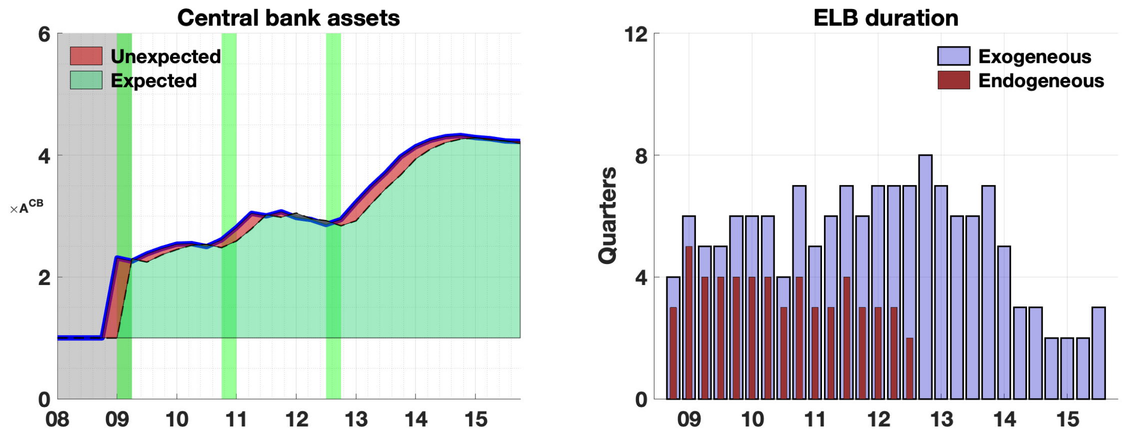
the Federal Reserve implemented a series of asset purchase programs and provides forward guidance on the policy rate. Income inequality remained elevated compared to its pre-crisis level during this period. While the income Gini index initially fell during the Great Recession, it rose as the economy recovered, remaining elevated throughout the sample period. In the following section, I examine the isolated impact of UMP on the evolution of aggregate variables and inequality during this period by conducting counterfactual experiments with parameters, including exogenous expected ELB durations, drawn from their posterior distribution.

4.2 Aggregate effects of UMP

During the aftermath of the Great Recession, the central bank in the model implemented two types of UMP. First, it significantly expanded its balance sheet, as illustrated in the left panel of Figure 4. By transforming the demand for non-productive assets, i.e., bonds and deposits, into the demand for productive assets, i.e., capital, the central bank stimulated economic activity, raising equity prices and boosting investment. Simultaneously, the central bank employed expansionary forward guidance. The right panel of Figure 4 shows the exogenous and endogenous ELB durations at the posterior mode of the estimated parameter distribution.[39] As shown in the figure, at the posterior mode, exogenous expected ELB durations exceeded endogenous ELB durations by 1 to 2 quarters, and at times by up to 8 quarters. Notably, since mid-2012, the policy rate remained at zero, despite economic

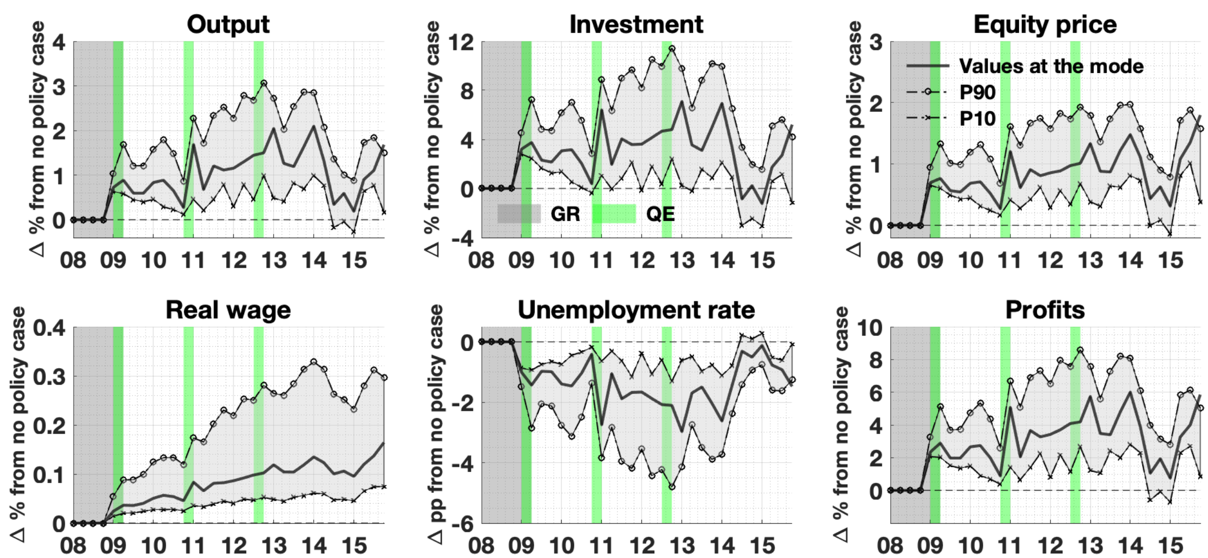
[39]Since inequality measures are not included in the set of observables, their dynamics vary depending parameter values.

Figure 4: UMP in the model



Notes: The left panel shows the evolution of the central bank's assets since the Great Recession as the ratio to its pre-crisis level. The area represents the unexpected evolution of the central bank's balance sheet, while the light green area indicates the expected evolution. The right panel shows both exogenous and endogenous expected ELBU durations since the Great Recession at the posterior mode.

Figure 5: Aggregate effects of UMP

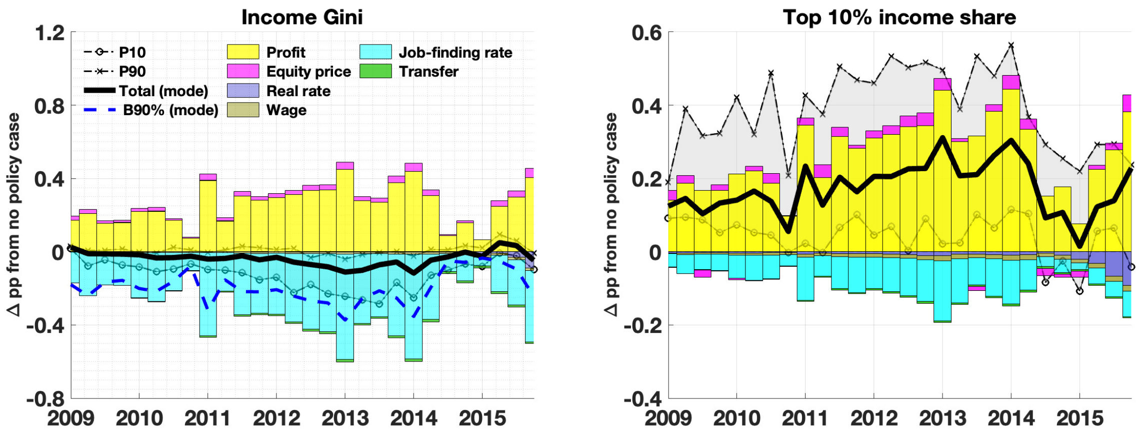


Notes: The variables are presented as percentage differences from their corresponding values in the alternative scenario without policy interventions, except for the unemployment rate, which is shown as percentage point differences from its corresponding value in the alternative case. The black solid lines indicate the effects at the posterior mode, while the shaded areas represent the 10th to 90th percentiles of the effects.

conditions justifying lift-offs of the policy rate into positive territory. This policy stance was equivalent to expected or contemporaneous expansionary interest rate shocks, which spurred economic activity by generating inflationary pressure and reducing the real interest rate.

Figure 5 shows the effects of UMP on aggregate variables. The black solid line represents the effects at the posterior mode, while the shaded area depicts the 10th to 90th percentile

Figure 6: Distributional effects of UMP: income inequality measures



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%, as the percentage point difference compared to the corresponding levels in the counterfactual case. The black solid lines indicate the effects at the posterior mode, while the shaded areas represent the 10th to 90th percentiles of the effects. The dotted blue line represents the differences in income Gini index among the bottom 90% wealthiest households at the posterior mode. The colored bars represent the contributions of different variables to the effects on each inequality measure.

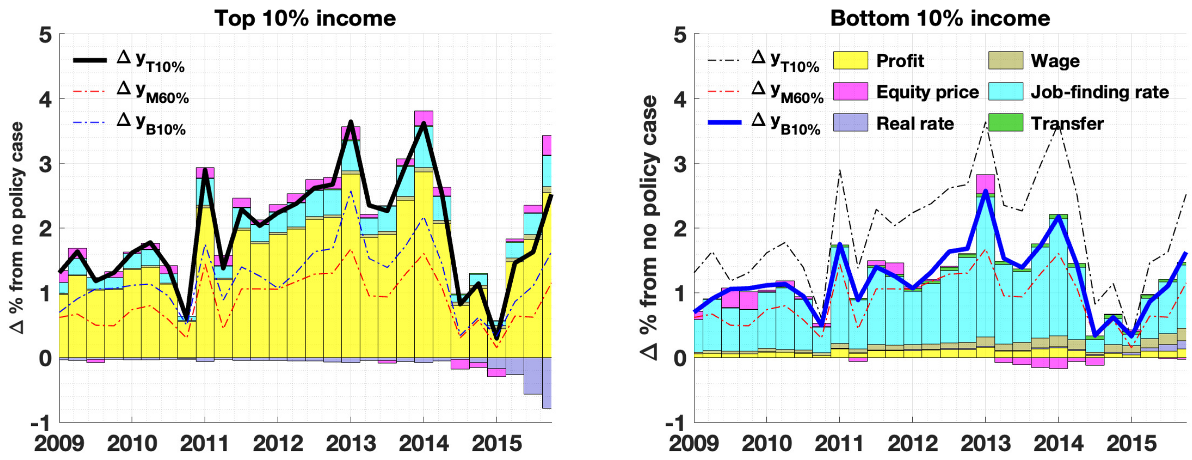
range of the effects, measured as percentage or percentage point differences relative to a counterfactual scenario without UMP. According to the model's simulation results, UMP had a stimulating effect on the aggregate economy during the ELB episode. At the posterior mode, asset purchase programs, combined with expansionary forward guidance helped mitigate the severity of the financial crisis, increasing equity prices and investment by approximately 1% and 3%, respectively, on average between 2009 and 2015. Similarly, output was about 1% higher on average compared to the scenario with no policy interventions. These stimulating effects on aggregate activity led to a notable increase in profits, with gains reaching up to 8% during the ELB period. The labor market also experienced significant improvements, with the unemployment rate being, on average, 1.4 percentage points lower at the posterior mode. In contrast, the differences in real wages were relatively small, with an average increase of only 0.1%, due to high wage rigidity.[40] In the next section, I investigate how these aggregate effects translate into distributional outcomes.

4.3 Distributional effects of UMP: inequality measures

Figure 6 shows the effects of UMP on two measures of income inequality, the Gini index and the top 10% income share, during the ELB episode. Interestingly, these two measures provide different perspectives on how QE and forward guidance affected income inequality

[40]The magnitudes of the effects of UMP in the model fall into the ballpark of the existing estimates found in the literature. See, for instance, Chung et al. (2012), Engen et al. (2015), Kiley (2014), and Rosa (2012).

Figure 7: Distributional effects of UMP: household income



Notes: Both panels show the evolution of the top 10% (black), middle 60% (red), and bottom 10% (blue) wealthiest households' incomes during the ELB episode. Additionally, the left panel illustrates the decomposition of the top 10% income evolution, while the right panel presents the decomposition of the bottom 10% income. The colored bars represent the contributions of different variables to the income evolution for each group.

after 2009. The left panel shows that UMP slightly reduced the income Gini, compared to a counterfactual scenario where the central bank took no action once nominal rates hit the ELB.[41] At the posterior mode, lower unemployment rates due to UMP reduced the Gini index by up to 0.6 percentage points, which aligns with policymakers' arguments that emphasized the positive impact of monetary policy measures on the labor market.[42] However, this reduction was largely offset, by about 80%, by the rise in profits and equity prices, leading to only a marginal improvement in the Gini index.[43]

Notably, the reduction in income inequality was more pronounced when the Gini index was calculated solely for households in the bottom 90% of the wealth distribution. This group's Gini dynamics closely mirrored the effect of higher job-finding rates on overall index, as households in this group rely predominantly on labor income and have similar income compositions. As a result, UMP reduced income inequality, particularly among households in the bottom 90% of the wealth distribution.

While UMP reduced inequality for the bottom 90% of households, it had a contrasting effect on the top 10%, where capital income plays a more significant role. Spefcifically, UMP

[41] At the posterior mode, QE accounts for a -2.5% change in the Gini index during the ELB episode, relative to its level at the start of the Great Recession.

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

[43] Capital gains are excluded from the model's income definition due to the difficulty of tracking the purchase price of illiquid assets for each household. However, higher equity prices still boost income in two ways: first, households can sell fewer shares when prices are higher, retaining a larger equity holding; second, households receive equity holdings from deceased individuals as part of an annuity arrangement, leading to higher income when equity prices rise.

widened the income gap between the top 10% and the rest of the households during the ELB episode, with an average increase of about 0.2 percentage points at the posterior mode. Decomposition results show that this widening was primarily driven by a significant rise in profits and equity prices due to UMP, underscoring the importance of capital income for wealthy households. While lower unemployment rates partially offset these gains, the effect was not large enough to reverse the overall increase in inequality. This outcome highlights the limitations of the Gini index in capturing the benefits of UMP for households at the very top of the wealth distribution.[44]

Having examined the overall effects of UMP on the Gini index and top 10% income share, I now turn to a more detailed breakdown of how these policies affected household income across different wealth brackets. As shown in Figure 7, the top 10% households experienced the largest income gains from UMP, leading to an increase in the top 10% income share. The bottom 10% households saw the second-largest income gains, while the middle 60% experienced the smallest increase in income, resulting in a decrease in the overall Gini index. On average, the top 10% households enjoyed about 2% higher income in the baseline scenario compared to the counterfactual without policy intervention. This increase was primarily driven by higher profits, as shown in the left panel of the figure, again underscoring the importance of business income for wealthy households.

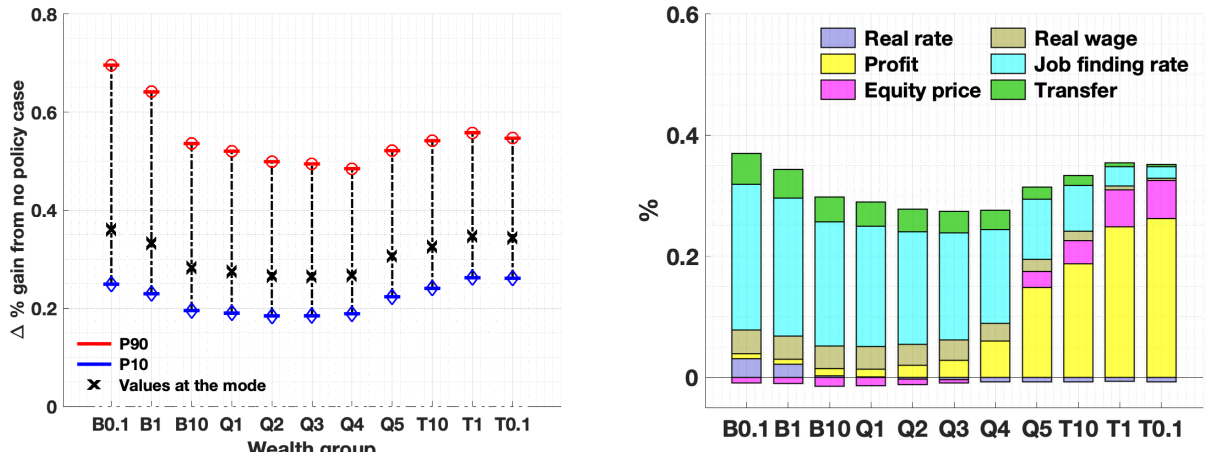
The bottom 10% wealthiest households also saw more than a 1% increase in income when the central bank implemented UMP, with these gains almost entirely due to higher job-finding rates. This is because the household group at the lower end of the wealth distribution hold little wealth and consists of a higher proportion of unemployed households.[45] Households in the middle of the wealth distribution, similarly holding little wealth, were less affected by the job-finding rate increase, as they consist of a higher proportion of already-employed households compared to the bottom 10%. Consequently, this group saw the smallest income gains among all wealth groups.

Thus, while UMP reduced income inequality among the bottom 90%, it simultaneously widened the gap between the top 10% and the rest of the households, reflecting the divergent effects of labor market improvements and capital income on different wealth groups.

[44] It is well known that many inequality measures, including the Gini index, do not guarantee subgroup consistency. For a detailed discussion on the properties of the Gini index, see, for instance, Jurkatis and Strehl (2014).

[45] In the model, at the beginning of Q1 2009, 8.75% of households in the bottom 10% the wealth distribution were unemployed, compared to only 6.54% in the middle quintile. Thus, even though the job-finding rate increased uniformly across the distribution, more households became employed at the bottom than in other groups.

Figure 8: Welfare effects of UMP: Consumption Equivalents



Notes: The figure displays the welfare gains from UMP in terms of consumption equivalents. The left panel shows the range of effects from the 10th to 90th percentile, while the right panel presents a decomposition of these effects at the posterior mode. The colored boxes indicate the contribution of each variable. B0.1 (T0.1), B1 (T1), and B10 (T10) represent the bottom (top) 0.1%, 1%, and 10% of the wealth distribution, respectively, while Q1 to Q5 correspond to the first through fifth quintiles.

4.4 Distributional effects of UMP: welfare gains

In this section, I assess which wealth groups gained the most from UMP during the ELB episode in terms of consumption equivalents, which represent the fraction of lifetime consumption that households in the counterfactual scenario would be willing to forgo to benefit from UMP. I define these wealth groups based on the distribution of wealth in 2009 Q1 and track them throughout the ELB episode, computing their consumption equivalents.[46] By comparing consumption in the baseline and counterfactual scenarios, I calculate the consumption equivalents for each group, representing the fraction of lifetime consumption that households in the counterfactual scenario would be willing to forgo to benefit from UMP.

Figure 8 shows that UMP benefited both wealthy and poor households the most, while providing the smallest welfare gains to the middle quintile.[47] The average welfare gain from UMP was equivalent to 0.27 percent of lifetime consumption at the posterior mode, ranging from about 0.2% to 0.5% depending on the parameter values. However, households at both

[46] Households' wealth distribution in 2009 Q1 is determined in 2008 Q4, and thus, is not affected by UMP. Note also that, since households' wealth and working status vary over time, the composition of wealth groups also changes. Thus, for instance, households in the fifth quintile in 2009 Q1 do not necessarily belong to the fifth quintile in 2013 Q4. In computing the consumption equivalents, I need to follow the same households, and thus fix wealth groups. Also, as the sample ends in 2018 Q4, I assume that there are no shocks beyond that period.

[47] Among different working statuses, business owners saw the highest welfare gains, equivalent to 0.82% of lifetime consumption, followed by the unemployed at 0.35%, and the employed at 0.27% of lifetime consumption at the posterior mode.

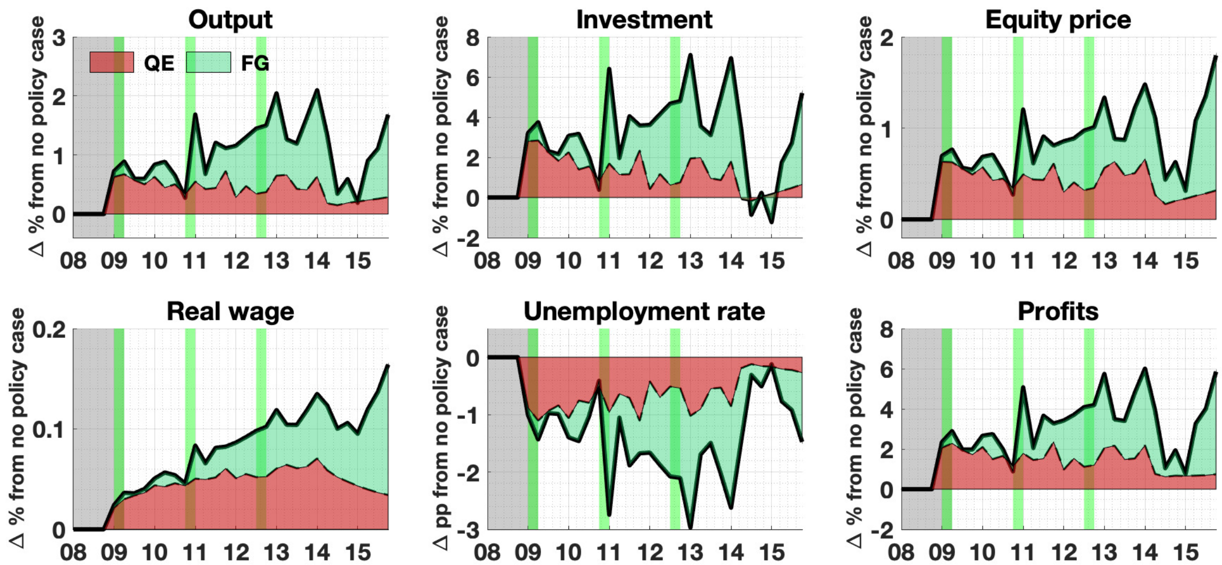
ends of the wealth distribution enjoyed above-average welfare gains, while the middle class saw the least benefit: the consumption equivalent for the bottom and top 1% of households was about 0.06 percentage points higher than that of the middle 60% at the posterior mode. These differences in welfare gains were driven by variations in the share of unemployed households across groups and the composition of their income and wealth. In the model, with parameter values at the posterior mode, the aggregate unemployment rate was 8.27% in 2009 Q1. However, the share of unemployed households in the bottom 10% of the wealth distribution was 8.75%, compared to only 6.54% in the middle quintile. Since a larger proportion of households at the bottom of the wealth distribution were unemployed, they experienced greater welfare gains from the improved job-finding rate. In contrast, households at the top of the wealth distribution, whose income and wealth heavily rely on profits and equity, saw above-average welfare gains due to rising profits and equity prices.

A noteworthy finding is that the differences in welfare gains for the top 10% relative to others are smaller than the differences in income gains. The consumption equivalents for both the bottom and top 10% are similar, with the largest welfare gains observed for the bottom 0.1%. This outcome is due to the anticipated effects of tapering in the periods beyond the sample. During the ELB episode, wealthier households experienced higher consumption gains that mirrored their income gains. However, as the economy moves into the tapering phase, households expect lower equity prices and profits.[48] Lower profits reflect the adverse effects of tapering on banks' net worth, but these are not accompanied by similarly higher unemployment rates. Additionally, tapering creates downward pressure on inflation, which is expected to lead to higher real wages in the future. As a result, the welfare gaps between the top 10% and the bottom 90% are smaller than the income gaps observed during the ELB episode.

Overall, I find that UMP had non-linear distributional effects during the ELB episode. UMP substantially benefited the top 10% of the wealth distribution by boosting profits and equity prices, while also significantly benefiting the bottom 10% by reducing unemployment rates. In contrast, the gains for the middle class were relatively small due to limited changes in real wages. Because of this non-linear effect, UMP can be seen as either increasing or decreasing inequality, depending on the measure of inequality used, such as the Gini index or top income shares.

[48] The evolution of key variables during the tapering phase is shown in Figure A17 in the appendix.

Figure 9: Aggregate effects of QE and forward guidance



Notes: The black solid line represents the effects at the posterior mode, while the red-shaded areas show the contribution of QE alone, and the green-shaded areas indicate the additional effects from forward guidance.

5 QE and forward guidance

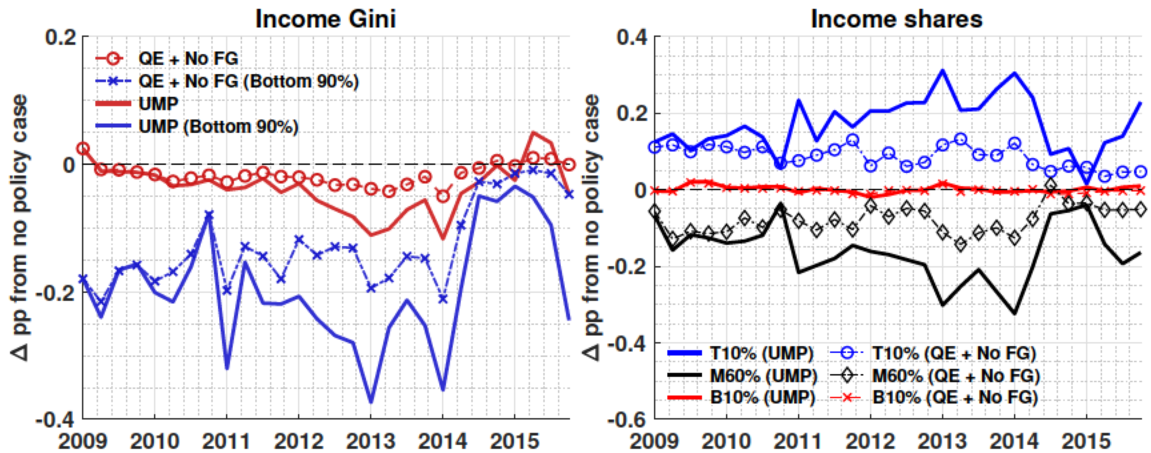
The results presented so far for UMP are driven not only by the central bank's asset purchases but also by the exogenous ELB durations and maintaining the policy rate at zero for longer than would be prescribed by an estimated Taylor rule. In this section, I decompose the total effects of UMP into the separate impacts of QE and forward guidance by simulating the model under the assumption that the expected ELB durations are endogenously determined and that the central bank sets a positive interest rate as soon as prescribed by the Taylor rule.[49] All results presented in the following sections are based on simulations that use the parameter values and exogenous expected ELB durations at the posterior mode of the parameter distribution.

5.1 Aggregate effects of QE and forward guidance

As discussed in the previous section, exogenous expected ELB durations were longer than endogenous durations during the ELB episode, indicating an additional stimulus from forward guidance. The gap between endogenous and exogenous durations began to widen in 2011, which aligns with the findings of Jones (2017). These extended ELB durations acted

[49]The endogenous ELB durations are computed using the OccBin method of Guerrieri and Iacoviello (2015).

Figure 10: Distributional effects of QE and forward guidance: income inequality measures



Notes: The left panel displays the differences in the income Gini index, with the red and blue solid lines representing the differences between the baseline and the counterfactual case with no policy interventions, and the red and blue dotted lines showing the differences between the case with QE and endogenous ELB durations compared to the counterfactual. The right panel shows the differences in income shares relative to the counterfactual.

similarly to anticipated future expansionary monetary policy shocks. Moreover, by keeping policy rates at zero, the central bank continued to provide stimulus to the economy until the end of the ELB episode.

When the additional stimulus from forward guidance was excluded, the aggregate effects of asset purchases alone were much smaller. Overall, the central bank's asset purchases accounted for about 45% of the total aggregate effects of UMP. Specifically, QE had a stronger impact at the beginning of the ELB episode, while forward guidance contributed a larger share of the stimulus in the later periods. This occurred because the central bank's balance sheet changes were largest at the onset of the ELB period, and much of its evolution was anticipated thereafter. Furthermore, as the gap between exogenous and endogenous ELB durations widened after 2011, forward guidance played an increasingly important role. In terms of relative magnitude, forward guidance was comparable to QE, with both having strong effects on profits, equity prices, and the unemployment rate. In summary, forward guidance amplified the aggregate effects of UMP, while maintaining the relative contributions of QE, and played an equally significant role in stimulating the economy.

5.2 Distributional effects of QE and forward guidance

Now I turn to the effects of the additional stimulus from exogenous ELB durations on inequality measures. The left panel of Figure 10 shows that, without additional stimulus from forward guidance, asset purchase programs alone had a smaller impact on the income Gini index during the ELB episode. While asset purchase programs initially reduced the income Gini index among households in the bottom 90% of the wealth distribution, their overall effect on the Gini index was quite small, with only a slight negative average impact. However, as expansionary forward guidance further reduced unemployment rates, the Gini index declined more significantly relative to the counterfactual case with no policy interventions. Similar to the effects of UMP discussed in the previous section, the distributional impacts of forward guidance were more pronounced among households in the bottom 90% of the wealth distribution. While forward guidance lowered the overall Gini index by about 0.05 percentage points around 2013 compared to QE alone, its impact on the Gini index for the bottom 90% was more substantial, reducing it by an additional 0.2 percentage points during the same period.

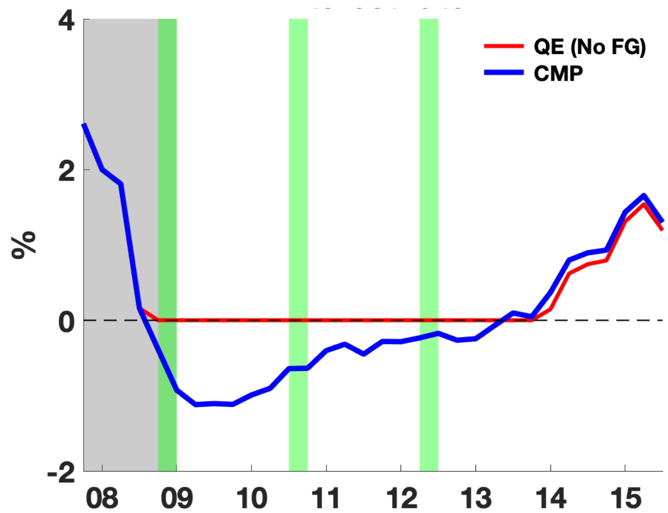
The effects of forward guidance on the top 10% income shares further confirm that the additional stimulus from forward guidance amplified the non-linear distributional effects of QE. As shown in the right panel of Figure 10, asset purchase programs alone increased the top 10% income shares by about 0.1 percentage point, particularly at the onset of the ELB episode, while reducing the middle 60% income shares and leaving the bottom 10% income shares virtually unchanged. The implementation of forward guidance further amplified these distributional effects, leading to an additional increase in the top 10% income shares, particularly between 2011 and 2014, when the gap between exogenous and endogenous expected ELB durations was largest. This occurred because the additional stimulus from forward guidance significantly increased profits and equity prices by about 2% and 1%, respectively, during that period. Simultaneously, forward guidance further reduced the middle 60% income shares while leaving the bottom 10% shares virtually unchanged, intensifying the "hollowing out" of the middle class.

To summarize, the results from this section show that forward guidance amplified both the aggregate and distributional effects of QE, strengthening its non-linear impact on inequality. This highlights the trade-off the central bank faces in stimulating the broader economy. According to the model, forward guidance was effective in boosting economic activity and benefiting vulnerable groups, such as the poor and unemployed. However, in doing so, the central bank also delivered substantial benefits to wealthy households, further concentrating income and wealth.

6 QE and conventional monetary policy

The persistent decline in the natural interest rate in recent decades has raised concerns about the increasing likelihood of ELB episodes going forward. As a result, the literature

Figure 11: The evolution of the policy rate under different scenarios



Notes: The red solid line represents the path of the policy rate in the baseline case with QE alone, while the blue solid line shows the path of the policy rate in the counterfactual scenario where the policy rate is allowed to drop below zero instead of implementing QE and forward guidance.

has started to discuss increasing the inflation target and consequently the steady-state nominal policy rate to provide more room for conventional monetary policy (CMP).[50] In this section, I compare QE and conventional monetary policy in terms of both their aggregate and distributional effects, to provide a reference for the benefit of avoiding the binding ELB constraint. Specifically, I ask what might have occurred if policymakers had been able to lower the policy rate further instead of relying on a package of unconventional policies. To model CMP, I assume that the central bank sets the policy rate according to the Taylor rule, without being constrained by the ELB, and does not implement UMP.[51] In this scenario, the policy rate follows the blue solid line in Figure 11 in the counterfactual case, while the red solid line represents the policy rate in the baseline case.[52]

6.1 Aggregate effects of QE and CMP

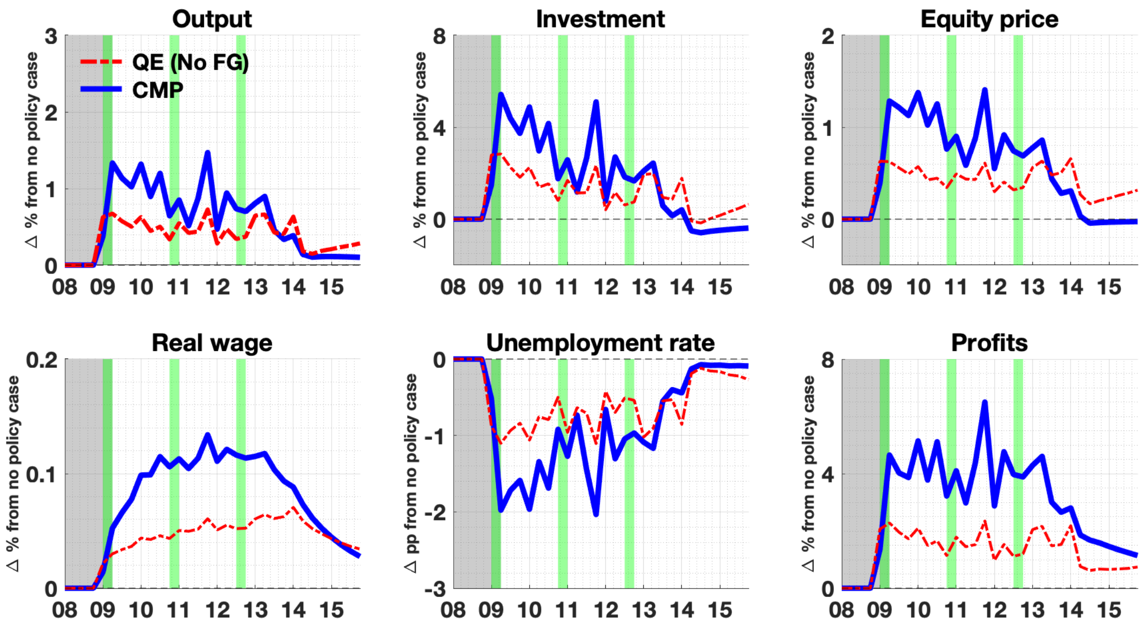
Figure 12 shows the aggregate effects of QE and CMP. As the policy rate fell below zero, reaching almost -1%, and remained negative for an extended period, the economy experienced significant boosts in economic activity. However, as the interest rate gradually rose

[50] See, for instance, Ball (2014),Blanchard et al. (2010), and Williams (2016).

[51] In the simulation, the nominal policy rate drops below zero. However, I do not interpret the results as the effects of negative interest rates, since saving in assets with negative nominal rates can be irrational in practice. Instead, I interpret the results as reflecting CMP with a higher nominal policy rate and inflation rate, where real interest rates remain the same as in the baseline case but the central bank has more room to lower the nominal policy rate.

[52] Since there is no forward guidance in the baseline case, it is also counterfactual.

Figure 12: Aggregate effects of QE and CMP



Notes: The solid red and blue lines represent the effects of QE and CMP, respectively. Except for the unemployment rate, all variables are shown as percentage differences from their corresponding values in the counterfactual scenario with no policy interventions. The unemployment rate is displayed as percentage point differences from its corresponding value in the counterfactual case.

from its low point in 2009 and returned to positive territory around 2014, these stimulating effects diminished. In the case of QE, as discussed in the previous section, there was a substantial stimulus effect in 2009 when the expansion of the central bank's balance sheet was unexpected, but the stimulating effects also gradually decreased over time.

There are two noteworthy findings in the results. First, despite the large volume of the central bank's asset purchases at the beginning of the ELB episode, the magnitude of the stimulus effects was smaller than those of CMP.[53] This is not because CMP had particularly strong stimulus effects.[54] Instead, a substantial portion of QE's stimulus effects was offset by general equilibrium responses, such as changes in private investment behavior. Specifically, QE crowded out private investment, particularly by banks. An increase in equity prices boosted banks' net worth but reduced the expected gross rate of return on equity, i.e., banks' profitability, which discouraged their investment. As shown in Figure 13, although QE increased banks's net worth by raising equity prices, it simulatenously discouraged banks'

[53] Note that the central bank's asset purchases in 2009 Q1 were equivalent to about 6.5% of steady-state output in the model. However, the initial impact on output in the model was less than 1% of steady-state output.

[54] As shown in Figure A6, interest rate shocks had a modest stimulus effect, in line with findings in the literature.

Figure 13: Effects of monetary policy on banks: QE vs CMP

Notes: The blue lines show banks' net worth and equity holdings in the case of CMP, relative to their respective levels in the counterfactual scenario with no policy interventions. The black dotted line with circles in the left panel represents the equity holdings of both financial institutions and the central bank, while the red line shows only the banks' equity holdings. In the right panel, the black dotted line represents banks' net worth.

equity investment.[55] In contrast, CMP increased the profitability of banks by lowering their financing costs, thereby crowding in bank investments and leading to greater stimulus effects than QE.

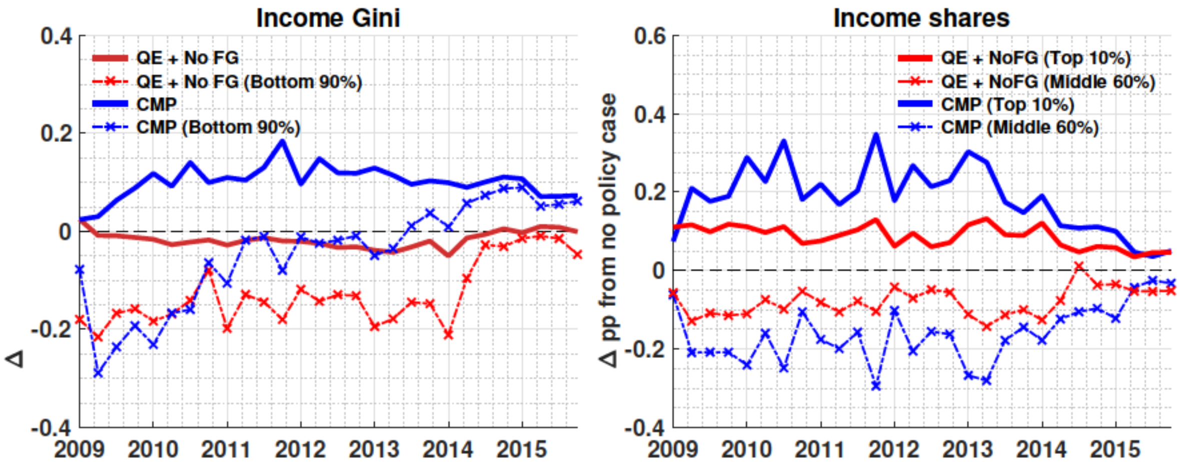
As a result, CMP had particularly strong effects on profits, with the positive impact more than double that of QE.[56] In contrast, CMP's effects on the unemployment rate, real wage, and equity prices were approximately 40% greater than those of QE. This is because CMP, in the model, was particularly beneficial for banks compared to QE. The right panel of Figure 13 shows that CMP had a greater and more persistent positive impact on banks' net worth, which constitutes profits in the model. This occured because CMP increased banks' profitability, helping the banking sector grow, while QE limited that growth by reducing the expected rate of return on their investments.

In summary, I find that CMP could have had greater stimulus effects than QE during the ELB episode if the ELB were not binding. In terms of the relative magnitudes of the effects on different variables, CMP had particularly strong effects on profits, driven by its positive impact on the banking sector, compared to QE. Additionally, CMP led to lower real interest rates during the same period compared to the baseline case, where the policy rate was constrained by the ELB.

[55] Figure 13 shows the value of equity holdings, which includes the effects of price changes.

[56] In the case of CMP, the average effect on profits was about 3.3%, while QE increased profits by only about 1.4% compared to the counterfactual case with no policy interventions.

Figure 14: Distributional effects of QE and CMP: Gini index and income shares



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.

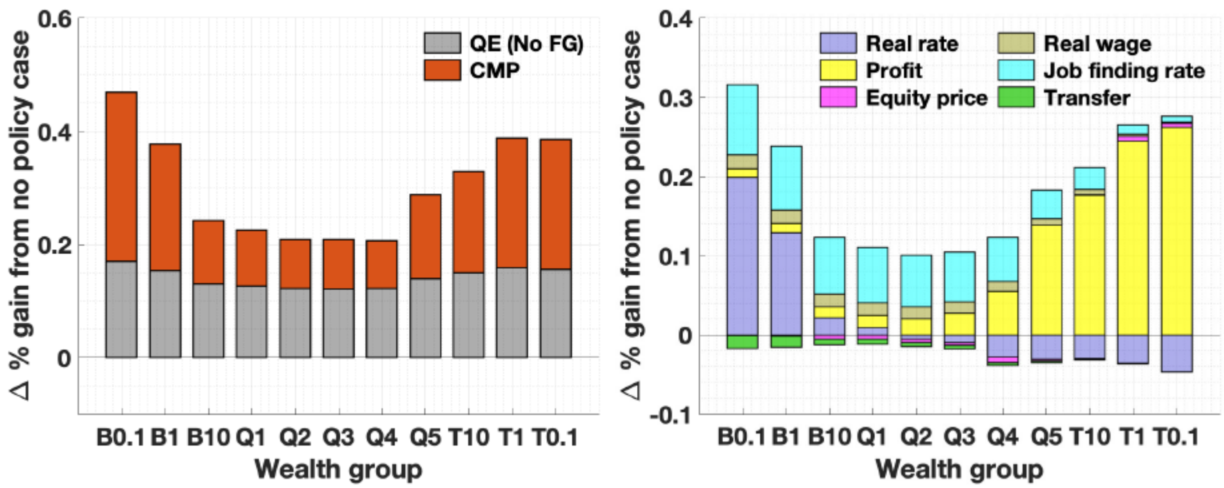
6.2 Distributional effects of QE and CMP

The particularly strong effects on profits, discussed in the previous section, led to more adverse distributional effects of CMP compared to those of QE. The left panel of Figure 14 shows that CMP increased the income Gini index relative to the counterfactual case with no policy interventions, whereas QE slightly reduced the index. Although CMP provided additional stimulus to the labor market, resulting in a slightly greater decrease in the Gini index for the bottom 90% of households than QE, especially in the early stages of the ELB episode, the substantial impact on profits was significant enough to reverse this trend, leading to an overall increase in the Gini index.

The effects on the top 10% income shares further confirm that CMP had more adverse distributional effects. CMP led to a greater increase in the top 10% income share compared to QE, primarily due to its particularly positive impact on the banking sector and profits. This concentration of income at the top contributed to widening inequality, as the gains from CMP disproportionately benefited wealthier households, exacerbating income inequality more than QE did.

Lastly, the comparison of welfare gains from QE and CMP shows that while CMP provided larger gains overall, it disproportionately benefited wealthier households, though it also delivered significant gains to households at the very bottom of the wealth distribution, which includes a substantial share of borrowers. As shown in the left panel of Figure A19, CMP provided larger welfare gains to all households compared to QE. However, when com-

Figure 15: Welfare gain comparison: QE vs CMP



Notes: The figure compares the welfare gains from QE and conventional monetary policy in terms of consumption equivalents. The gray bars represent welfare gains from QE across households, while the dark orange bars indicate the additional gains from conventional monetary policy compared to the welfare gains from QE. The right panel provides a breakdown of the additional gains from CMP, relative to the gains from QE.

paring the gains for the bottom and top 10% of households, CMP delivered larger gains to wealthier households overall. Interestingly, households in the bottom 0.1% and 1% of the wealth distribution experienced greater gains than those in the top 0.1% and 1%. This indicates that CMP amplified the non-linear distributional effects seen with QE.

The decomposition of these additional welfare gains highlights two key points. First, lower real interest rates, driven by CMP, indeed redistributed the benefits of stimulus from savers to borrowers, resulting in the largest gains for households at the very bottom of the wealth distribution, many of whom are borrowers. However, in the model, the banking sector also benefited from lower real interest rates, leading to higher profits, which in turn benefited households at the top of the wealth distribution.

To summarize, CMP had more adverse effects on inequality than QE alone, even though it was more effective in stimulating the aggregate economy. This again highlights the tension between stimulating economic activity and exacerbating inequality, as policies like CMP tend to disproportionately benefit wealthier households while still providing important gains to lower-wealth borrowers.

7 Conclusion

In this paper, I examine the distributional consequences of UMP during the ELB episode following the Great Recession in the U.S. To this end, I develop a medium-scale Heterogeneous Agent New Keynesian (HANK) model incorporating portfolio choice, wage rigidity, labor market frictions, banks, and a zero lower bound on the policy rate. I model QE as central bank private asset purchases, following Gertler and Karadi (2011), and forward guidance as exogenous expected ELB durations, based on Jones (2017). The model is calibrated to match micro-level data on households' wealth and income composition, and estimated using U.S. macroeconomic data and Bayesian methods to capture the dynamics of key aggregate variables such as real wages, unemployment, and profits.

The estimated model generates empirically plausible dynamics in response to exogenous shocks. In particular,it generates a procyclical response of profits to an expansionary monetary policy shockto expansionary monetary policy shocks, a feature absent in many New Keynesian models. This allows the model to capture the substantial benefits that wealthy households gained from expansionary monetary policy, which existing models typically fails to capture.

A counterfactual analysis revealed that UMP reduced the income Gini index during the ELB episode, primarily through positive effects on employment. However, QE also widened the income gap between the top 10% and the bottom 90% by significantly increasing profits and equity prices. In comparing different types of UMP, I found that QE and forward guidance had similar aggregate and distributional effects, amplifying each other. Lastly, conventional monetary policy (CMP) had more adverse distributional effects, particularly benefiting the banking sector, even though it was more effective in stimulating the economy than QE.

The results of this paper suggest that the criticisms of UMP, particularly QE, for exacerbating inequality, as well as the counterargument emphasizing its positive effects on labor markets, are both valid depending on the focus. If one focuses on the gap between the top 10% and the rest of the population, UMP appears to increase inequality. However, if one looks at the welfare improvements for the bottom 10%, UMP reduces inequality by narrowing the gap between the bottom and middle of the wealth distribution. Crucially, the findings imply that models which fail to account for the positive impact of monetary policy on profits may provide a misleading or incomplete picture of monetary policy's effects on inequality, since the dynamics of profits are critical in shaping the distributional effects of monetary policy.

References

Altavilla, C., Boucinha, M., and Peydró, J.-L. (2018). Monetary policy and bank profitability in a low interest rate environment. Economic Policy, 33(96):531–586.

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

Amaral, P. S. (2017). Monetary Policy and Inequality. Economic Commentary, (January).

Anderson, E., Rebelo, S., and Wong, A. (2018). Markups across space and time. Working Paper 24434, National Bureau of Economic Research.

Atkinson, T., Richter, A. W., and Throckmorton, N. A. (2019). The zero lower bound and estimation accuracy. Journal of Monetary Economics.

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

Auclert, A., Bardóczy, B., and Rognlie, M. (2023). MPCs, MPEs, and Multipliers: A Trilemma for New Keynesian Models. The Review of Economics and Statistics, 105(3):700–712.

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

Ball, L. M. (2014). The Case for a Long-Run Inflation Target of Four Percent. IMF Working Papers 2014/092, International Monetary Fund.

Bank of England, B. o. E. (2012). The distributional effects of asset purchases. Bank of England Quarterly Bulletin, 52(3):254–266.

Bayer, C., Born, B., and Luetticke, R. (2020). Shocks, frictions, and inequality in us business cycles. CEPR Discussion Papers 14364.

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

Bayer, C., Luetticke, R., Pham-Dao, L., and Tjaden, V. (2019). Precautionary savings, illiquid assets, and the aggregate consequences of shocks to household income risk. Econometrica, 87(1):255–290.

Bernanke, B. S. (2015). Monetary policy and inequality. Brookings Blog.

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

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

Blanchard, O. J., Dell'Ariccia, G., and Mauro, P. (2010). Rethinking Macroeconomic Policy. IMF Staff Position Notes 2010/003, International Monetary Fund.

Borio, C., Gambacorta, L., and Hofmann, B. (2017). The influence of monetary policy on bank profitability. International Finance, 20(1):48–63.

Broer, T., Harbo Hansen, N.-J., Krusell, P., and Öberg, E. (2019). The New Keynesian Transmission Mechanism: A Heterogeneous-Agent Perspective. The Review of Economic Studies, 87(1):77–101.

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

Casiraghi, M., Gaiotti, E., Rodano, L., and Secchi, A. (2018). A "reverse robin hood"? the distributional implications of non-standard monetary policy for italian households. Journal of International Money and Finance, 85:215 – 235.

Chetty, R., Guren, A., Manoli, D., and Weber, A. (2011). Are micro and macro labor supply elasticities consistent? a review of evidence on the intensive and extensive margins. American Economic Review, 101(3):471–75.

Chib, S. and Ramamurthy, S. (2010). Tailored randomized block MCMC methods with application to DSGE models. Journal of Econometrics, 155(1):19–38.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (1999). Chapter 2 monetary policy shocks: What have we learned and to what end? volume 1 of Handbook of Macroeconomics, pages 65 – 148. Elsevier.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. Journal of Political Economy, 113(1):1–45.

Christiano, L. J., Eichenbaum, M. S., and Trabandt, M. (2016). Unemployment and business cycles. Econometrica, 84(4):1523–1569.

Chung, H., Laforte, J.-P., Reifschneider, D., and Williams, J. C. (2012). Have We Underestimated the Likelihood and Severity of Zero Lower Bound Events? Journal of Money, Credit and Banking, 44:47–82.

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

Coibion, O., Gorodnichenko, Y., Kueng, L., and Silvia, J. (2017). Innocent bystanders? monetary policy and inequality. Journal of Monetary Economics, 88:70 – 89.

Cuba-Borda, P., Guerrieri, L., Iacoviello, M., and Zhong, M. (2019). Likelihood evaluation of models with occasionally binding constraints. Journal of Applied Econometrics, 34(7):1073–1085.

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

den Haan, W. J., Ramey, G., and Watson, J. (2000). Job destruction and propagation of shocks. American Economic Review, 90(3):482–498.

Domanski, D., Scatigna, M., and Zabai, A. (2016). Wealth inequality and monetary policy. BIS Quarterly Review.

Draghi, M. (2016). Stability, equity, and monetary policy. 2nd DIW Europe Lecture, German Institute for Economic Research (DIW).

Engen, E. M., Laubach, T., and Reifschneider, D. L. (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 (U.S.).

Furceri, D., Loungani, P., and Zdzienicka, A. (2018). The effects of monetary policy shocks on inequality. Journal of International Money and Finance, 85:168 – 186.

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

Gornemann, N., Kuester, K., and Nakajima, M. (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, E. (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, J., Hercowitz, Z., and Huffman, G. W. (1988). Investment, capacity utilization, and the real business cycle. The American Economic Review, 78(3):402–417.

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

Guerrieri, L. and Iacoviello, M. (2017). Collateral constraints and macroeconomic asymmetries. Journal of Monetary Economics, 90:28–49.

Hafemann, L., Rudel, P., and Schmidt, J. (2018). Moving closer or drifting apart: Distributional effects of monetary policy. The Manchester School, 86(S1):110–136.

Heathcote, J., Perri, F., and Violante, G. L. (2010). Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006. Review of Economic Dynamics, 13(1):15–51.

Inui, M., Sudo, N., and Yamada, T. (2017). Effects of Monetary Policy Shocks on Inequality in Japan. Bank of Japan Working Paper Series 17-E-3, Bank of Japan.

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

Jones, C., Kulish, M., and Rees, D. M. (2018). International Spillovers of Forward Guidance Shocks. IMF Working Papers 18/114, International Monetary Fund.

Jurkatis, S. and Strehl, W. (2014). Gini decompositions and Gini elasticities: On measuring the importance of income sources and population subgroups for income inequality. Technical report.

Justiniano, A., Primiceri, G. E., and Tambalotti, A. (2011). Investment shocks and the relative price of investment. Review of Economic Dynamics, 14(1):102 – 121. Special issue: Sources of Business Cycles.

Kaplan, G., Moll, B., and Violante, G. L. (2018). Monetary policy according to hank. American Economic Review, 108(3):697–743.

Kiley, M. T. (2014). The Response of Equity Prices to Movements in Long-Term Interest Rates Associated with Monetary Policy Statements: Before and After the Zero Lower Bound. Journal of Money, Credit and Banking, 46(5):1057–1071.

Kulish, M., Morley, J., and Robinson, T. (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, D. (2019). The Effects of Monetary Policy on Consumption and Inequality. Working paper.

Lenza, M. and Slacalek, J. (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.

Luetticke, R. (2020). Transmission of monetary policy with heterogeneity in household portfolios. American Economic Journal: Macroeconomics, forthcoming.

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

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

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

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

Ravenna, F. and Walsh, C. (2008). Vacancies, unemployment, and the phillips curve. European Economic Review, 52(8):1494–1521.

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

Rosa, C. (2012). How "unconventional" are large-scale asset purchases? The impact of monetary policy on asset prices. Technical report.

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

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

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

Smets, F. and Wouters, R. (2007). Shocks and frictions in us business cycles: A bayesian dsge approach. American Economic Review, 97(3):586–606.

Storesletten, K., Telmer, C., and Yaron, A. (2004). Cyclical dynamics in idiosyncratic labor market risk. Journal of Political Economy, 112(3):695–717.

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

Williams, J. C. (2016). Monetary Policy in a Low R-star World. FRBSF Economic Letter.

Winberry, T. (2018). A method for solving and estimating heterogeneous agent macro models. Quantitative Economics, 9(3):1123–1151.

Yellen, J. 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, E. 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):36 – 41. Computational Suite of Models with Heterogeneous Agents: Incomplete Markets and Aggregate Uncertainty.

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

Appendix

A Further details on the model description

A.1 Households

Let V_a and V_b denote the partial derivative of the value function with respect to illiquid and liquid asset holding, respectively. Similarly, u_c denotes the partial derivative of the utility function with respect to consumption. By Envelope Theorem, I have the following expressions for the partial derivatives of the value function.

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

$$V_b(a_t, b_t) = \begin{cases} \left(\frac{1+\tilde{i}_t}{\pi_t}\right) u_c(c_t^A, n_t) & \text{if adjust} \\ \left(\frac{1+\tilde{i}_t}{\pi_t}\right) u'(c_t^N, n_t) & \text{if not adjust} \end{cases} \tag{A.2}$$

where c_t^A and c_t^N are the optimal consumption when the household chooses to adjust its illiquid asset holding or not, respectively.[57] Households choose to adjust their equity holdings if the following conditions are satisfied.

$$V^A(a_t, b_t) - \chi_t \geq V^N(a_t, b_t) \tag{A.3}$$

where V^A and V^N denote the value of households when they adjust and do not adjust their illiquid asset holding respectively. Then, the probability of adjustment $P^*(a_t, b_t)$ can be computed as follows.

$$\begin{aligned} P^*(a_t, b_t) &= P\Big[\chi_t \leq V^A(a_t, b_t) - V^N(a_t, b_t)\Big] \\ &= F\Big[V^A(a_t, b_t) - V^N(a_t, b_t)\Big] \end{aligned} \tag{A.4}$$

[57] Households' optimal hours worked is not affected by the household's portfolio choice.

Given the probability of adjustment, the household's Euler equation with respect to each asset holding can be described as follows.

$$q_t u_c(c_t, n_t) \geq \beta \mathbb{E}\Big[P^*(a_{t+1}, b_{t+1})\Big\{q_{t+1} + r^a_{t+1}\Big\}u_c(c^A_{t+1}, n_{t+1}) + \Big\{1 - P^*(a_{t+1}, b_{t+1})\Big\}r^a_{t+1}u_c(c^N_{t+1}, n_{t+1})$$
$$+ \Big\{1 - P^*(a_{t+1}, b_{t+1})\Big\}\mathbb{E}\Big[V_a(a_{t+1}, b_{t+2})\Big]\Big] \text{ with equality if } a_{t+1} > 0 \quad \text{and} \quad a_{t+1} \neq a_t \tag{A.5}$$

$$u_c(c_t, n_t) \geq \beta \mathbb{E}\Big[P^*(a_{t+1}, b_{t+1})\Psi^l_t\Big(\frac{1 + \tilde{i}_{t+1}}{\pi_{t+1}}\Big)u_c(c^A_{t+1}, n_{t+1}) + \{1 - P^*(a_{t+1}, b_{t+1})\}\Psi^l_t\Big(\frac{1 + \tilde{i}_{t+1}}{\pi_{t+1}}\Big)u_c(c^N_{t+1}, n_{t+1})\Big]$$
$$\text{with equality if } b_{t+1} > 0 \tag{A.6}$$

Note that, as explained in the main text, households' optimality condition regarding liquid assets is perturbed by liquidity preference shocks.

A.2 Banks

As long as the expected equity premium $R^a_{t+i} - R_{t+i}$ is positive, a bank's optimal choice is to purchase assets to the extent possible. If there is no limit in taking deposits, either a bank expands its assets indefinitely, or the premium becomes zero. To limit the bank's ability to borrow, I assume a moral hazard/costly enforcement problem, as in Gertler and Karadi (2011). Specifically, at the beginning of the period, a bank can divert the fraction Δ of the bank's asset and transfer it to business owners. Once the bank diverts the funds, the depositors force the bank into bankruptcy but can recover only the remaining $1 - \Delta$ fraction of assets. It is too costly for the depositors to recover all the funds that the banker diverted. Taking into account this incentive problem, investors will make deposits only to the point the following constraint holds.

$$J^b(N_{jt}) \geq \Delta q_t A^b_{jt+1} \tag{A.7}$$

where the left-hand side is the cost for the bank when it diverts a fraction of assets, i.e., the franchise value of the bank. The right-hand side is the value of diverting. To further specify the above condition, one needs to compute the value of the bank. Using the guess and verify approach, one can show that the bank j's value $J^b(N_{jt})$ is linear in its assets and net-worth.

$$J^b(N_{jt}) = \vartheta^a_t q_t A^b_{jt+1} + \vartheta^n_t N_{jt} \tag{A.8}$$

with

$$\vartheta_t^a = \mathbb{E}_t\Big[(1-\theta_b)\Psi_t^b\Lambda_{t,t+1}(R_{t+1}^a - R_{t+1}) + \theta_b\Psi_t^b\Lambda_{t,t+1}x_{t,t+1}\nu_{t+1}\Big] \tag{A.9}$$

$$\vartheta_t^n = \mathbb{E}_t\Big[(1-\theta_b)\Psi_t^b\Lambda_{t,t+1}R_{t+1} + \theta_b\Psi_t^b\Lambda_{t,t+1}z_{t,t+1}\eta_{t+1}\Big] = (1-\theta_b) + \mathbb{E}_t\Big[\theta_b\Psi_t^b\Lambda_{t,t+1}z_{t,t+1}\eta_{t+1}\Big] \tag{A.10}$$

where $x_t = q_{t+1}A_{jt+2}^b/q_tA_{jt+1}^b$ is the gross growth rate in assets between t and $t+1$ and $z_t = N_{jt+1}/N_{jt}$ is the gross growth rate of net worth. Ψ_t^b is the aggregate risk premium shock, which follows an AR(1) process as below.

$$\log\Psi_t^b = \rho_b\log\Psi_{t-1}^b + \epsilon_{b,t} \ \ , \ \epsilon_{b,t} \ \sim \ N(0,\sigma_b^2) \tag{A.11}$$

where $\epsilon_{b,t}$ is a normally distributed shock, and σ_b is its standard deviation. An increase in Ψ_t^b leads to an increase in the value of banks' assets and net-worth by making banks value future more. Thus, a positive shock to Ψ_t^b leads to an expansion of banks' balance sheet.

With the value function derived above, I can re-write the incentive constraint as follows.

$$\vartheta_t^a q_t A_{jt+1}^b + \vartheta_t^n N_{jt} \geq \Delta q_t A_{jt+1}^b \tag{A.12}$$

If the constraint binds, the value of assets that the banker can purchase will be determined by the level of his or her net worth. By re-arranging the above equation, we have

$$q_tA_{jt+1}^b = \frac{\vartheta_t^n}{\Delta - \vartheta_t^a}N_{jt} = \Theta_tN_{jt} \tag{A.13}$$

where Θ_t is the bank's leverage ratio, i.e., the ratio of assets to its net worth.[58] When the constraint binds, I can express the law of motion for net worth as follows.

$$N_{jt+1} = \Big\{(R_{t+1}^a - R_{t+1})\Theta_t + R_{t+1}\Big\}N_{jt} \tag{A.14}$$

In addition, it follows that

$$z_{t,t+1} = N_{jt+1}/N_{jt} = \Big\{(R_{t+1}^a - R_{t+1})\Theta_t + R_{t+1}\Big\} \tag{A.15}$$

$$x_{t,t+1} = q_{t+1}A_{jt+2}^b/q_tA_{jt+1}^b = \Theta_{t+1}N_{jt+1}/\Theta_tN_{jt} = (\Theta_{t+1}/\Theta_t)z_{t,t+1} \tag{A.16}$$

Note that all components of Θ_t do not depend on bank-specific variables. Thus, I can sum

[58]Note that, given $N_{jt} > 0$, the constraint binds only if $0 < \vartheta_t^a < \Delta$. Under the parametrizations used in this paper, the constraint always binds.

across banks to obtain

$$q_t A_{t+1}^b = \Theta_t N_t \tag{A.17}$$

where A_{t+1}^b is the aggregate quantity of the equity held by banks and N_t denote the aggregate bank net worth.

Finally, I describe a law of motion for N_t. First, note that N_t is the sum of the net worth of surviving banks, N_{ot} (old), and the net worth of entrants, N_{nt} (new). Regarding the latter, I assume that the value of start-up funds for new bank is equal to the value of assets that exiting banks had intermediated in the previous period, which equals $(1-\theta_b)q_{t-1}A_t^b$. Specifically, for each new bank, the equity mutual fund gives $\omega/(1-\theta_b)$ fraction of this value. Then, I have

$$N_t = N_{ot} + N_{et} = \theta_b\{(R_t^a - R_t)\Theta_{t-1} + R_t\}N_{t-1} + \omega q_{t-1}A_t^b \tag{A.18}$$

Finally, profits from the financial sector are the sum of net-worth of existing banks, net of start-up funds for new banks.

$$\Pi_t^b = (1-\theta_b)\{(R_t^a - R_t)\Theta_{t-1} + R_t\}N_{t-1} - \omega q_{t-1}A_t^b \tag{A.19}$$

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).[59] The methodology enables a fast solution that is necessary for

[59] 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.

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.[60] 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.

$$\mathbb{E}_t\left[\mathcal{F}(X_{t+1}, Y_{t+1}, X_t, Y_t)\right] = 0 \tag{A.20}$$

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 the end of period t.[61] 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 (A.20) as follows.

$$\mathbb{E}_t\left[\mathcal{F}(X_{t+1}, Y_{t+1}, X_t, Y_t)\right] = \left[\mathcal{F}_{1,t}, \mathcal{F}_{2,t}, \mathcal{F}_{3,t}, \mathcal{F}_{4,t}, \mathcal{F}_{5,t}, \mathcal{F}_{6,t}, \mathcal{F}_{7,t}\right]' \tag{A.21}$$

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

[60] 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.

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

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.[62] 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 (A.21), we know that the Jacobian has the following form.

$$\mathcal{J}_t = \begin{bmatrix} \frac{\partial \mathcal{F}_{1,t}}{\partial X_{t+1}} & \frac{\partial \mathcal{F}_{1,t}}{\partial Y_{t+1}} & \frac{\partial \mathcal{F}_{1,t}}{\partial X_t} & \frac{\partial \mathcal{F}_{1,t}}{\partial Y_t} \\ \frac{\partial \mathcal{F}_{2,t}}{\partial X_{t+1}} & \frac{\partial \mathcal{F}_{2,t}}{\partial Y_{t+1}} & \frac{\partial \mathcal{F}_{2,t}}{\partial X_t} & \frac{\partial \mathcal{F}_{2,t}}{\partial Y_t} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \mathcal{F}_{7,t}}{\partial X_{t+1}} & \frac{\partial \mathcal{F}_{7,t}}{\partial Y_{t+1}} & \frac{\partial \mathcal{F}_{7,t}}{\partial X_t} & \frac{\partial \mathcal{F}_{7,t}}{\partial Y_t} \end{bmatrix} \tag{A.22}$$

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 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.

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

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, ϵ_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.

$$X_t = P_t X_{t-1} + D_t + Q_t \epsilon_t \tag{A.23}$$

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

$$X_t = P X_{t-1} + Q \epsilon_t \tag{A.24}$$

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.[63] Then,

$$Y_t = H_t X_t = H_t P X_{t-1} + H_t Q \epsilon_t \tag{A.25}$$

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.

$$\epsilon_t = (H_t Q)^{-1} (Y_t - H_t P X_{t-1}) \tag{A.26}$$

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.

$$\epsilon_t(\tilde{T}_t) = \{H_t Q(\tilde{T}_t)\}^{-1} \{Y_t - H_t P(\tilde{T}_t) X_{t-1} - H_t D(\tilde{T}_t)\Big\} \tag{A.27}$$

where $\tilde{T}_t$ is the expected ZLB durations in period t. Note that the solution and the corre-

[63] 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.

sponding 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 as follows.

$$\log p\Big(\mathbb{Y}_{\{1:T\}}\Big) = -\frac{T n_y}{2}\log(2\pi) - \frac{T}{2}\log(\det(\Sigma)) - \frac{1}{2}\sum_{t=1}^{T}\epsilon_t'\Sigma^{-1}\epsilon_t + \sum_{t=1}^{T}\log\left(|\det\frac{\partial\epsilon_t}{\partial Y_t}|\right) \quad \text{(A.28)}$$

where $\frac{\partial\epsilon_t}{\partial Y_t} = \Big\{H_t Q_t\Big\}^{-1}$.[64]

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

C Further details on the calibration

The model is a quarterly model. I set the inverse of the elasticity of inter-temporal 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% in the data.[65] 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 inequality of illiquid asset holding in the model. The calibrated adjustment costs imply an average adjustment frequency of 6.7% per quarter at the steady state, which is close to 6.5%, the value used in Bayer et al. (2020). Also, with the calibrated adjustment costs, the top 10% wealthiest households hold 73% of total illiquid assets in the model, compared to 74% in the data.

The income process, which is the ultimate source of inequality in the model, is reverse-engineered to match asset holding and wealth inequality in the data. First, I set the income process for s_t as a standard AR(1) process with three states, using the Tauchen (1986) method for discretization. I set the autocorrelation and 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 (super low-skilled and super high-skilled) to match the wealth inequality in the data. I fix the probability of becoming a business owner P_e to 0.05%, which is similar to the value used in Bayer et al. (2019). Then, I calibrate the probability of leaving the business owner state, which represents top-income earners' income risk, to match the top 10% wealthiest households' share of liquid asset. The resulting value for $\tilde{P}_e$ is 20.6%.

I set exogenous job separation rate at 10%, following den Haan et al. (2000). Also, the steady state real wage is set to 1.2112 to have a ratio of labor income to output, net of fixed costs, of 60% at the steady state. I target a vacancy filling rate of 70%, based on den Haan et al. (2000), Ravenna and Walsh (2008), and Christiano et al. (2016). The target for the steady state unemployment rate is set to 5.5%, which is the average unemployment rate before the Great Recession in my sample. Matching these targets, for the given job separation rate, the steady state real wage, and the vacancy posting cost implies a matching efficiency of 1.7127 and the matching maintenance cost of 0.0076.[66]

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

[66] I estimate the vacancy posting cost and adjust the cost Ξ^L to ensure that the free entry condition is satisfied for given labor market parameter values. The value presented in Table 1 for Ξ^L corresponds to the

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 share of the fixed cost in production. For the given value of labor agencies and other firms' profits, I set the fixed cost to match the capital to output ratio of 3.03 in the data.[67][68] The exponent of capital in the production function is set to 0.27, which implies the capital share, i.e., the sum of profits of intermediate good firms and capital rental payment, to output, net of fixed costs, of 40%.

The parameters associated with variable capital utilization are calibrated to match two targets; the steady state utilization rate and the depreciation rate. As is standard, I set the steady state utilization rate to 1. Then, I target a steady state depreciation rate of 6% (annualized), a standard value used in the literature. Matching these two targets results in $\delta_0 = 0.015$ and $\delta_1 = 1.0025$.

For the financial sector parametrization, I mainly follow Gertler and Karadi (2011). I target a steady state leverage ratio of 3, which implies $\Delta = 0.3304$. The survival rate of banks is 0.97, and ω is set to 0.0076 to match the banks' equtiy share of 55%. The money market mutual fund's discount factor is set to ensure that the steady state inter-temporal optimality condition holds for a given real rate of return on liquid assets.[69] The fraction of tax revenues that is given to the fund is set to 5.33% to ensure a tax rate of 30%, while matching the share of lump-sum transfers in the income of bottom 80%. Finally, the fraction of firms' profits that is given to business owners is set to 23.89%, 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%, which is a standard value used in the literature. The tax rate is 30%. The levels of government purchases and lump-sum transfers are set to match the share of transfer income in the bottom 80% households' income and the tax rates of 30%. The borrowing premium of 2.53% is chosen to help match the mass of households with zero assets. Also, the borrowing limit is set to match the fraction of households with debt in the data.

The central bank's inflation target is set to 1.005, which is the current quarterly inflation

value of the vacancy posting cost at the posterior mode.

[67] I measure aggregate capital as the current-cost net stock of private fixed assets from the Bureau of Economic Analysis. Consumer durables are not included.

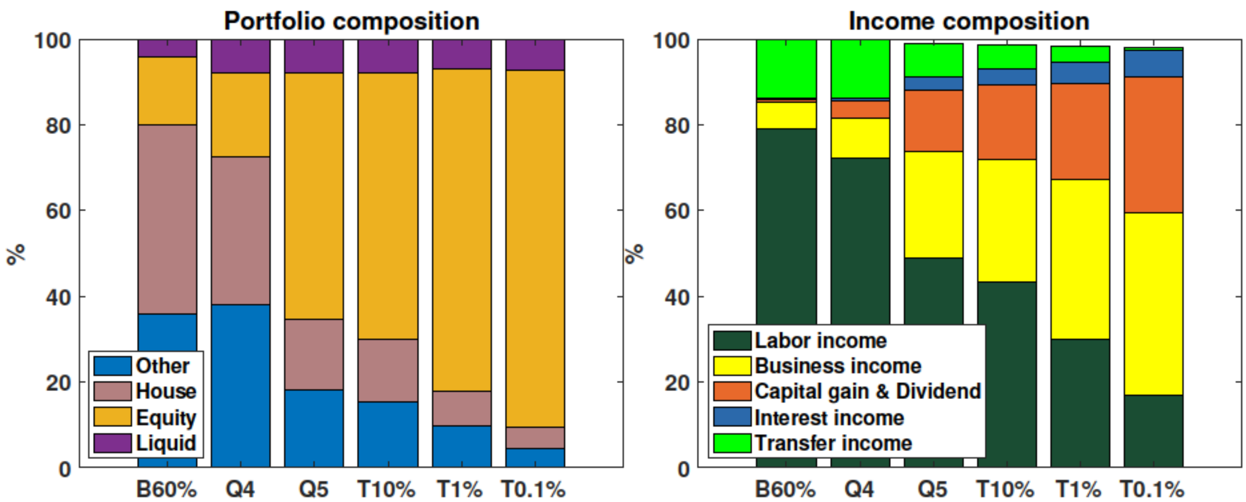
[68] In the estimation, the vacancy posting cost varies. To ensure that the free entry condition holds, I adjust Ξ^L. However, adjusting Ξ^L changes the value of labor agencies at the steady state, which also affects the level of aggregate profits and the dividend rate. Thus, to maintain the steady state dividend rate, I also adjust the fixed cost of production for intermediate good firms, along with Ξ^L. The value presented in Table 1 is the level of the fixed cost that corresponds to the posterior mode of the vacancy posting cost.

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

target of the Federal Reserve. The steady state policy rate is calibrated to match households' liquid to illiquid asset ratio in the data. Also, I assume that the central bank's assets are equal to 5% of output at the steady state, based on the historical average before the implementation of QE. Finally, the auto-correlation of the central bank's assets is set to 0.99.

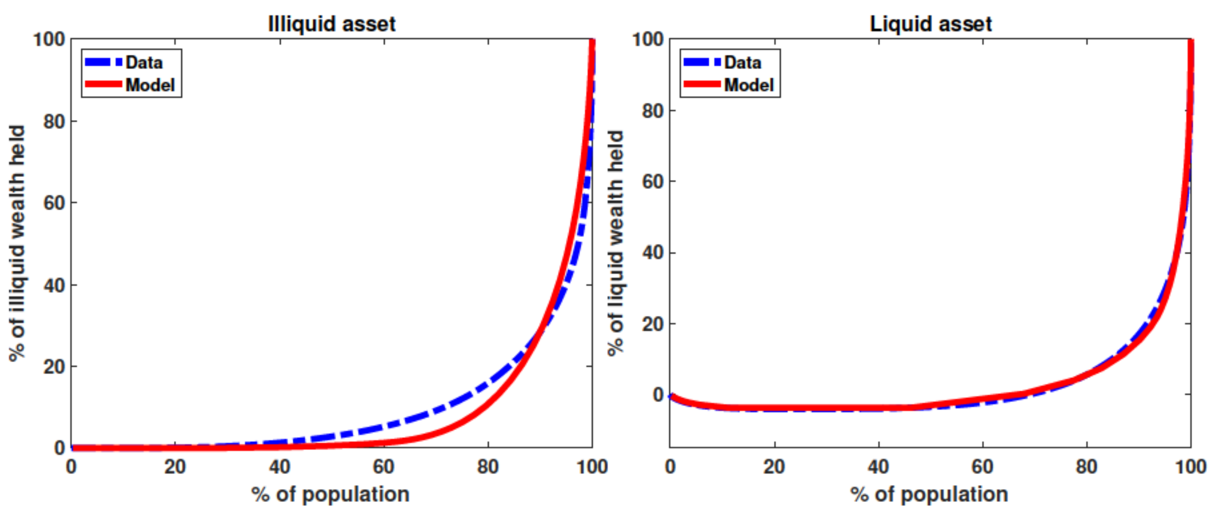
D Fit of the model

Figure A1: Portfolio and income composition in the data



Notes: The figure shows more detailed decomposition of households portfolio and income composition in the data. For the description of each item, see the main text.

Figure A2: Lorenz curves in the data and the model



Notes: The figure shows asset 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.

E Further details on the estimation

E.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.

E.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 civil-

ian 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

E.3 Structural shocks

1. Total factor productivity shock

$$\log Z_t = \rho_z \log Z_{t-1} + \epsilon_{Z,t} \ , \epsilon_{Z,t} \ \sim \ N(0, \sigma^2_{\epsilon_{Z},t}) \tag{A.29}$$

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

$$\Lambda^b_{t,t+1} = \Psi^b_t \Lambda_{t,t+1} \tag{A.30}$$

$$\log\left(\frac{\Psi^b_t}{\Psi^b}\right) = \rho_b \log\left(\frac{\Psi^b_{t-1}}{\Psi^b}\right) + \epsilon_{b,t} \ , \epsilon_{b,t} \ \sim \ N(0, \sigma^2_{b,t}) \tag{A.31}$$

3. Price mark-up shock

$$\Psi^p_t = \frac{\eta_t}{\eta_t - 1} \tag{A.32}$$

$$\log(\Psi^p_t) = \rho_p \log(\Psi^p_{t-1}) + \epsilon_{p,t} \ , \ \epsilon_{p,t} \ \sim \ N(0, \sigma^2_p) \tag{A.33}$$

4. Investment technology shock

$$\log(\Psi_t^k) = \rho_k \log(\Psi_{t-1}^k) + \epsilon_{k,t} \quad , \sim \; N(0, \sigma_k^2) \tag{A.34}$$

5. Liquidity preference shock

$$\log(\Psi_t^l) = \rho_l \log(\Psi_{t-1}^l) + \epsilon_{l,t} \; , \; \epsilon_{l,t} \; \sim \; N(0, \sigma_l^2) \tag{A.35}$$

6. Wage shock

$$\frac{w_t}{w} = \left(\epsilon_{w,t} \frac{r_t^l}{r^l}\right)^{\vartheta_w(1-\rho_w)} \left\{\frac{w_{t-1}}{w} \times \left(\frac{\pi}{\pi_t}\right)\right\}^{\rho_w} \quad , \quad 0 < \rho_w < 1 \; , \; \vartheta_w > 0 \tag{A.36}$$

$$\tag{A.37}$$

7. Lump-sum transfer shock

$$T_t^g = \left(1 - \frac{1}{\Psi_t^g}\right) Y \tag{A.38}$$

$$\log\left(\frac{\Psi_t^g}{\Psi^g}\right) = \rho_g \log\left(\frac{\Psi_{t-1}^g}{\Psi^g}\right) + \epsilon_{g,t} \; , \epsilon_{g,t} \; \sim \; N(0, \sigma_g^2) \tag{A.39}$$

8. Monetary policy shock

$$1 + \hat{i}_{t+1} = (1 + \hat{i}) \left(\frac{1 + \hat{i}_t}{1 + \hat{i}}\right)^{\rho_R} \left[\left(\frac{\pi_t}{\pi}\right)^{\phi_\pi} \left\{\exp(u_t - u)\right\}^{\phi_u}\right]^{1-\rho_R} \exp(\epsilon_{R,t}) \quad , \; \epsilon_{R,t} \; \sim \; N(0, \sigma_R^2) \tag{A.40}$$

$$i_{t+1} = \min\{0, \hat{i}_{t+1}\} \tag{A.41}$$

9. Fixed cost shock

$$\Psi_t^F = \rho_F \Psi_{t-1}^F + (1 - \rho_F)\Psi^F + \epsilon_{F,t} \quad , \; \epsilon_{F,t} \; \sim \; N(0, \sigma_F^2) \tag{A.42}$$

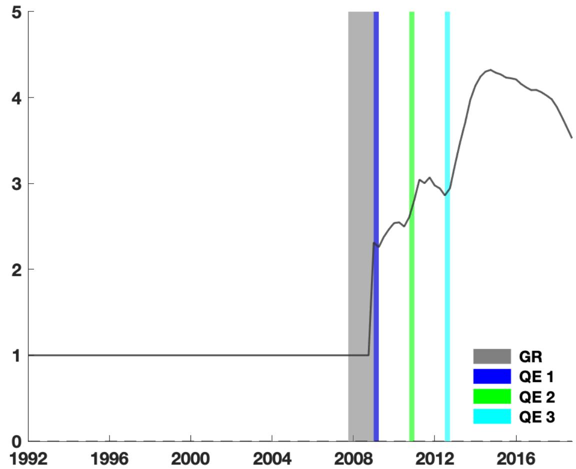
$$\tag{A.43}$$

10. QE shock

$$A_{t+1}^{\mathrm{CB}} = \Psi_{\mathrm{QE},t} Y \; , \quad \log(\Psi_t^{\mathrm{QE}}) = \rho_{\mathrm{QE}} \log(\Psi_{t-1}^{\mathrm{QE}}) + \epsilon_{\mathrm{QE},t} \quad \epsilon_{\mathrm{QE},t} \; \sim \; N(0, \sigma_{\mathrm{QE}}^2) \tag{A.44}$$

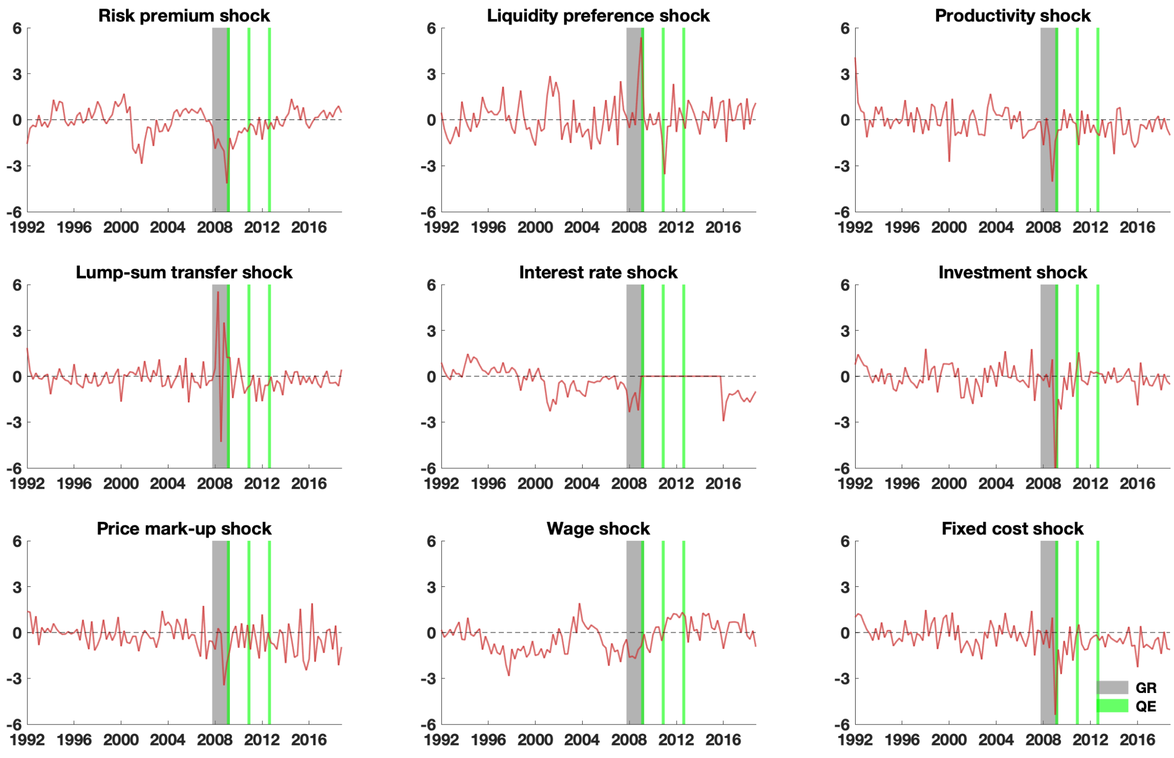
E.4 Additional figures and tables

Figure A3: The central bank's assets



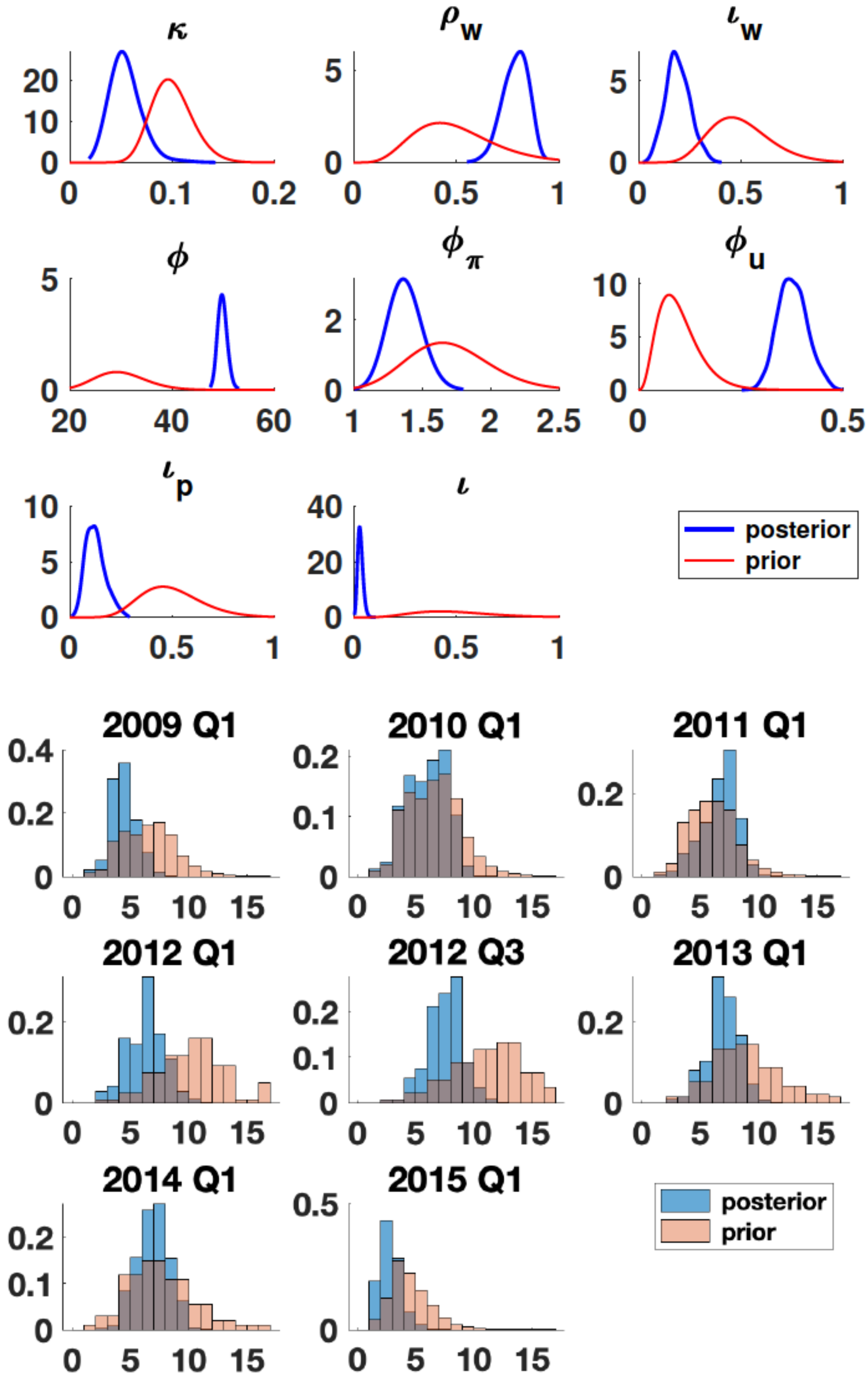
Notes: The figure shows the central bank's asset as the ratio to its end of 2007 level. Green, blue, green, and sky blue area depict the Great Recession periods, the period in which QE 1, 2, and 3 are announced.

Figure A4: 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.

Figure A5: Posterior distributions of estimated parameters and expected ZLB durations



F 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.[70] 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 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.

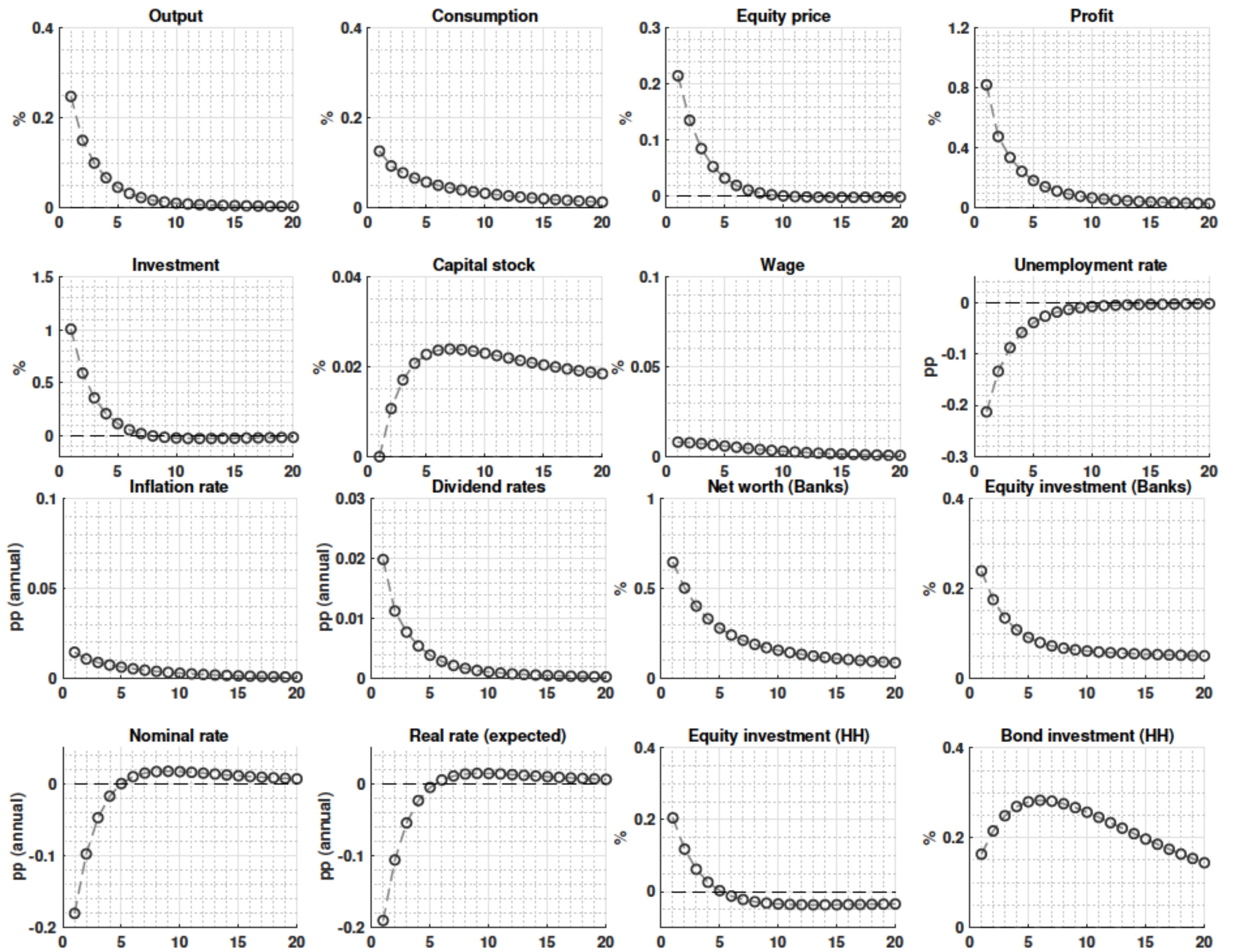
In the following subsections, I show the model's impulse responses, including a procyclical response of profits, to an expansionary monetary policy shock, and discuss how the model generates such a response.

F.1 Procyclical profits

Figure A6 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

[70] 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.

Figure A6: Impulse responses to an expansionary monetary policy shock

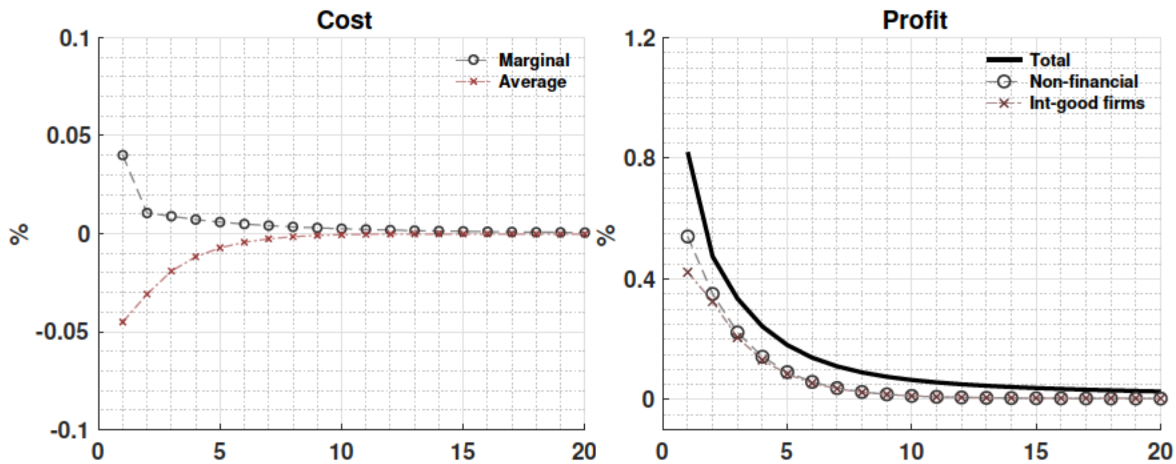


Notes: The figure shows the model's impulse responses to a negative 25 basis points (annualized) interest rate shock. All variables are shown as the percentage deviations from their respective steady state values except for the nominal rate, the inflation rate, the dividend rates and the unemployment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

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.[71]

[71] 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.

Figure A7: 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.

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.[72] 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. [73]

Besides, based on a recent finding of Anderson et al. (2018), I assume that the fixed

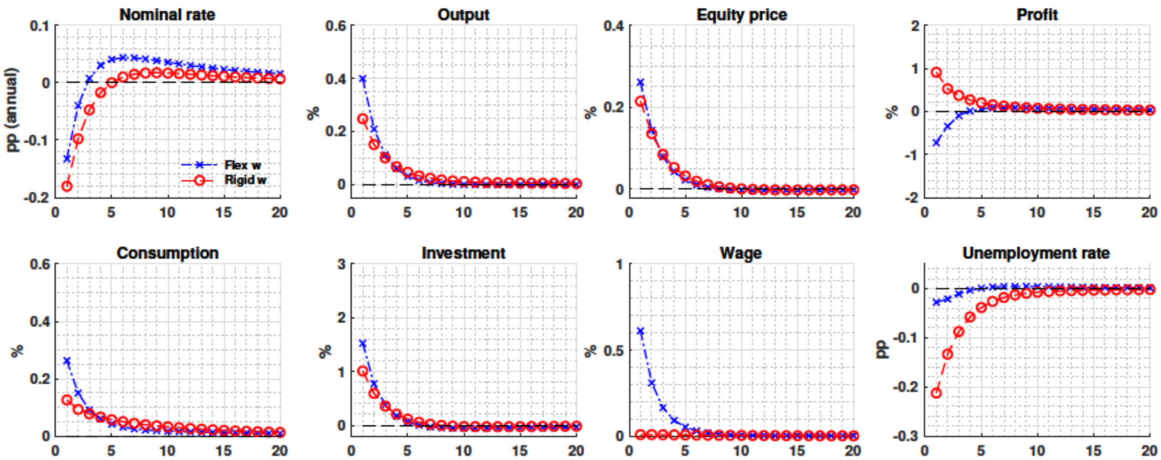
[72]In a standard New Keynesian model, the degree of price rigidity should be high for a monetary policy shock to have real effects. A high degree of price rigidity implies, in the absence of the factor price rigidity, a strong countercylicality of profits or markups, the latter of which has been often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

[73]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.

cost accounts for a significant proportion of the total production cost.[74] The presence of the fixed cost helps the model generate a procyclical profit response as well. What matters for firms' 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.[75] Thus, as Figure A7 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.[76] When the 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.[77] 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.

Figure A8: 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.

F.2 Comparison with a model with the flexible wage

Figure A8 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 A1 shows the values of key parameters at the posterior mode.

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

[75] 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.

[76] 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 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.

[77] 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.

Table A1: 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

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 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.[78]

G 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.

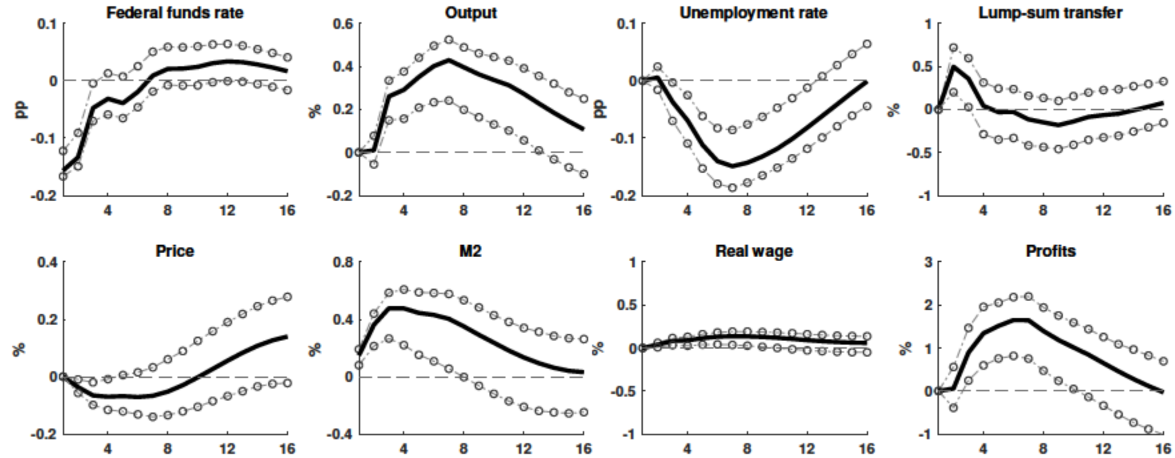
$$\mathrm{FF}_t = f(\Omega_t) + \epsilon_{r,t} \tag{A.45}$$

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.

$$\mathcal{Y}_t = \begin{bmatrix} \log(\text{Output}_t) \\ \log(\text{Price index}_t) \\ \log(\text{Commodity price index}_t) \\ \log(\text{Real wage}_t) \\ \text{Unemployment rate}_t \\ \log(\text{Profits}_t) \\ \log(\text{Lump-sum transfer}_t) \\ \text{FF}_t \\ \log(\text{Total reserves}_t) \\ \log(\text{Non-borrowed reserves}_t) \\ \log(\text{M2}_t) \end{bmatrix} \tag{A.46}$$

[78] 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.

Figure A9: 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.

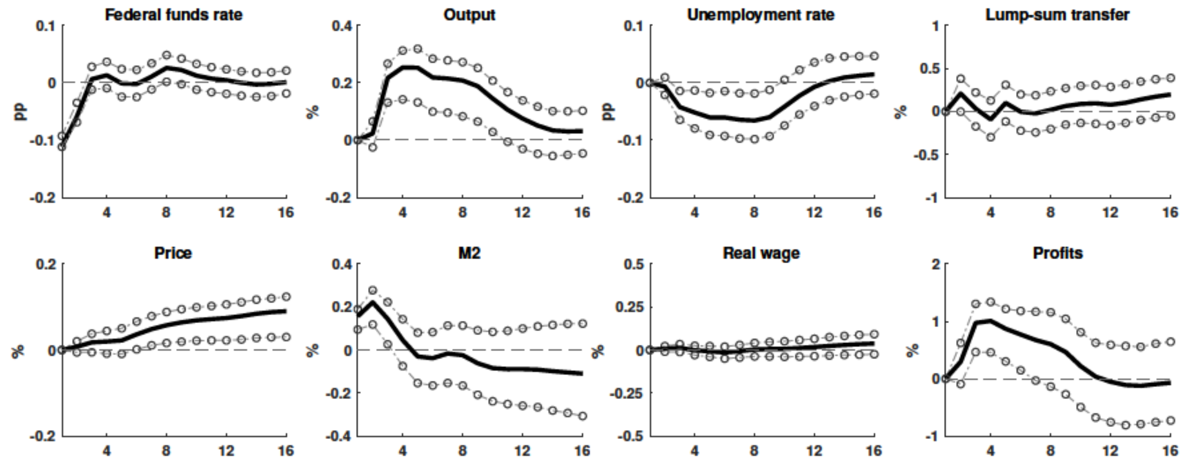
The information set available to the monetary authority includes the data on output, price index, commodity price, index, real wage, unemployment rate, profits, and lump-sum transfer. As in Christiano et al. (1999), I assume that the innovation $\epsilon_{r,t}$ is orthogonal to all variables in the central 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.[79] 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 A9 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 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

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

Figure A10: 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.

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.[80] Most of variables in the SVAR model peaks between 4th and 8th quarters after the shock. In contrast, in the model, responses are immediate.

H Further details on the results

H.1 The decomposition method

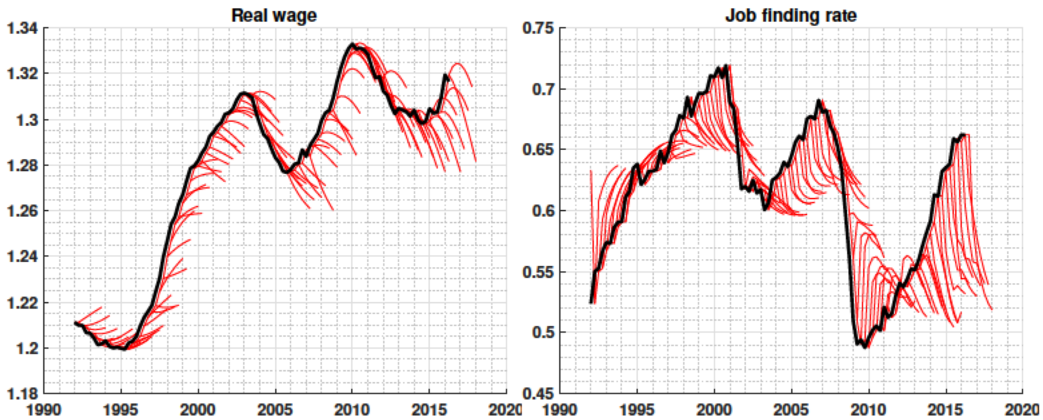
To evaluate the relative contribution of various channels to the evolution of inequality and heterogeneous welfare effects, I compute foresight paths of the following variables each period in the sample .

$$\{ w_{t,t+j} \ , \ i_{t,t+j} \ , \pi_{t,t+j} \ , \ q_{t,t+j} \ , \ r^a_{t,t+j}, \ \Pi_{t,t+j} \ , T_{t,t+j} \ , f_{t,t+j} \}_{j=1}^{N} \tag{A.47}$$

where $x_{t,t+j}$ is the expected value of x in period $t+j$ given the information in period t. N is a very large number that ensures that $x_{t,t+N}$ converges to its steady state value in N

[80]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.

Figure A11: 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.

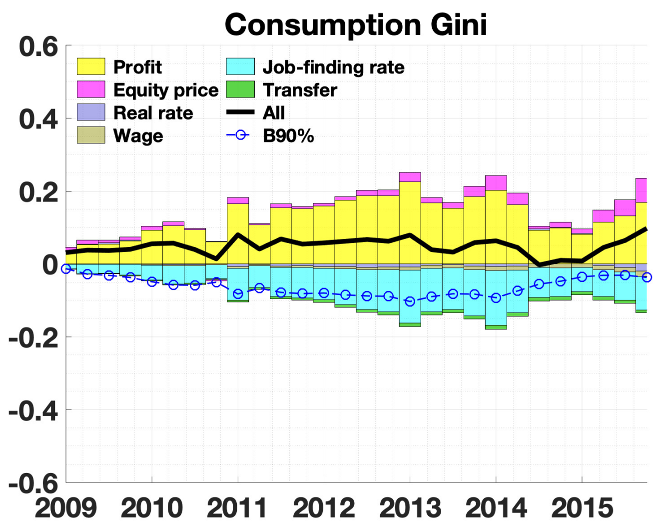
periods. The above eight variables, i.e., real wage, nominal rate, inflation rate, equity price, dividend rates, total lump-sum transfer, and the job-finding rate, are what determine the household's optimal decisions and welfare together with the expected future value (utility) of households' choices. 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.[81] Using different combinations of these paths, I solve the household's problem from $t+N$ 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.[82] Figure A11 shows the realized path of the real wage and the job-finding rate along with each period's household expectations on it.

[81]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. That is, $x_{t,t+1}$ is, in general, different from $x_{t+1,t+1}$.

[82]For the complete decomposition of the households' behavior and the associated welfare, I examine the following eight combinations. In the first combination, all variables follow the path in the counter-factual case. In the second combination, all variables follow the paths in the baseline case. In the third case, only the profit and dividend rates follow the paths in the baseline case, while all others follow paths in the counter-factual case. In the fourth combination, only the nominal rate and the inflation rate follow the baseline paths. In the fifth combination, only the real wage follows the baseline path. In the sixth combination, only the job-finding rate follows the baseline path. In the seventh combination, only the equity price follows the baseline path. Finally, in the eighth combination, only the total transfer follows the baseline path, and all others follow the paths in the counter-factual case.

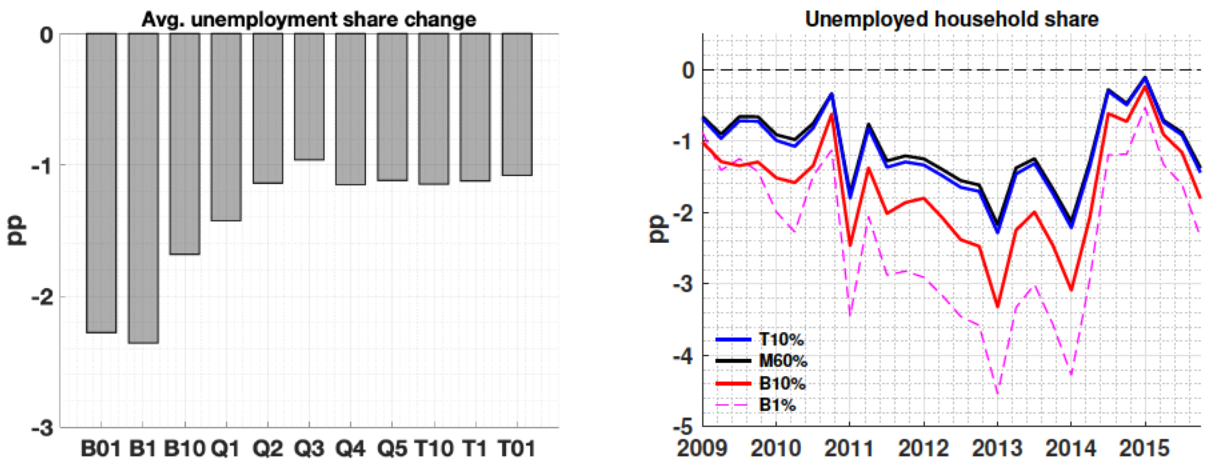
H.2 Additional figures

Figure A12: Distributional effects of UMP: Gini index



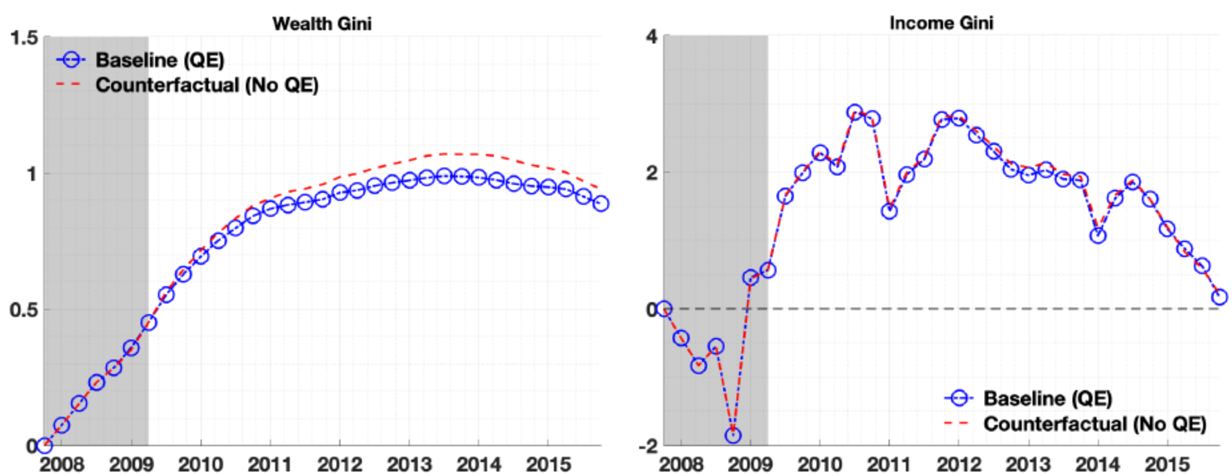
Notes: The figure shows relative degrees of inequality in the model during the ELB episode as differences in the Gini index between the baseline and the counterfactual case. The thick black line shows the overall effects of UMP, while each bar shows the contribution of each variable to the overall effects. 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 A13: Unemployed household shares across wealth groups



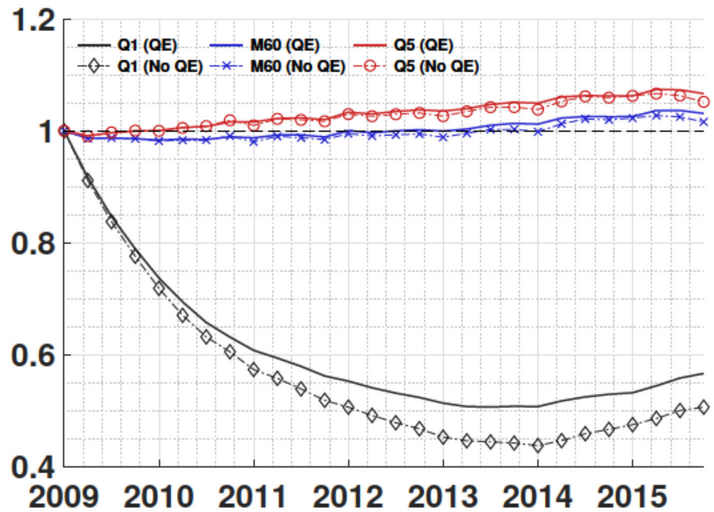
Notes: The left panel shows average changes in the share of unemployed households induced by UMP across wealth groups. The right panel shows the evolution of unemployed household shares during the ELB episode, as percentage point difference from the corresponding values in the counterfactual case with 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 A14: Wealth and income inequality during the ELB period: Gini index



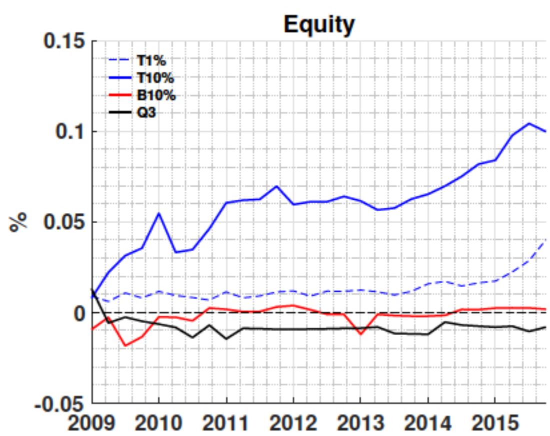
Notes: The figure shows the evolution of the wealth and income Gini indices during the ELB episode, as differnces of the index relative to its 2007 Q4 level. The blue lines with circles show the Gini indices in the baseline case. The dashed red lines show the Gini indices in the counterfactual case.

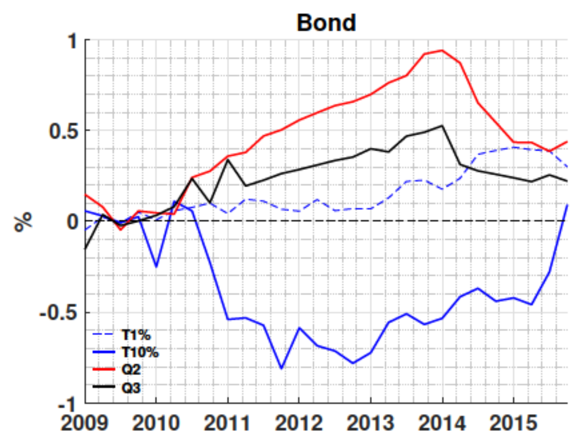
Figure A15: Effects of UMP on households' wealth



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.

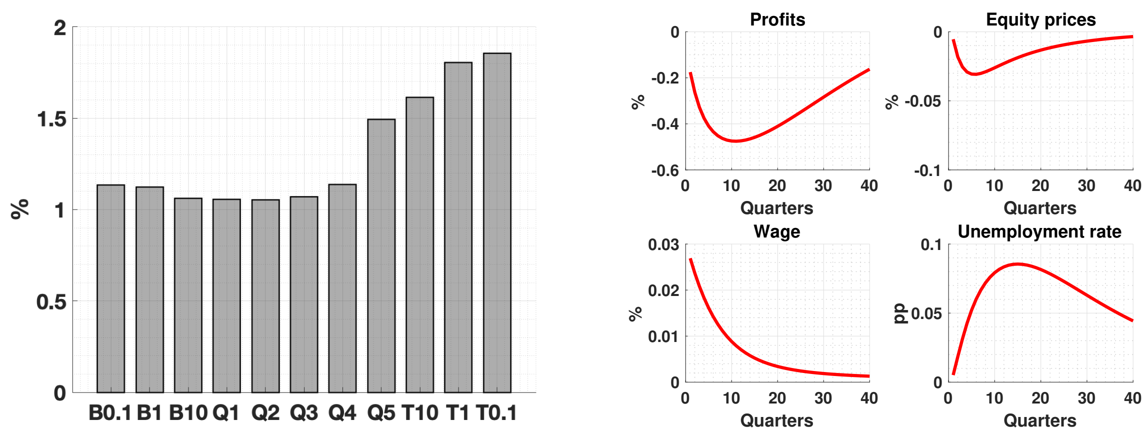
Figure A16: Effects of UMP on equity and bond shares





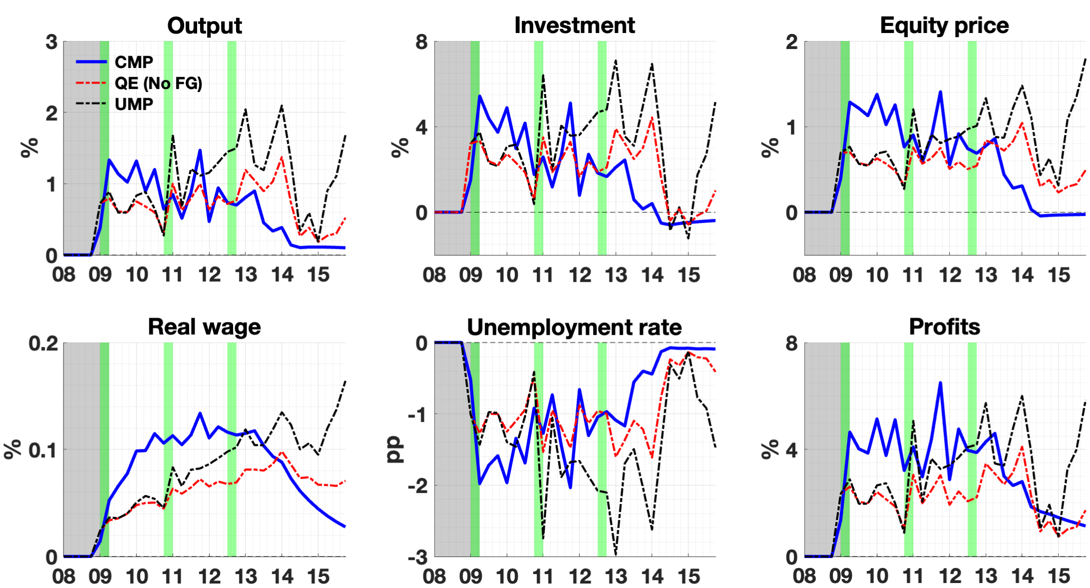
Notes: The figure shows different wealth groups' equity and bond shares during the ELB episode. T1%, T10%, B10%, Q2 and Q3 refers to the top 1% and 10%, the bottom 10%, the second and the middle quintile, respectively. The unit is the difference in the share of equity and bond between the baseline and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

Figure A17: Average consumption gain and households' expectations beyond the sample



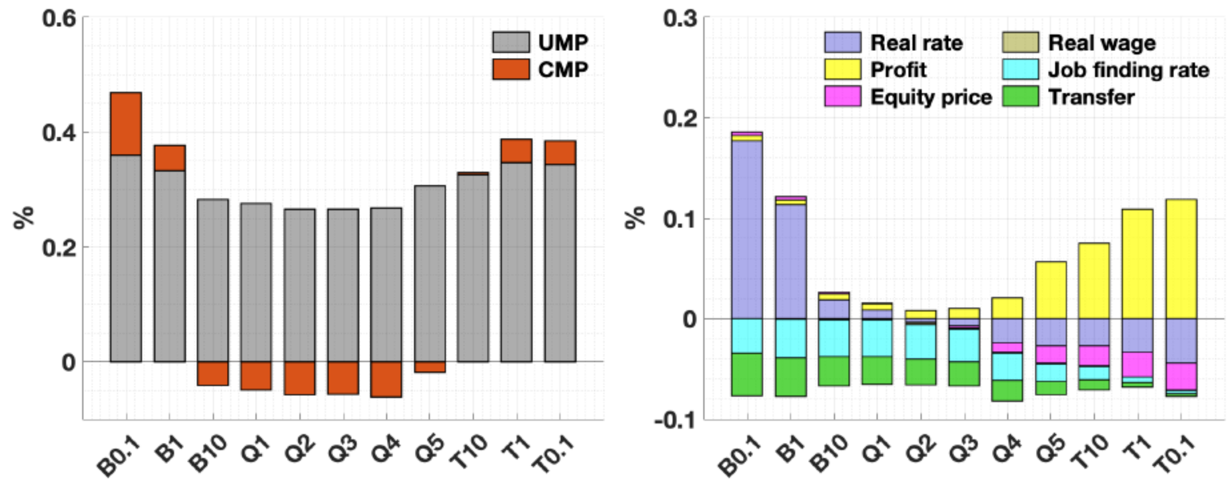
Notes: The left panel shows relative levels of consumption in the baseline case of UMP during the ELB episode, relative to the corresponding consumption levels in the counterfactual case of no UMP across households' wealth groups. The right panel shows households' expectations on profits, equity prices, wages, and unemployment rates from 2019 Q1 onwards.

Figure A18: Aggregate effects of UMP and CMP



Notes: The solid red and blue lines represent the effects of UMP and CMP, respectively. Except for the unemployment rate, all variables are shown as percentage differences from their corresponding values in the counterfactual scenario with no policy interventions. The unemployment rate is displayed as percentage point differences from its corresponding value in the counterfactual case.

Figure A19: Welfare gain comparison: UMP vs CMP



Notes: The figure compares the welfare gains from QE and conventional monetary policy in terms of consumption equivalents. The gray bars represent welfare gains from QE across households, while the dark orange bars indicate the additional gains from conventional monetary policy compared to the welfare gains from QE. The right panel provides a breakdown of the additional gains from CMP, relative to the gains from QE.