Quantitative Easing and Inequality*

Donggyu Lee†
University of Maryland

March 4, 2021
The latest draft is available here

This paper studies how quantitative easing (QE) affects household welfare across the wealth distribution. I build a heterogeneous agent New Keynesian (HANK) model with household portfolio choice, wage and price rigidities, endogenous unemployment, frictional financial intermediation, an effective lower bound (ELB) on the policy rate, forward guidance, and QE. To quantify the contribution of the various channels through which monetary policy affects inequality, I estimate the model using Bayesian methods, explicitly taking into account the occasionally binding ELB constraint and the QE operations undertaken by the Federal Reserve during the 2009-2015 period. I find that QE program unambiguously benefited all households by stimulating economic activity. However, it had non-linear distributional effects. On the one hand, it widened the income and consumption gap between the top 10% and the rest of the wealth distribution, by boosting profits and equity prices. On the other hand, QE shrank inequality within the lower 90% of the wealth distribution, primarily by lowering unemployment. On net, it reduced overall wealth and income inequality, as measured by the Gini index. Surprisingly, QE has weaker distributional consequences compared with conventional monetary policy. Lastly, forward guidance and an extended period of zero policy rates amplified both the aggregate and the distributional effects of QE.

Keywords: unconventional monetary policy, inequality, heterogeneous agent New Keynesian model, quantitative easing, Bayesian estimation, effective lower bound
JEL Codes: E12, E30, E52, E58

*I am very grateful to my advisors Borağan Aruoba, Luminita Stevens, Thomas Drechsel and John Shea for their invaluable guidance and support. I also thank Pierre De Leo, Guido Kuersteiner, Seth Murray, Gonzalo Garcia-Trujillo, Jun Hee Kwak, and Eugene Oue for helpful comments.
†Department of Economics, University of Maryland. Email: dglee81@umd.edu

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 quantitative easing (QE) affects welfare across the wealth distribution. Though effective at stimulating aggregate economic activity, the quantitative easing policies 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] At the same time, the persistent decline in the natural interest rate has increased the likelihood that the economy will often find itself at the effective lower bound (ELB) on the policy rate, and that monetary authorities will need to turn to QE as a stabilization tool.[3] Yet, whether or how QE raises inequality remains an open question and a topic of heated debate.

Gauging QE's distributional effects is a challenging task, as various forces compete in determining the net effects of QE on inequality. First, 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.[4] Finally, higher inflation induced by QE re-distributes wealth from savers to debtors by lowering real rates. A proper evaluation of the net effect of QE on inequality needs to take into account these channels comprehensively.[5]

This paper provides a structural evaluation of the aggregate and distributional consequences of the Federal Reserve's QE program using a medium-scale HANK model that can capture and quantify the dynamics of the channels mentioned above. Two key re-

[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] See, among others, Laubach and Williams (2016) and Holston et al. (2017).

[4] 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 QE has stronger effects on unemployment rates than on real wages, it can reduce income inequality.

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

quirements are necessary for a model to fulfill this task. First, a model should match households' wealth and income composition in the data. Second, it should generate empirically plausible responses of variables that affect households' wealth and income, such as profits, asset prices, wages, and the unemployment rate. The interaction of these two factors, i.e., wealth/income components and their responses to QE, will determine the winners and losers from the QE policy.

To meet the first requirement, the model features portfolio choice and endogenous unemployment. Households can hold two types of assets (deposits/equity), and their working status (employed/unemployed) varies endogenously over time. As in Bayer et al. (2019), I introduce an additional working status under which households receive a fraction of profits as income without supplying labor. Because of these features, households in the model have a heterogeneous composition of wealth (deposits/equity) and income (labor, assets, and business income) and heterogeneous exposures to unemployment risk. In steady state, the top 10% wealthy households hold about 70% of total wealth, mostly in the form of equity, and the sum of business and asset income accounts for about 50% of their total income, consistent with U.S. data. In contrast, households in the lower 80% of the wealth distribution rely mostly on labor income, and a larger share of the households at the bottom of the distribution are 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 for studying QE's distributional effects, 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 procyclical profit response to monetary policy shocks.

[6]According to the 2007 SCF data, labor income, i.e., wage and salary, accounts for about 80% of the total income of the bottom 80% of the wealth distribution, while most of the remaining income consists of transfer income. In stark contrast, for the top 0.1% wealthy households, the share of labor income is only 16%, and transfer income accounts for less than 1% of their total income. The remaining 85% consists mostly of income sources related to profits, i.e., business income and dividends. For the top 10% households, the ratio is 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.

I estimate the model using Bayesian methods to estimate the shock processes that pushed the economy to the ELB 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 QE operations between 2009 and 2015. Because of highly volatile profits and relatively stable and sluggish wage dynamics in the data, the estimated parameter values suggest a high degree of wage rigidity and a relatively flexible extensive margin adjustment of labor supply. Thus, the estimated model generates a strongly procyclical response of profits and employment to conventional monetary policy shocks while the real wage hardly changes, consistent with the existing empirical evidence.

In the model, QE takes the form of the central bank's direct private asset purchase as in Gertler and Karadi (2011). By transforming the demand for a non-productive asset (government bonds or deposits) into a productive asset (capital), QE increases aggregate demand and undoes the contractionary effects of the binding ELB constraint, which are equivalent to those of a series of contractionary interest rate shocks. The model suggests that, between 2009 and 2015, QE on average generated 3.3% and 0.9% higher profits and equity prices, a 1.5% lower unemployment rate, but only 0.1% higher real wages, compared to the counterfactual of no QE 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 and consumption shares increased by 0.17 and 0.06 percentage points during the ELB episode, mainly because of higher profits and equity prices. However, at the same time, QE reduced the wealth and income Gini indices by 0.05 and 0.04 percentage points 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.

Interestingly, the simulation results suggest 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. The difference between QE and conventional monetary policy reflects the dynamics of real rates. Lower real rates disproportionately benefit households at the bottom of the wealth distribution, who are mostly debtors, while hurting the remaining savers. However, the financial sector also substantially benefits from the lower financing costs.[9] The benefits of the financial sector are transferred mostly to the top of the wealth distribution as a part of aggregate profits. Hence, households in the top decile of the wealth distribution end up experiencing higher welfare gains under conventional monetary policy despite the direct welfare loss from lower real rates.

Finally, I find that about half of QE's aggregate effects were due to the expansionary forward guidance and the prolonged periods of zero policy rates. In the model, forward guidance takes the form of exogenously imposing an expected ELB duration, i.e., the number of periods during which the central bank commits to maintaining the policy rate at zero or the effective lower bound, as in Jones (2017). If the expected ELB duration is longer than the duration determined by the policy rule and the expected evolution of the economy, the effects of forward guidance are equivalent to those of anticipated future expansionary interest rate shocks. I estimate an expected ELB duration that is longer than the expected duration implied by fundamentals alone for the entire period between 2009 and 2015. Moreover, I estimate additional stimulus, in the form of rates that were lower for longer: the Fed could have set a positive rate as early as 2014 Q3, but maintained the rate at zero until the end of 2015. Consequently, the economy experienced additional stimulus that accounts for about 55% of the total stimulus effects of unconventional monetary policies. However, forward guidance and the extended periods of zero policy rates also amplified the adverse effects of QE on inequality, further increasing the top 10%'s income share by 0.09 percentage points on average.

Related Literature

(Empirical literature)

My work contributes to a growing literature studying the distributional consequences of conventional and unconventional monetary policy measures. Regarding conventional monetary policy, Coibion et al. (2017) show, using the Consumer Expenditure Survey (CEX), that a conventional monetary tightening raises the measures of income and consumption inequality, including the Gini index and the inter-decile ratio. Mumtaz and Theophilopoulou (2017) and Furceri et al. (2018) find similar results for the U.K. and for a panel of advanced and emerging countries, respectively. In contrast, Inui et al. (2017) show, using the microdata on households' income in Japan, that an expansionary inter-

[9] QE boosts banks' net worth by increasing equity prices. However, simultaneously, it decreases the expected gross rate of return on their assets, i.e., profitability. In contrast, conventional monetary policy does not directly increase banks' net worth, but in the absence of general equilibrium feedbacks, it increases the profitability of banks by lowering their financing costs.

est rate shock does not have a significant impact on income inequality. Using structural VAR models, Davtyan (2017) and Hafemann et al. (2018) find that a contractionary (expansionary) interest rate shock reduces (increases) income inequality, as measured by the Gini index. My results underscore the sensitivity of conclusions to the inclusion of the wealthiest households: to the extent that the CEX under-represents this group, it reflects what the model predicts for the lower 90% of the wealth distribution, for whom it is indeed the case that tightening raises inequality.

In the case of unconventional monetary policies, the empirical literature is relatively limited, and the results are even more mixed. In the U.S., Bivens (2015) argues that QE significantly reduced income inequality by helping the economy achieve full employment, but had negligible effects on wealth inequality. Conversely, Montecino and Epstein (2015) examine the contribution of QE to the observed changes in households' income in 2010 and 2013 SCF data, and they conclude that QE increased income inequality mainly via its strongly positive impact on equity returns.

There are also cross-region differences in the results. Saiki and Frost (2014) show that an increase in assets held by the Bank of Japan (BOJ) raised income inequality in Japan, as measured by the top to the bottom quintile income ratio. Similarly, Taghizadeh-Hesary et al. (2020) argue that BOJ's unconventional monetary policies, including zero and negative interest rates, raised the top 10% to the bottom 10% income ratio. The existing work on European countries suggests that QE reduces income inequality mainly via its positive impact on labor markets. Casiraghi et al. (2018) estimate the aggregate effects of QE using the semi-structural model of the Italian economy and apply them to households' distribution in the data. They conclude that QE mostly benefits households at the bottom of the wealth and income distribution because their earnings respond more strongly and positively to QE operations. Using a similar methodology, Lenza and Slacalek (2018) also find that the ECB's QE programs reduce the income Gini index in European countries mainly by lowering unemployment rates, while leaving the wealth Gini largely unchanged. In contrast, Bank of England (2012) and Domanski et al. (2016) find that QE increases wealth inequality mainly by increasing equity prices.

Competing results in the empirical literature show the importance of a comprehensive examination of competing channels as well as the consideration of country-specific characteristics, such as the preexisting degree of inequality, the flexibility of labor markets, and the structure of equity markets and the financial sector. I therefore adopt a dynamic structural general equilibrium (DSGE) analysis approach to examine the impact of QE on inequality. In terms of the methodology, my work is closely related to Casiraghi et al. (2018) and Lenza and Slacalek (2018) since I estimate the aggregate effects of QE and apply them to the household distribution to evaluate the contribution of each channel as well as the overall distributional consequences. However, while they estimate the aggregate and distributional consequences of QE separately with an auxiliary assumption on how the aggregate effects are distributed across the distribution of households, I evaluate both aggregate and distributional effects at the same time with a unified framework. Moreover, I compare the effects of QE with those of conventional monetary policy and the other type of unconventional monetary policy, i.e., forward guidance.

(Theoretical literature)

Regarding my theoretical contribution, this paper lies at the intersection of three key literatures: 1) the literature studying macroeconomic fluctuations using a HANK framework, 2) the literature studying unconventional monetary policies, and 3) the literature on Bayesian estimation of DSGE models.

In the HANK literature, my model builds on existing work that has focused on the transmission mechanisms or distributional consequences of conventional monetary policy shocks. The seminal paper of Kaplan et al. (2018) shows that the presence of wealthy hand-to-mouth households drastically changes transmission mechanisms of monetary policy shocks in a HANK model compared to those in a representative agent New Keynesian (RANK) model. Similarly, Luetticke (2020) and Auclert et al. (2020b) examine the transmission mechanisms of conventional monetary policy shocks and emphasize the importance of investment in the transmission of these shocks. Unlike these papers, Gornemann et al. (2016) examine the distributional consequences of interest rate shocks and find that a contractionary monetary policy shock increases inequality by reducing employment while increasing the return on assets.

My work is also closely related to the literature studying unconventional monetary policies. The model in this paper draws on Gertler and Karadi (2011), who propose a framework for the analysis of a central bank's large scale asset purchase program, which features frictional financial intermediation to break the irrelevance result of Wallace (1981). Also, as in Del Negro et al. (2017) and Chen et al. (2012), I evaluate the effects of the Federal Reserve's large-scale asset purchase programs during the ELB episode, but the focus is on distributional consequences rather than the programs' aggregate effects. Jones (2017) studies the effects of forward guidance in the form of exogenous ELB durations, and I adopt the same approach in simulating the ELB episode.

In addition, Hohberger et al. (2020) study the distributional consequences of unconventional monetary policy, using a small open economy two agents (saver-spender) New Keynesian model (TANK). They conclude that an expansionary QE shock reduces income inequality, measured by the income shares of the two agents, in the medium-run. Cui and Sterk (2018) also examine the effects of QE but focus on the efficacy and the stability of the policy as a regular monetary policy tool rather than on its distributional effects.

Finally, 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. I contribute to this literature by estimating a HANK model with an occasionally binding constraint and unconventional monetary policies. 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 counterfactuals to examine the aggregate and distributional effects of QE during the ELB episode. Section 5 compares the effects of QE with the effect of conventional monetary policy. Section 6 examines the aggregate and distributional effects of forward guidance during the ELB episode. Section 7 checks the robustness of the results. Section 8 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, and wage and price rigidities. These features help the model generate empirically plausible dynamics for key aggregate variables such as wages, unemployment, equity prices, and profits. They also shape the relative contributions of different transmission channels of unconventional monetary policy and determine its distributional consequences.

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 are employed, unemployed, or business owners. There are exogenous and endogenous transitions between working status, as I will explain below.

Households derive utility from consumption and disutility from labor, and die with exogenous probability ζ each period.[10] Conditional on surviving, they discount the future at rate $\beta \in (0,1)$, solving

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

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.[11] $\mathbb{1}_{\{a_{it+1} \neq a_{it}\}}$ is an indicator function equal to 1 in periods in which the household changes its holdings a_{it} of illiquid assets. 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 of labor supply, and ψ is a parameter that scales the steady state hours worked to 1. As in Bayer et al. (2019), I assume that the disutility from supplying labor is proportional to the productivity level. Under this preference specification, all employed households choose the same amount of hours worked, as a function solely of the real wage.[12] This feature significantly facilitates the computation.

[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.

[11] Without adjustment costs, both assets become perfect substitutes.

[12] Auclert et al. (2020a) point out that the consumption-labor complementarity embedded in GHH preferences leads to unreasonably high fiscal and monetary policy multipliers in a model with frictionless labor markets and flexible wages. In contrast, Broer et al. (2019) show that, under the separable preferences, a countercyclical response of profits to accommodative monetary policy can lead to an undesirable amplification because of its negative income effect on labor supply. Since my model features frictional labor markets and sticky wages and generates procyclical responses of profits to demand shocks, it is free from

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

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

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

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 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.[13] The period income y_{it} depends on the household's working status,

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

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. Idiosyncratic productivity 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 υ. 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 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.[14]

Households transfer wealth inter-temporally via two assets. Liquid assets b are subject to an exogenous borrowing limit $\underline{b}$, and pay a real rate that depends on whether the

---

these problems.

[13]In the model, taxes are levied on households' income and firms' profits. For simplicity, I assume that tax rates are the same for both types of tax bases. Thus, I use the same notation for both types of taxes.

[14]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.

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 return r_t^a.[15] Adjusting illiquid asset holding incurs a utility cost, which, for tractability, I assume is stochastic.[16] Following Bayer et al. (2019), the independently and identically distributed adjustment costs are drawn from the logistic distribution, with cumulative probability

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

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

2.2 Final good firm

The final good is a standard CES aggregator,

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

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

[15] In Kaplan et al. (2018), households do not earn any income from their illiquid assets if they do not make withdrawals. Instead, the dividends are automatically re-invested in the illiquid asset. This difference in the treatment of illiquid asset returns is mainly due to the difference in our perspective on illiquid asset. Kaplan et al. (2018) view housing as the major type of illiquid assets. Since most housing is for residential purposes, it is natural to assume that households receive no pecuniary income from their own house. However, in this paper, I focus on claims on profits, such as stocks and proprietorship of a business, as the main type of illiquid asset. In the case of housing, I mainly focus on property that yields rental income, which is disproportionately held by wealthier households in the data. Therefore, it is natural to assume that households enjoy a liquid stream of dividends by holding illiquid assets in my model.

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

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 in a monopolistically competitive environment. The production technology is characterized by a standard Cobb-Douglas 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).

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

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. As I will discuss later, the fixed cost of operation plays an important role in generating procyclical responses of profits to demand shocks in the model. The fixed cost is a random proportion of the steady state output and follows an AR(1) process.

Firms use business owners' average marginal rate of substitution, which is denoted by $\Lambda_{t,t+1}$, to discount future cash flows. Price adjustment costs a lá Rotemberg (1982) are 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.[17]

The first order conditions for labor and capital rental services are standard and given by

$$r_t^k = \theta \mathrm{MC}_t\left(\frac{Y_{jt}}{K_{jt}}\right), \ \ r_t^l = (1-\theta)\mathrm{MC}_t\left(\frac{Y_{jt}}{L_{jt}}\right) \ , \tag{14}$$

where MC_t is the Lagrange multiplier on the production constraint of the cost minimization problem, which represents the real marginal cost,

$$\mathrm{MC}_t = \frac{(r_t^k)^\theta (r_t^l)^{1-\theta}}{Z_t}\left(\frac{1}{\theta}\right)^\theta \left(\frac{1}{1-\theta}\right)^{1-\theta} \ , \tag{15}$$

The optimality conditions regarding firms' price setting under the symmetric equilibrium assumption yield the following Phillips curve.

$$\log \pi_t - \log \pi_{t-1}^{\iota_p} \pi^{1-\iota_p} = \mathbb{E}_t\Big[\Lambda_{t,t+1}\Big(\log \pi_{t+1} - \log \pi_t^{\iota_p}\pi^{1-\iota_p}\Big)\Big] + \kappa\left(\mathrm{MC}_t - \frac{1}{\Psi_t^p}\right), \tag{16}$$

where $\Psi_t^p = \frac{\eta_t}{\eta_t - 1}$ 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.

$$(r_t^l - w_t - \Xi^L) s_{it} n_{it} \tag{17}$$

where Ξ^L is the cost for maintaining a match.[18]

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

[17] 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.

[18] 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.

sume a function of the form.[19]

$$\frac{w_t}{w} = \left\{\epsilon_{w,t}\left(\frac{r_t^l}{r^l}\right)\right\}^{(1-\rho_w)} \times \left\{\frac{w_{t-1}}{w} \times \left(\frac{\pi_{t-1}}{\pi_t}\right)^{\iota_w} \times \left(\frac{\pi}{\pi_t}\right)^{1-\iota_w}\right\}^{\rho_w} \quad , \quad 0 < \rho_w \; , \; \iota_w < 1 \tag{18}$$

Equation (18) 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.[20] 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.[21] 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.[22]

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

$$J^L(s_{it}) = (r_t^l - w_t - \Xi^L)s_{it}n_{it} + \mathbb{E}\Big[\Lambda_{t,t+1}(1-\zeta)(1-\lambda)(1-P_e)J^L(s_{it+1})\Big] \; , \tag{19}$$

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,

$$\iota = \frac{M_t}{V_t}\int J^L(s_t)d\mu_t(s_{it}) \; , \tag{20}$$

[19]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 infeasible. 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.

[20]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.

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

[22]Assuming $\epsilon_{w,t}$ as an autoregressive process does not affect the estimation results. Thus, to reduce the number of parameters to be estimated, I assume that the wage shock is i.i.d. Note that even though the shock is i.i.d, it propagates as the wage exhibits inertia.

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 following matching function

$$M_t = \frac{\Big(U_t + \lambda N_t\Big)V_t}{\left\{\Big(U_t + \lambda N_t\Big)^{\alpha} + V_t^{\alpha}\right\}^{\frac{1}{\alpha}}} \quad , \quad \alpha > 0 \quad , \tag{21}$$

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.[23] 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.[24] For a given capital stock K_t, the capital firm earns the following profit

$$r_t^k v_t K_t - \delta(v_t) K_t \quad , \tag{22}$$

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,

$$r_t^k = \delta'(v_t) \tag{23}$$

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

$$\delta(v_t) = \delta_0 v_t^{\delta_1} \quad , \quad \delta_1 > 1 \; , \tag{24}$$

[23] 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.

[24] 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.

where δ_0 is the depreciation rate under full utilization and δ_1 governs the degree of acceleration of depreciation.[25]

Regarding capital accumulation, I assume that the capital firm purchases new capital from its investment department on behalf of investors.[26] The investment department has 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.

$$q_t K_{t+1} - \Psi_t^k \left\{ K_{t+1} + \frac{\phi}{2} \left(\log \frac{K_{t+1}}{K_t} \right)^2 K_{t+1} \right\}, \tag{25}$$

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.

$$q_t = \Psi_t^k \left\{ 1 + \phi \log \frac{K_{t+1}}{K_t} + \frac{\phi}{2} \left(\log \frac{K_{t+1}}{K_t} \right)^2 \right\} - \mathbb{E}_t \left[\Lambda_{t,t+1} \Psi_{t+1}^k \phi \left(\log \frac{K_{t+2}}{K_{t+1}} \right) \frac{K_{t+2}}{K_{t+1}} \right] \tag{26}$$

Finally, investment expenditure is defined as

$$\tilde{I}_t = \Psi_t^k K_{t+1} \left\{ 1 + \frac{\phi}{2} \left(\log \frac{K_{t+1}}{K_t} \right)^2 \right\} - \{1 - \delta(v_t)\} K_t \tag{27}$$

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

[25] As I follow standard practice and set the steady state utilization rate to 1, δ_0 is equal to the steady state depreciation rate.

[26] For simplicity, I assume that these two entities operate independently from each other. Thus, one does not take into account the effects of its own decision on the other.

cash-flow constraint of the equity mutual fund is as follows:

$$(1-\tau)(1-\nu)\Pi_t - q_t(K_{t+1}-K_t) + q_t(A_{t+1}-A_t) = r_t^a A_t \quad , \tag{28}$$

where Π_t is the sum of all firms' profits and ν is the share of profits that is given to business owners.[27] The tax rate on firms' profits is denoted by τ. Given that the amount of aggregate 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

$$r_t^a = (1-\tau)(1-\nu)\Pi_t/K_t \ , \tag{29}$$

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

In practice, central banks' unconventional liquidity provision takes the form of purchases of assets held by financial institutions. To introduce such a policy into the model, I model banks as in Gertler and Karadi (2011).

There is a continuum of banks indexed by $j \in (0,1)$. Each bank takes deposits from savers and purchase equity. Bank j's balance sheet is given by

$$q_t A_{jt+1}^b = N_{jt} + B_{jt+1}^b \quad , \tag{30}$$

where A_{jt+1}^b and B_{jt+1}^b are bank j'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.

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

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.

As in Gertler and Karadi (2011), each period, only a θ_b fraction of banks continue to operate, while the remaining $1-\theta_b$ fraction exit the market. Let $J^b(N_{jt})$ denote the value of a surviving bank j. Under the environment described so far, the value of bank j is

[27] 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.

given by

$$J^b(N_{jt}) = \max_{\{A^b_{jt+1}, B^b_{jt+1}, N_{jt+1}\}} \mathbb{E}_t \Big[\Psi^b_t \Lambda_{t,t+1} \Big\{ (1-\theta_b) N_{jt+1} + \theta_b J^b(N_{jt+1}) \Big\} \Big] \tag{32}$$

s.t.

$$q_t A^b_{jt+1} = N_{jt} + B^b_{jt+1} \quad , \quad N_{jt+1} = R^a_{t+1} q_t A^b_{jt+1} - R_{t+1} B^b_{jt+1} \tag{33}$$

$$J^b(N_{jt}) \geq \Delta q_t A^b_{jt+1} \tag{34}$$

where Ψ^b_t denotes the aggregate risk premium shock, which follows an AR(1) process.[28] Like other firms in the model, banks are owned by the equity mutual fund and discount future cash flows, using the average MRS of business owners $\Lambda_{t,t+1}$. Howevevr, as shown in the above equation, banks' discount factor is perturbed by an exogenous risk premium shock Ψ^b_t. When a positive risk premium shock occurs, banks value future returns more and thus, demand more assets for a given equity premium.

Equitation (34) is the incentive compatibility constraint, which reflects a moral hazard problem assumed in Gertler and Karadi (2011). Specifically, banks would purchase assets to the point that the equation (34) holds with equality.[29] With a guess and verify approach, one can show that a bank j's value has the following expression.

$$J^b(N_{jt}) = \vartheta^a_t q_t A^b_{jt+1} + \vartheta^n_t N_{jt} \tag{35}$$

where ϑ^a_t and ϑ^b_t are the expected value of assets and net-worth, respectively.[30] Given that the incentive constraint always binds, as in Gertler and Karadi (2011), I have the following relationship between the amount of asset purchases and a bank's net-worth.

$$q_t A^b_{jt+1} = \frac{\vartheta^n_t}{\Delta - \vartheta^a_t} N_{jt} = \Theta_t N_{jt} \tag{36}$$

where Θ_t is the leverage ratio of banks. Since the leverage ratio does not depend on bank-specific variables, the above relation can be aggregated. That is, I have

$$q_t A^b_{t+1} = \Theta_t N_t \tag{37}$$

where A^b_{t+1} and N_t are the financial sector's equity holding and net-worth, respectively.

Given the law of motion for individual bank's net-worth, exogenous survival rate, and

[28]The shock plays a similar role to the role of risk premium shocks in Smets and Wouters (2007) and the same shock applies to all banks.

[29]For a more detailed description of the incentive problem, see the appendix.

[30]For detailed description of these variables, see the appendix.

the assumption that existing banks are replaced with new banks with a seed fund given by the equity mutual fund, the law of motion for the aggregate net-worth of banks can be described as follows.

$$N_t = \theta_b\Big\{(R_t^a - R_t)\Theta_{t-1} + R_t\Big\}N_{t-1} + \omega q_{t-1}A_t^b \tag{38}$$

where the last term is the seed fund for new banks, which is a ω fraction of the existing banks' asset holdings.

Finally, profits from the financial sector are the sum of the net-worth of existing banks, net of the seed fund given to new banks. For a more detailed description of the banking sector, see the appendix.

$$\Pi_t^b = (1-\theta_b)\Big\{(R_t^a - R_t)\Theta_{t-1} + R_t\Big\}N_{t-1} - \omega q_{t-1}A_t^b \ , \tag{39}$$

2.8 Money market mutual fund

The model features another type of hypothetical mutual fund, which I call the money market mutual fund. Its main role is to provide liquidity to the financial sector.[31] It receives contributions from the government and invests in liquid assets. With these contributions and the proceeds from its assets, the fund makes lump-sum transfers to households. Specifically, I assume that the fund smoothes out the flow of lump-sum transfers with the following objective.

$$\max_{\{T_t^m, B_{t+1}^m\}} \mathbb{E}_0\left[\sum_{t=0}^{\infty} \Psi_t^l \beta_m^t \frac{(T_t^m)^{1-\sigma}}{1-\sigma}\right] \tag{40}$$

subject to

$$T_t^m + B_{t+1}^m = C_t^g + (1+r_t^b)B_t^m \ , \tag{41}$$

[31] Since banks are levered investors, there should be an equivalent amount of liquid assets that correspond to banks' illiquid asset holdings in the model. If I assume that households are the sole entities that provide funds to banks, the share of liquid assets in households' portfolio should be high. However, the SCF data shows that households hold only about 10% of their total assets as liquid assets. Moreover, according to the Financial Account data (previously known as 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 includes deposits, bonds, open market paper, loans, and other liabilities, has been about 25% since 2000. Based on these facts, I assume that there is a significant non-household liquidity provider, in the form of the money market mutual fund.

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.[32] The contribution that the fund receives from the 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.[33] Finally, the MMMF is subject to an AR(1) liquidity preference shock Ψ_t^l.[34]

2.9 Monetary authority

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

$$i_{t+1} = \min\{0, \hat{i}_{t+1}\} \ \text{ with } \ \frac{1+\hat{i}_{t+1}}{1+\hat{i}} = \left(\frac{1+\hat{i}_t}{1+\hat{i}}\right)^{\rho_R} \left[\left(\frac{\pi_t}{\pi}\right)^{\phi_\pi} \exp\{-\phi_u(u_t - u)\}\right]^{1-\rho_R} \exp(\epsilon_{R,t}), \quad (42)$$

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 effective lower bound. Thus, if the shadow rate becomes negative, the policy rate can no longer respond to inflation and the unemployment gap.[35] However, in the model, the central bank can affect the economy even when the ELB binds, by adjusting its holdings of illiquid assets.

Central bank asset purchases (QE) have been modeled in different ways in the literature. For instance, Gertler and Karadi (2011) modeled QE as the central bank's direct purchase of private assets, i.e., capital, following a rule based on the equity premium. In contrast, Chen et al. (2012) modeled QE as an exchange between long-term and short-term government debt and did not propose any rule.

In this paper, I follow Gertler and Karadi (2011) and assume that the central bank directly purchases illiquid assets from the private sector. However, unlike Gertler and

[32] 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.

[33] 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.

[34] If a positive liquidity preference shock occurs, the MMMF increases liquid asset investment and reduced transfers. Thus, consumption and the inflation rate fall.

[35] The effective lower bound 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 effective lower bound is zero.

Karadi (2011), I model QE as a purely discretionary policy,

$$A_{t+1}^{\text{CB}} = \Psi_t^{\text{QE}} A^{\text{CB}} \ , \ B_{t+1}^{\text{CB}} = q_t A_{t+1}^{\text{CB}} \ , \tag{43}$$

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. As in Gertler and Karadi (2011), the central bank issues government bonds to finance its asset purchases.[36] 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 exogenous expected ELB durations if the economy is at the ELB.[37] 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 central bank's policy rule), 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

[36]As the formulation of QE policy implies, the central bank transforms demand for non-productive liquid assets (bonds) into demand for productive illiquid assets (capital/equity). Thus, QE policy directly increases investment. By increasing investment, QE policy also increases equity prices, which inflates banks' net worth. As banks' investment is proportional to their net worth, an increase in net worth can increase banks' equity purchases. However, a rise in equity prices also implies a fall in the expected gross rate of return on equity, which discourages equity investment. Similarly, an increase in equity prices can increase consumption via wealth effects. But, a fall in expected returns can reduce households' equity investment. The general equilibrium effects of QE depend on the relative magnitudes of these forces.

[37]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.

its debt according to the following simple autoregressive rule, as in Woodford (1995).

$$\frac{B^g_{t+1}}{B^g} = \left(\frac{R_t/\pi_t \times B^g_t}{R/\pi \times B^g}\right)^{\rho_B} \ , \ 0 \leq \rho_B < 1 \ , \tag{44}$$

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.

$$B^g_t = \sum_{l \geq t}^{\infty} \left\{ \prod_{i=t}^{l} \left(\frac{\pi_i}{R_i}\right) \right\} \left\{ \mathbb{T}_l - (G_l + T^g_l + D_l + T^{CB}_l + C^g_l) \right\} \ , \tag{45}$$

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. $T^{\text{CB}}_t = q_t A^{\text{CB}}_{t+1} - (q_t + r^a_t) A^{\text{CB}}_t + R_t B^{\text{CB}}_t - B^{\text{CB}}_{t+1}$ is the transfer from (or to) the monetary authority.

Equation (45) 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 according to the following stochastic process.[38]

$$T^g_t = \left(1 - \frac{1}{\Psi^g_t}\right) Y \ , \tag{46}$$

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

[38]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.

2.11 Market clearing conditions

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

$$\underbrace{\underbrace{A^h_{t+1}}_{\text{households}} + \underbrace{A^b_{t+1}}_{\text{banks}} + \underbrace{A^{CB}_{t+1}}_{\text{central bank}}}_{\text{equity demand}} = \underbrace{K_{t+1}}_{\text{equity supply}} , \tag{47}$$

where $A^h_{t+1} = \int a_{t+1} d\mu_t$ 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.[39]

$$\underbrace{\underbrace{B^h_{t+1}}_{\text{households}} + \underbrace{B^m_{t+1}}_{MMMF}}_{\text{liquid asset demand}} = \underbrace{\underbrace{B^b_{t+1}}_{\text{bank deposits}} + \underbrace{B^g_{t+1}}_{\text{gov. bond}} + \underbrace{B^{QE}_{t+1}}_{\text{bond for QE}}}_{\text{liquid asset supply}} , \tag{48}$$

where $B^h_{t+1} = \int b_{t+1} d\mu_t$ 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 period must equal the capital services demanded by the intermediate goods producers:

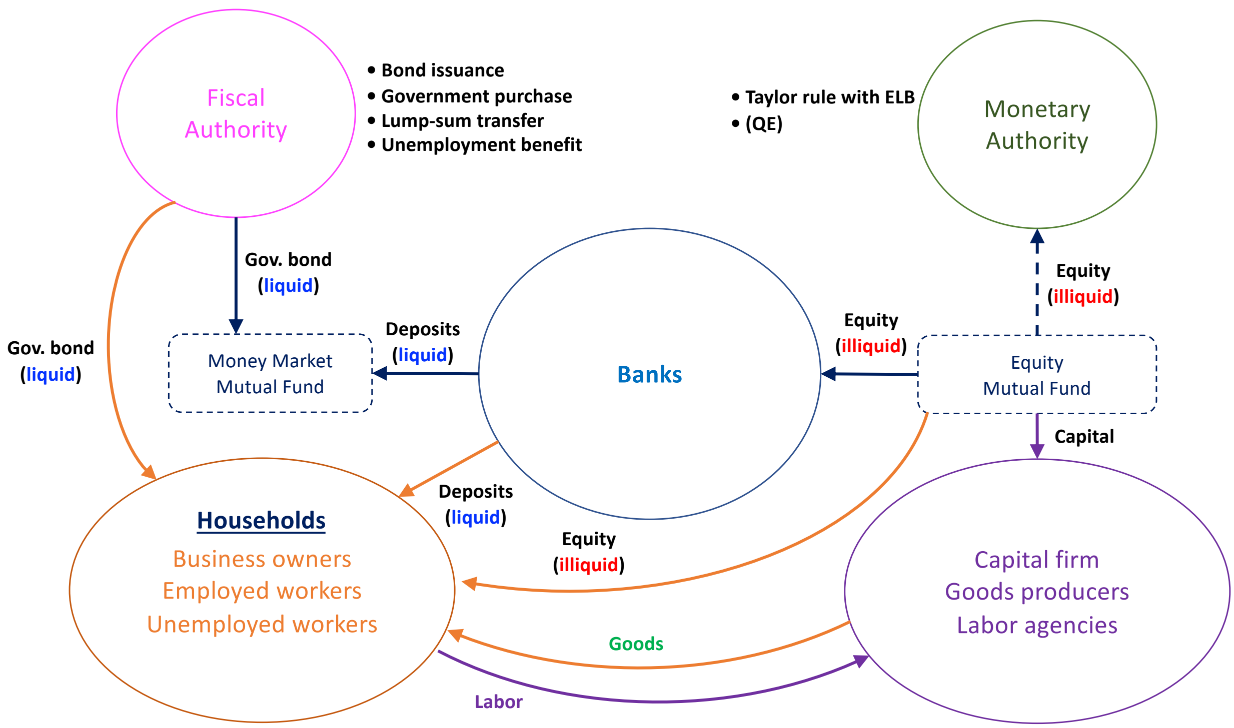
$$v_t K_t = \int_0^1 K_{j,t} dj = K^I \tag{49}$$

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

$$\int \mathbb{1}_{\{e_t=1\}} s_t n_t d\mu_t = \int_0^1 L_{j,t} dj \, . \tag{50}$$

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

Figure 1: The model overview



Notes: The figure shows economic agents in the model and the goods and asset flows among them. Arrows start from the entity that sells a certain good (good or labor) or an asset (liquid or illiquid). The dashed line indicates transactions associated with QE operations.

If the above-mentioned markets clear, by Walras' law, the goods market also clears. Figure 1 summarizes the model.

2.12 Solution method

I solve the model using a perturbation method developed by Reiter (2009) and extended by Winberry (2018) and Bayer and Luetticke (2020). First, I solve the steady state of the model using an endogenous grid method developed by Carroll (2006). Then, I linearize the model around the steady state and apply a perturbation method. However, since the model features many idiosyncratic states, i.e., liquid and illiquid asset holdings, skill level, and working status, the linearized system's dimension is too large. Thus, a state-space reduction is required. For the state-space reduction, I adopt the method used by Bayer et al. (2019) and Bayer and Luetticke (2020). For the value function, I use Chebyshev polynomials with sparse grids to approximate deviations from the steady state. For the idiosyncratic distribution, I use a fixed copula, as suggested by Bayer and Luetticke

(2020). I assume a time-invariant functional relationship between the joint and marginal distributions and use it to approximate the evolution of the distribution.

Also, to handle the occasionally binding constraint on the policy rate, I adopt the methodology of Kulish et al. (2014) and Jones (2017). Specifically, I treat the model with a binding ELB as a temporary alternative regime, while treating the model with a positive policy rate as a reference regime, and assume exogenous durations of the alternative regime during the estimation. For comparison with the results from the model with endogenous ELB durations, I adopt the method of Guerrieri and Iacoviello (2015). Further details on the numerical method and its application during the estimation procedure can be found in the appendix.

3 Calibration and Estimation

I adopt a two-stage approach to parametrize the model. First, I set a subset of parameters so that the model's steady state matches moments of households' wealth distribution and income composition in the micro-data. I then estimate the remaining parameters with full information Bayesian methods, using time-series data on aggregate macro variables. Importantly, I explicitly take into account the incidence of the binding ELB constraint and QE operations during the estimation.

3.1 Data for calibration

For the calibration, I mainly use data from the Survey of Consumer Finances (SCF), since it has detailed information on households' wealth and income composition.[40] I use the 2007 SCF, as it is the last survey before the implementation of QE.[41]

To map the model to the data, I first define liquid assets in the data as the sum of checking, savings and money market deposits, call accounts, certificates of deposit, bonds net of credit card balances, and other lines of credit. Illiquid assets are defined as the sum of all financial assets other than liquid assets, plus net housing wealth, business interest in corporate and non-corporate businesses, minus installment loans. I include only 40% of net housing wealth in illiquid assets, following Kaplan et al. (2018), to take into ac-

[40] The SCF is well-suited for the study of inequality as it over-samples wealthy households. Specifically, two-thirds of respondents comprise a representative sample of U.S. households, while the remainder of respondents are over-sampled from wealthy households.

[41] Using a more recent survey for computing target moments would not lead to significant differences in parametrization as different surveys produce similar target moments.

Table 1: 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

count the presence of assets that are owned for purely residential purposes. Consumer durables, such as vehicles net of non-revolving debt, are also excluded from illiquid assets. Households with negative illiquid assets are excluded, as short positions in the illiquid asset are not allowed in the model.[42]

I decompose income into three categories; labor income, capital income, and transfer income. In the data, wages and salaries constitute labor income. I define capital income as the sum of business income and asset income.[43] Business income consists of profits from running businesses or farms. Asset income includes fixed interest on financial assets, dividends, and capital gains. Transfer income consists of miscellaneous transfer income, social security benefits, and pension income.

As Kaplan et al. (2018) have shown, the household wealth distribution matters for the responsiveness of a HANK model to monetary policy shocks. Hence, I target moments related to the household wealth distribution, such as the shares of borrowers, wealthy hand-to-mouth households, and households with zero assets.[44] However, gains and losses from monetary policy ultimately depend on households' income composition and the relative response of each income component to monetary policy.

Tables 1 and 2 show moments in the data and their model counterparts. The targeted moments are shown in blue and bold text. As shown in the table, the model matches key moments in the data successfully. Specifically, the model is capable of matching the mass of households with zero liquid wealth in the data, which determines the overall

[42]Including such households does not make significant differences in the target moments.

[43]In the model, firms' profits constitute both business owners' income (business income) and income from equity holding (asset income). Thus, there is no clear distinction between business income and asset income in the model.

[44]As in Kaplan et al. (2018), wealthy hand-to-mouth households are defined as households with zero liquid but a positive amount of illiquid assets.

Table 2: 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.,

responsiveness of consumption to changes in the interest rate. The overall amount of saving, the portfolio composition, wealth inequality, and indebtedness, as reflected in the capital to output ratio, aggregate liquid to illiquid asset ratio, the net worth Gini, and the share of households with debt, are also close between the model and the data. However, the model produces a larger mass of households with zero illiquid assets than in the data, because of the lack of housing in the model.

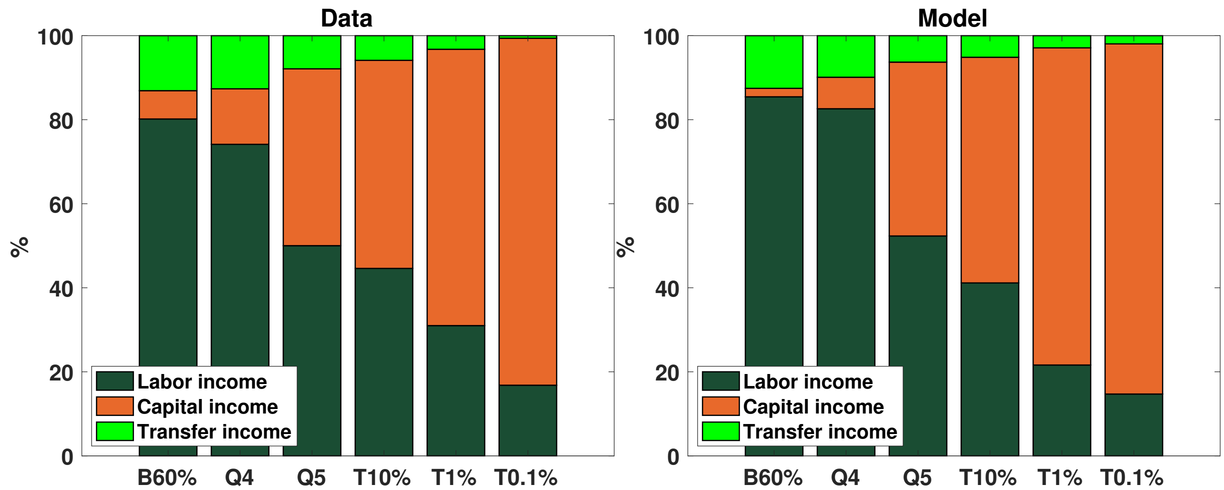
In Table 2, only the top 10% shares of liquid and illiquid wealth are targeted. Given that it is notoriously difficult to match the top end of the wealth distributions, the model does a reasonably good job generating an asset distribution close to the data.[45] As the table shows, more than 70% of each type of asset is held by the top 10% households.

Next, I discuss the income composition of households in different wealth groups. Inequality in households' wealth translates into heterogeneous household income composition in the model. As Figure 2 shows, the model closely matches households' income composition in the data. Both in the data and the model, the share of capital income, i.e., the sum of business and asset income, increases in households' wealth.[46] In contrast, the share of labor income, i.e., wages and salaries, decreases in households' wealth. For households in the bottom 60% of the wealth distribution, labor income accounts for about 80% of total income in the data and the model. For households with top 0.1% wealth, which I targeted, the labor income share is 16% both in the data and the model. Their capital income share is 81% in the model and 83% in the data. Though the model is less successful in matching the wealthiest households' relative asset holdings, it successfully matches their income composition. In the model, the top 10% wealthy households' labor and capital income shares are 43% and 49%, respectively. In the data, the corresponding

[45] These shares are defined, in the model and in the data, relative to households' total asset holding, not to aggregate asset holding, since the moments are computed solely from the SCF, which only contains data on households.

[46] For more detailed income composition in the data, see Figure A1 in the appendix.

Figure 2: 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.

values are 42% and 53%.

3.2 Calibration

Table 3 shows the calibrated parameters. 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.[47] 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% wealthy households hold 73% of total illiquid assets in the model, compared to 74% in the data.

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

Table 3: 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 4: 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 5: 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

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% wealthy households' share of liquid asset. The resulting value for $\tilde{P}_e$ is 20.6%. Tables 4 and 5 show the values for idiosyncratic productivity and the state transition matrix for workers and business owners.[48]

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.[49]

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

[48]Employment status transition probabilities, i.e., job-finding and separation rates, are discussed in the text.

[49]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 3 for Ξ^L corresponds to the value of the vacancy posting cost at the posterior mode.

the capital to output ratio of 3.03 in the data.[50][51] 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.[52] 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 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.

[50] 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.

[51] 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 3 is the level of the fixed cost that corresponds to the posterior mode of the vacancy posting cost.

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

3.3 Data for estimation

For quantitative evaluation of the effects of QE, I estimate the remaining model parameters with Bayesian methods, using the following set of ten observables.[53]

$$\left[\Delta \log Y_t, \Delta \log C_t, \Delta \log \tilde{I}_t, \log \pi_t, \log(1+i_t), \log u_t, \Delta \log w_t, \log T_t^g, \log \Pi_t, \log A_t^{\mathrm{CB}}\right] \quad (51)$$

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} are 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.

I measure output as real GDP and consumption as real personal consumption expenditure on non-durable goods and services. I define investment as the sum of private fixed investment on all types of fixed assets and personal consumption expenditure on durable goods. The inflation rate is defined as the quarterly percentage change of the GDP deflator. For the nominal interest rate, I use the effective Federal funds rate. The real wage in the model corresponds to the average hourly wage of production and non-supervisory employees in total private sector. The unemployment rate is the headline U-3 rate computed by the BLS. I measure lump-sum transfers as the sum of government's net current transfer payment and net capital transfer payment. For profits, I use after-tax corporate profits with inventory value adjustment and capital consumption adjustment. Lastly, I use all Federal Reserve bank assets to measure the central bank's asset in the model. The time period is from 1992 Q1 to 2018 Q4.[54]

I assume the following shock processes: 1) the MMMF's liquidity preference 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.[55]

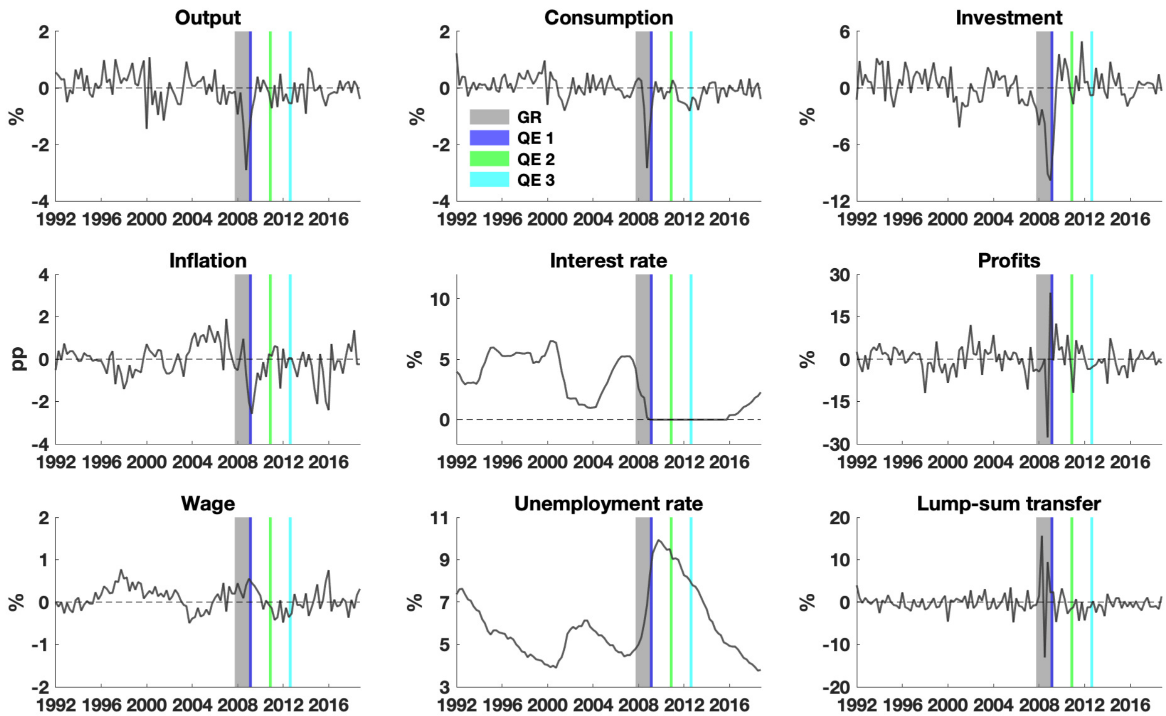
Figure 3 shows the dynamics of observables during the sample period. As shown in

[53] For a more detailed description of observables, see the appendix.

[54] I use relatively a short-sample period to avoid the periods with high interest rates, since households' optimal behavior is not consistent with the case in which the liquid asset return is higher than the return of illiquid assets. Thus, the dynamics of observables during the Great Recession and the ensuing ELB period are likely to affect the parameter values more significantly compared to when a longer sample is used. However, since the focus of this paper is on the dynamics of the economy during those periods, such an influence of the Great Recession is not a weakness in this paper.

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

Figure 3: 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.

the figure, output, consumption, and investment experienced the biggest drops in their growth rates during 2009. Following the implementation of QE 1, investment recovered, showing consecutive positive growth rates during the ELB episode. However, output and consumption still showed very low or negative growth rates even after the end of the Great Recession.

The inflation rate did not change much at the beginning of the crisis. However, it fell significantly during 2009, and then started to recover after QE 1 was implemented. Likewise, the unemployment rate soared to 10%, and then started to decrease gradually after 2009. Profits exhibit their most volatile dynamics during the Great Recession. The quarterly growth rate of profits fell to almost minus 30% in 2009, but recovered very quickly and remained mostly positive afterwards. In contrast, the real wage exhibits stable dynamics. During the Great Recession, real wages rose slightly, possibly because of a siginificant drop in the inflation rate.

As the contraction of the economy became severe and the inflation rate fell, conventional monetary policy reached its limit and the Fed embarked on so-called unconventional monetary policy, i.e., QE. As shown in the figure presented in the appendix, the central bank's balance sheet started to expand significantly starting in 2009. Compared to thieir pre-crisis level, the Federal Reserve's assets more than doubled during 2009 and continued to expand until the end of 2015, when the policy rate returned to a positive level for the first time since the Great Recession.

3.4 Estimation procedure

The biggest challenge associated with estimating a HANK model is to update the solution quickly for a new set of parameters. Even when using a perturbation method with a state-space reduction, it can take several minutes to find a new solution as the size of the equilibrium system is usually still large. Given that the estimation requires several hundreds of thousands of evaluations, estimation is infeasible with that amount of computation time. If one solves the model globally, then it takes immensely longer to solve a model, and thus calibration is the only viable option.

However, as Bayer et al. (2020) have shown, one can quickly update the solution if one restricts the set of parameters to be estimated to those that do not affect the steady state and uses auxiliary variables that summarize the effects of the household distribution over the idiosyncratic states on the aggregate variables.[56] The most time consuming part of the computation is the linearization of the model, which requires computing the Jacobian of the system. However, most of the equations are associated with the household's value function and the evolution of the distribution. Thus, if parameters do not directly affect the households' problem or the evolution of the distribution, these parts of the Jacobian do not need to be updated.[57] Thus, the number of elements in the Jacobian that need to be re-evaluated drastically decreases. Exploiting this idea, as proposed by Bayer and Luetticke (2020), I can update the solution very quickly in only several seconds, and thus the estimation is feasible.

When the model features an occasionally binding constraint, an evaluation of the likelihood requires additional steps as the solution depends on the expected ELB duration while the model is at the ELB. In this paper, I adopt the approach of Kulish et al. (2014), Jones (2017) and Jones et al. (2018) and assume a sequence of expected ELB durations

[56]To be precise, if the households' optimal policies at the steady state do not change, the solution can be obtained quickly for a given set of parameters.

[57]For instance, the discount factor or the household's relative risk aversion affect the way that households respond to a given set of price variables, i.e., wages and asset returns. Thus, if one wants to estimate these parameters, the Jacobian related to the household's problem also needs to be updated.

during the estimation.[58] I estimate these durations along with other structural parameters of the model. Specifically, I apply the randomized blocking scheme developed by Chib and Ramamurthy (2010) and used in Kulish et al. (2014).[59] For the likelihood evaluation, I use the inversion filter instead of the Kalman filter to speed up the estimation process.[60] For the estimation, 1,000,000 draws were evaluated. The first 200,000 draws were discarded as burn-in, and the remaining 800,000 draws were used to construct the posterior distributions of the structural parameters and the expected ELB durations.

[58]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.

[59]During the estimation, I make a 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 the previously accepted values and vice versa. For the expected ELB duration draws, I first randomly sample the number of quarters to update from the 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 multinominal 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 bu up to four quarters. Based on the ratio of the likelihoods, the acceptance is determined. For the other block with structural parameters, a standard Metropolis-Hastings algorithm is used.

[60]When using the Kalman filter, I need to keep updating the state transition matrix, which takes a considerable amount of time given a large size of the equilibrium system.

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.

3.5 Prior and posterior distributions

For structural parameters, I mostly follow the literature and use standard priors. For the slope of the Phillips curve, I assume a gamma prior distribution with mean 0.1 and standard deviation 0.02, which is equivalent to a Calvo price contract with an average price duration of about one year. For the degree of indexation to previous inflation in price and wage setting, I use gamma priors with means and standard deviations of 0.50 and 0.15, respectively, following Smets and Wouters (2007) and Justiniano et al. (2011). Since the existing literature mostly uses investment adjustment costs, the prior distribution for capital adjustment costs is chosen in a heuristic way. Specifically, I assume a normal distribution with mean 30 and standard deviation 5 for the degree of capital adjustment frictions in the model. The prior for the autocorrelation of the real wage is set to a beta distribution with mean 0.5 and standard deviation 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 structural parameters imply a high degree of wage and price rigidity, a relatively low vacancy posting cost, and significant capital adjustment frictions.[61] The posterior distribution for the slope of the Phillips curve is centered around a relatively low value, implying a high degree of price rigidity in the model. The value of this parameter at the mode corresponds to an average price duration of 6 quarters in an equivalent Calvo price setting. Similarly, wage rigidity is also estimated to be very high, implying that only a fifth of the real wage adjusts in proportion to changes in labor productivity, as proxied by the labor rental rate. In contrast, the estimated vacancy posting cost is low.

[61] This estimation result is due to the relative dynamics of profits, wages, and unemployment rates in the data. As I showed, the real wage is very stable in the data. Moreover, it does not strongly comove with output. In contrast, profits are volatile and positively comove with output. The correlation between the growth rates of profits and output is about 0.3, while the correlation of output with the real wage is -0.1. Thus, in fitting the data, the model favors a high degree of wage rigidity. Also, the model requires a low vacancy posting cost to generate strongly procyclical profits in response to demand shocks.

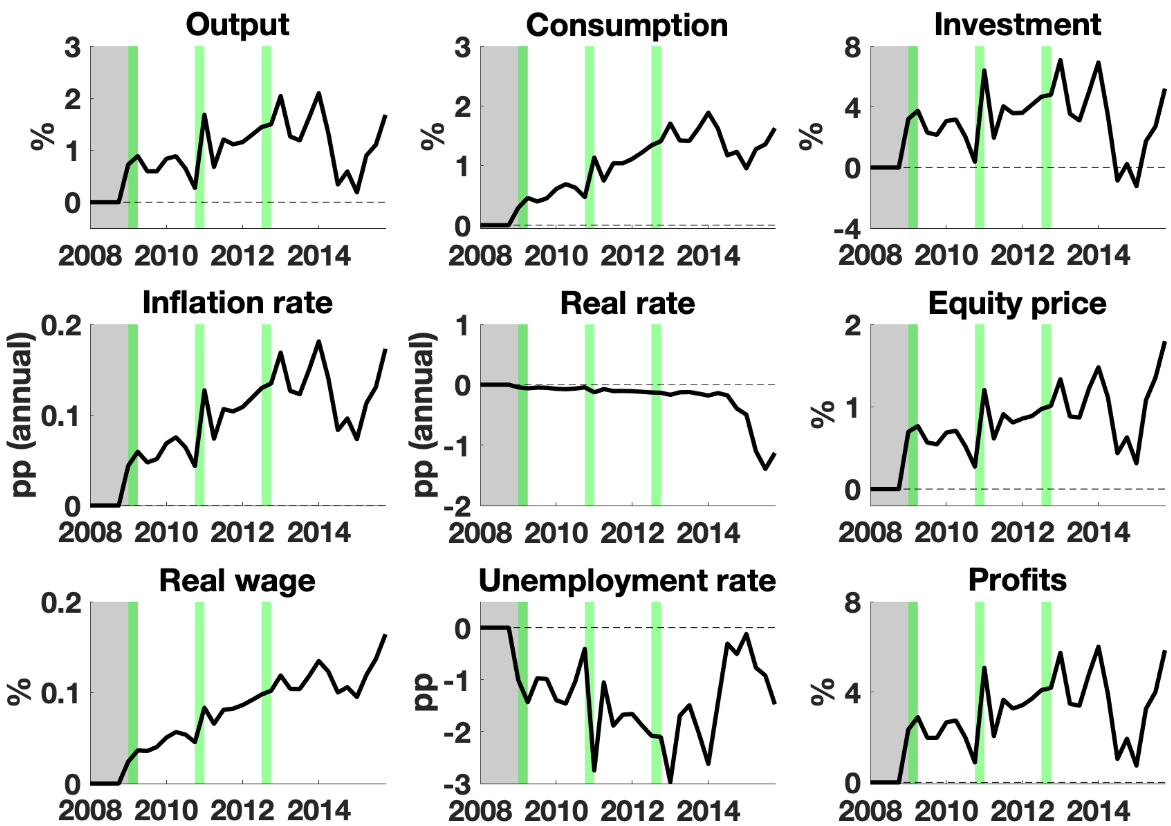
The estimated parameter value for the capital adjustment cost is particularly high. This is because of the presence of banks in the model.[62] As I show in the appendix, these parameters imply that the model generate a strong and procyclical response of profits, equity prices, unemployment rates, and an almost acyclical real wage response to monetary policy shocks, consistent with the empirical evidence.[63]

The estimated policy parameters are fairly standard except for the autocorrelation of the government bond issuance, which is low, and close to its prior mean, which is relatively low. However, such a low value of this parameter does not imply strong fiscal responses to exogenous shocks because of the assumption that the fiscal authority adjusts contributions to the MMMF. Since the MMMF smooths its transfer flows to households by assumption, its liquid saving fluctuates instead of government purchases or the government lump-sum transfer.

4 Quantitative Easing during the ELB episode

Now, I answer the central question of this paper: did QE raise inequality in the U.S. during the ELB episode? To this end, I conduct a counterfactual analysis that compares the economy's actual outcomes (as the baseline) to an alternative without QE. Since I use the inversion filter to extract structural shocks, 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. However, the central bank does not conduct unconventional monetary policies 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. Moreover, 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 unconventional monetary policy relative to the scenario of a passive central bank.[64]

Figure 4: Aggregate effects of QE



Notes: Except for inflation, the real rate, and the unemployment rate, variables are shown as percentage differences from the corresponding values in the alternative case with no policy interventions. The inflation rate, the real rate, and the unemployment rate are shown as the percentage point differences from their corresponding value in the alternative case.

4.1 Aggregate effects of QE

Figure 4 shows that, in the baseline case with QE output and consumption are about 1% higher on average than in the counterfactual case with no policy interventions. Likewise, investment is about 3% higher on average in the baseline case. The effects of QE on profits and the unemployment rate are particularly strong. On average, profits are about 3.3% higher during the ELB episode than in the case without QE. Similarly, on average,

[62] As in Gertler and Karadi (2011), a financial accelerator channel applies to banks in the model. Thus, bank assets respond strongly to changes in the equity price and the expected return, leading to high volatility of investment. Thus, a high degree of capital adjustment frictions is required to generate volatility of investment that is consistent with the data.

[63] Impulse response functions are reported in the appendix.

[64] For this section, I use the term 'QE' as a shorthand to refer to both asset purchases and forward guidance.

the unemployment rate is about 1.4% lower relative to what would have happened in the counterfactual case without QE. The effect of QE on equity prices is also quite significant. On average, the equity price is about 1% higher because of QE relative to the level when there were no unconventional policy interventions.[65] In contrast, the real wage differences are relatively small, with an average difference of 0.1% between the two cases. Importantly, these relative magnitudes of aggregate effects are similar to those of the effects of conventional monetary policy shocks, which implies that QE has the aggregate effects similar to those of expansionary interest rate shocks, and thus undoes the adverse effects of the binding ELB constraint.[66]

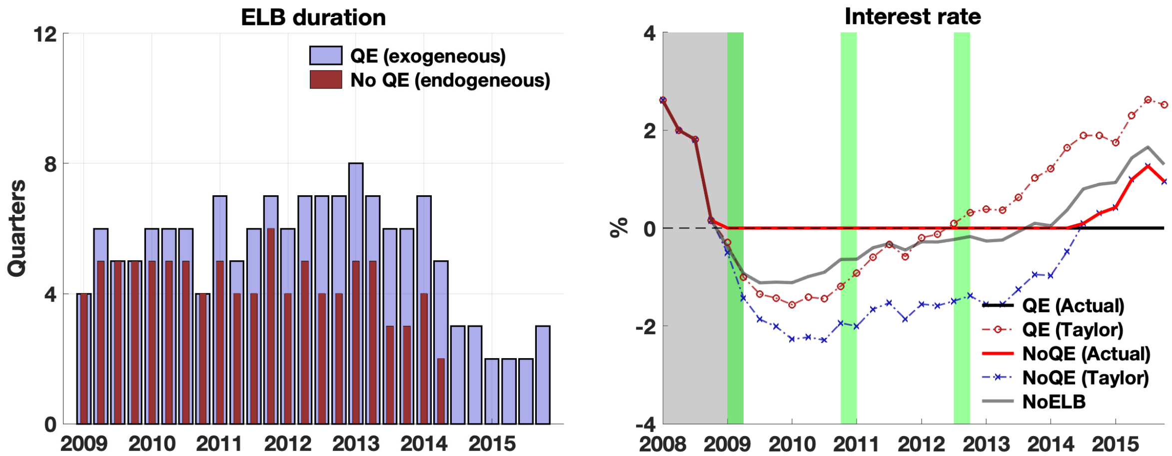
Notably, the aggregate effects of QE increase over time. This is because the effects of the central bank asset purchases accumulate and interact with the effects of forward guidance, i.e., longer exogenous expected ELB durations and policy rates staying at zero in later periods of the ELB episode. In Figure 5, the gap between the actual rate (solid red and black) and the policy rate prescribed by Taylor rule (red circle and blue cross) shows how severely the economy is constrained by the ELB if the actual rate is higher than the rate prescribed by the rule.[67] Conversely, if the actual rate is lower than the prescribed rate, the gap reflects the degree of expansionary monetary policy. As shown in Figure 5, the estimated monetary policy rule prescribes higher policy rates in the baseline case than in the counterfactual case at the beginning of the ELB episode, which implies that the economy is less severely affected by the binding ELB constraint because of the central bank's asset purchases. Besides, the central bank maintains the policy rate at zero until the end of 2015, even though the Taylor rule implies positive rates starting in 2012. The central bank also maintains an expected ELB duration of two to three quarters until the end of the ELB episode in the baseline case, while it sets positive rates from 2014 Q4 in the counterfactual case. Because of this combination of unconventional monetary policies, the economy experiences further stimulus effects, especially during the later periods of the ELB episode. In a later section of this paper, I isolate the effects of the central bank's asset purchases from the effects of other types of unconventional policies. But, in this

[65] The magnitudes of the effects of unconventional monetary policies in the model fall into the ballpark of the existing estimates found in the literature. For instance, using a set of existing empirical and DSGE models, Chung et al. (2012) finds that the unemployment rate would be lower by 1.5% points compared to what would have happened absent policy interventions. Similarly, using the FRB/US model, Engen et al. (2015) report estimated effects of QE on unemployment rates ranging from a 0.8% point to 1.5% point decrease. Regarding the effects of QE on asset prices, Kiley (2014) finds that a policy-induced 100 bp decline in 10 year Treasury yields is associated with 1.5-3% increase in the equity price. Also, Rosa (2012) finds that an unanticipated expansionary QE announcement is associated with a stock price increase of 0.9%.

[66] The model's impulse responses to an expansionary interest rate shock are shown in Figure A6 in the appendix.

[67] The Taylor rule prescribed policy rates reflect the aggregate economic state, which is affected by unconventional monetary policies.

Figure 5: ELB durations and interest rates



Notes: The left panel shows the expected ELB durations in the baseline case with QE and the counterfactual case with no policy interventions. In the counterfactual case, the endogenous ELB durations are obtained by applying the OccBin method of Guerrieri and Iacoviello (2015) for given shocks. The dark red bars show the endogenous ELB durations while the light purple bars show the exogenous durations. The right panel shows the dynamics of the actual policy rate and the policy rate prescribed by the estimated policy rule (Taylor rule) in the baseline and the counterfactual case. The thick black line and the red dotted line with circles show the actual policy rate and the Taylor rule prescription in the baseline case, respectively. The thick red line and the blue dotted line with crosses show the actual and the Taylor rule prescription in the counterfactual case. The light gray line shows the policy rate in the case in which the ELB does not bind.

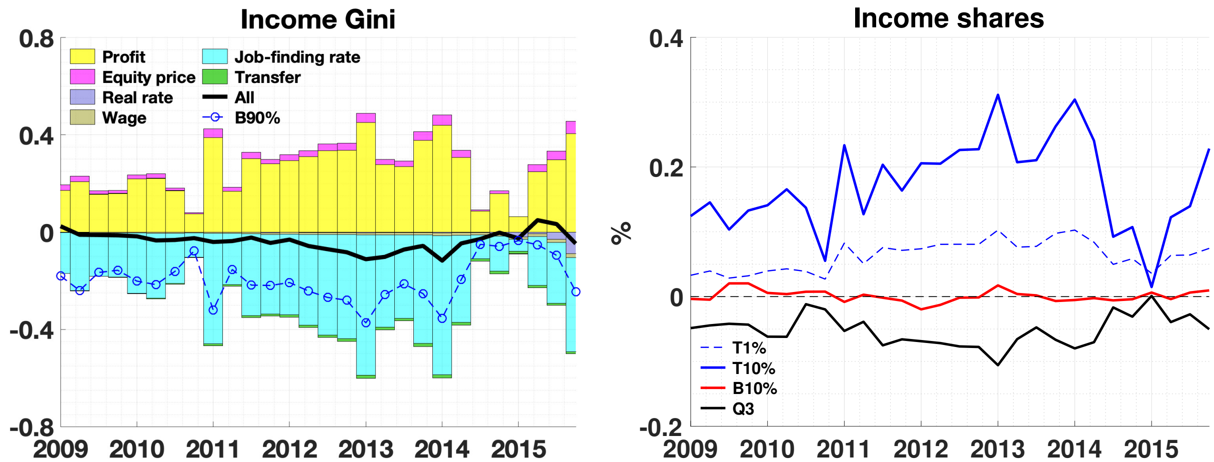
section, I do not distinguish them and consider these various types of unconventional policies as a whole.

4.2 Distributional effects of QE

In this section, I evaluate the effects of QE on inequality in detail, using the Gini index, the top 10% share, and welfare gains, as measured by consumption equivalents. Iin addition to examining the overall effects, I compute the contribution of each variable that affects households' wealth and income, including the job finding rate to understand the underlying mechanisms.[68]

[68]The decomposition method is similar to the microsimulation used in Casiraghi et al. (2018) and Lenza and Slacalek (2018). Specifically, I compute the evolution of inequality by feeding the expected paths of profits, real interest rates, wages, job-finding rates, lump-sum transfers, and equity prices in isolation. The household's optimal responses are computed based on each expected path. For more detail, see the appendix.

Figure 6: Distributional effects of QE: Income inequality



Notes: The left panel shows the differences in the model-implied income Gini index (0 to 100) between the baseline and the counterfactual case. The thick black line shows the overall effect of QE, while each bar shows the contribution of each variable to the overall effect. The blue dotted line with circles shows the Gini index computed from households in the bottom 90% of the wealth distribution. The right panel shows the income share of the top 1% (dashed blue), 10% (blue), the middle quintile (black), and the bottom 10% (red) households, as the percentage point difference compared to the corresponding levels in the counterfactual case. Income is the sum of labor, capital, and transfer income. Capital gain is not included.

Given the aggregate effects of QE examined in the previous section, increases in profits and equity prices due to QE are much higher than those of the real wage, which pushes towards higher inequality. In contrast, the lower unemployment rate is likely to reduce inequality as it benefits households at the bottom of the wealth distribution.[69] Given that changes in the real interest rate are not so large because of the binding ELB constraint, redistribution through real interest rate is unlikely to be large. Thus, the relative magnitudes of the first two channels are likely to determine the net effects of QE on inequality during the ELB episode.

Overall, QE had non-linear distributional effects during the ELB episode. QE benefited the top 10% of the wealth distribution substantially, by boosting profits and equity

[69]Since unemployed households make only a small amount of labor income, they mostly belong at the bottom of the wealth distribution. In the model, at the beginning of 2009 Q1, the share of unemployed households at the bottom 10% of the wealth distribution is 8.75%, while the share is only 6.54% in the middle quintile. Accordingly, even though the job-finding rate increases uniformly across the household distribution, a larger number of households become employed at the bottom of the wealth distribution than in other groups. Thus, an increase in average income is larger than in other groups. Moreover, poor households are mostly hand-to-mouth. An increase in income leads to a larger increase in consumption and welfare than other groups.

prices. At the same time, it also benefited the bottom 10% of the wealth distribution significantly by reducing unemployment rates. In contrast, the gains for the middle class were relatively small because of the small changes in the real wage. Because of this non-linear effect, QE can be seen as either increasing or decreasing inequality, depending on the measure of inequality used.

4.2.1 Income inequality

Figure 6 shows the evolution of the income Gini index and of income shares during the ELB episode. As shown in the left panel, QE modestly reduced the income Gini, relative to the counterfactual in which the Fed became passive once nominal rates reached the ELB.[70] Lower unemployment rates reduced the income Gini by up to 0.6 percentage points, consistent with policymakers' arguments that emphasized the positive impact of QE on the labor market.[71] Higher profits and equity prices offset about 80% of this effect, while transfers and wage growth had negligible effects.[72]

An interesting result is that the decrease in income inequality is larger when the income Gini index is computed only using households in the bottom 90% of the wealth distribution. The dynamics of the Gini index among the bottom 90% households closely follow the effects of QE on the job-finding rate. This is because the bottom 90% households have similar income composition and mostly rely on labor income.

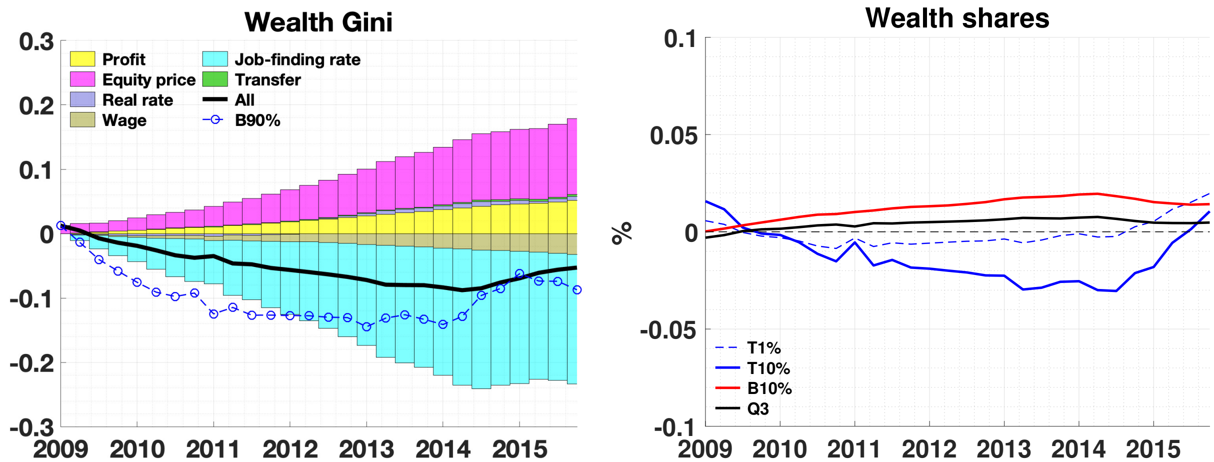
The right panel of Figure 6 shows that QE widened the income gap between the top 10% and the rest of households during the ELB episode. As shown in the figure, QE increased the top 1% and 10% income shares in aggregate income. Changes in the bottom 10% income share fluctuate around zero, while the income share of the middle quintile fell. This result underscores the failure of the income Gini to capture non-linear distri-

[70] During the ELB episode, the income Gini was higher compared to its level at the beginning of the Great Recession, as shown in Figure A14 in the appendix. QE accounts for only -2.5% of total changes in the income Gini index's, relative to the level at the beginning of the Great Recession, during the ELB episode. A small net effect is due to the result that the effects of higher profits and equity prices mostly offset the effects of lower unemployment rates.

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

[72] Note that capital gains are not included in the definition of income in the model because it is hard to keep track of the purchasing price of illiquid assets for each household. Though capital gain is not included, higher equity prices contribute to an increase in income for two reasons. First, when households sell their equity, they sell a lesser amount when the price is higher. Thus, after selling, households hold a larger amount of equity compared to when equity prices were lower. Besides, since households receive deceased households' equity holdings as a part of annuity arrangement, higher equity price increases equity holders' income from the annuity arrangement. If the capital gain is defined as the value of equity sold minus the steady state equity price, the income Gini is higher in 2015 in the baseline case than in the counterfactual case. However, for other years, income Gini is still lower in the baseline case.

Figure 7: Distributional effects of QE: Wealth inequality



Notes: The left panel shows the differences in the model-implied wealth Gini index (0 to 100) between the baseline and the counterfactual case. The thick black line shows the overall effect of QE, while each bar shows the contribution of each variable to the overall effect. The blue dotted line with circles shows the Gini index computed from households in the bottom 90% of the wealth distribution. The right panel shows the model-implied income shares of the top 1% (dashed blue), 10% (blue), the middle quintile (black), and the bottom 10% (red) households, as the percentage point difference compared to the corresponding levels in the counterfactual case. Income is the sum of labor, capital, and transfer income. Capital gain is not included.

butional effects of QE.[73] The income Gini falls mainly due to the reduced income gap between the bottom 10% and the middle class, which is large enough to offset the higher income gains for the top 10% in the calculation of the index.

4.2.2 Wealth inequality

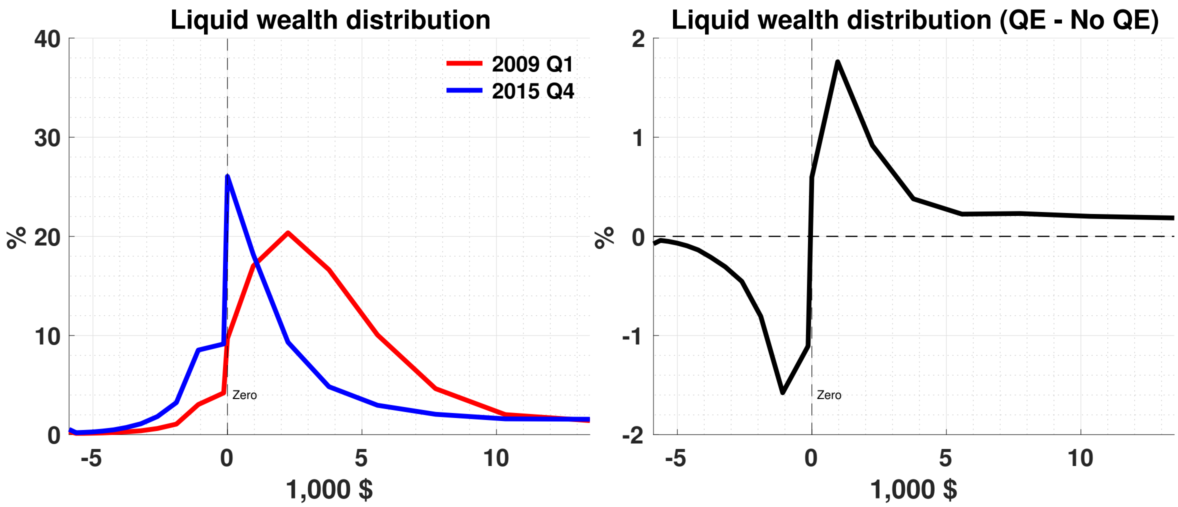
Despite its positive impact on equity prices, QE reduces the wealth Gini index slightly, and the main driver is also lower unemployment rates.[74] The literature has found similar results, but the existing work emphasizes competing effects of equity prices versus the savings redistribution or the role of house prices.[75] In the model, a redistribution

[73]It is well known that many existing inequality measures, including Gini indices, do not guarantee subgroup consistency. Such a problem is particularly pronounced in the case of the consumption Gini in the model, which is presnted in the appendix. For a detailed discusion on the properties of the Gini index, see, for instance, Jurkatis and Strehl (2014).

[74]As shown in Figure A14 in the appendix, the wealth Gini index increased by about one percentage point during the ELB episode, compared to its level at the beginning of the Great Recession. Changes in the wealth Gini index induced by QE accounts for about -5% of the total change during the ELB episode.

[75]Casiraghi et al. (2018) and Inui et al. (2017) both find that QE has only negligible effects on wealth inequality and argue that the effects of *portfolio composition channel* and *savings redistribution channel* cancel

Figure 8: Effects of QE on the liquid wealth distribution



Notes: The left panel shows model-implied households' liquid wealth distribution at the beginning and the end of the ELB episode. The red line shows the distribution in 2009 Q1. The blue line shows the distribution in 2015 Q4 in the baseline case. The black line in the right panel shows the differences in the liquid wealth distribution in 2015 Q4 between the baseline and the counterfactual case. A positve value implies that there is a larger mass in the baseline case than in the counterfactual case. Units are converted to dollar values, assuming that the steady state output is equal to real GDP per capita in 2009. Negative liquid wealth implies debt.

from real interest rate changes is weak due to limited changes in the real rate, and there is no housing. Instead, households dynamically adjust their liquid savings to smooth consumption, and such behavior leads to a fall in wealth inequality. As shown in the left panel of Figure 8, during the ELB episode, a substantial share of households exhaust their liquid savings, and some of them become indebted mostly because of the higher unemployment rate.[76] By reducing the unemployment rate, QE helped households maintain their liquid wealth or prevented them from accumulating debt. The right panel of Figure 8 shows that QE increases the mass of households with a positive amount of liquid wealth and reduces the mass of indebted households at the end of the ELB episode. This result shows the importance of taking into account the dynamic responses of households' balance sheets, including debt, in evaluating changes in household wealth inequality.[77]

out each other. Lenza and Slacalek (2018) also find a similar result, but they emphasize the role of housing prices in offsetting the adverse effects of equity price on inequality.

[76] Figure A16 shows that, during the ELB episode, households in the first quintile of the wealth distribution experienced dramatic declines in wealth during the ELB episode, and they are the most vulnerable households to unemployment risks.

[77] Domanski et al. (2016) highlights the effects of QE on equity prices and argues that QE increases wealth inequality in European countries. However, Domanski et al. (2016) assume that household portfolios is fixed during the simulation.

The results for wealth shares are consistent with the wealth Gini dynamics, unlike in the case of income inequality. That is, the top 1 and 10% wealth shares fall during the ELB episode in the baseline case, relative to their levels in the counterfactual case. At first, this result seems at odds with the effects of QE on income inequality. However, further analysis reveals that higher income of the top 10% induced by QE translates into higher inequality in illiquid asset holding, while benefits of the lower unemployment rate translate into lower inequality in liquid asset holding. As shown in the appendix, the latter effect mainly determines the dynamics of the overall wealth inequality. The top 1% and the top 10% equity holding share increases due to higher equity prices during the ELB episode, but these effects are swamped by reduced inequality in liquid asset holdings.[78]

4.2.3 Welfare effects

In this section, I examine who gains most in terms of welfare from QE during the ELB episode. To this end, I compute the consumption equivalents across different wealth groups. Specifically, I define wealth groups based on the distribution of wealth in 2009 Q1 and I keep track of these groups during the ELB episode, computing their consumption equivalents.[79] By comparing these groups' consumption in different cases, I compute different groups' consumption equivalents, defined as the fraction of lifetime consumption in the counterfactual case agents would be willing to give up to benefit from QE.

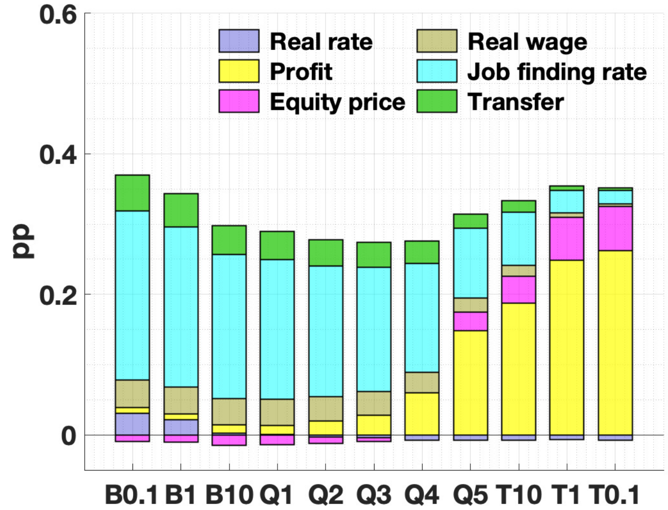
Figure 9 shows that QE benefits wealthy and poor households the most while leaving the smallest welfare gains for the middle quintile.[80] The average welfare gain from QE is equivalent to 0.27 percent of lifetime consumption. However, households at both ends of the wealth distribution enjoy higher than average welfare gains from QE while the middle class enjoys the least benefits: the consumption equivalent for the bottom and the top 1% households is about 0.06 percentage points higher than that of the middle 60%. The differences in welfare gains are due to different shares of the unemployed in each group and the groups' income and wealth composition. In 2009 Q1, the aggregate unem-

[78] Though the overall wealth inequality has fallen, a rise in equity holding inequality, together with higher profits, leads to higher income inequality in later periods of the ELB episode.

[79] Households' wealth distribution in 2009 Q1 is determined in 2008 Q4, and thus, is not affected by QE. 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.

[80] Across the working status, business owners enjoy the highest welfare benefit equivalent to 0.82% of lifetime consumption, followed by the unemployed with 0.35% of lifetime consumption. The welfare gain for the employed is equivalent to 0.27% of lifetime consumption.

Figure 9: Welfare effects of QE: Consumption Equivalents



Notes: The figure shows the welfare gains from QE in terms of consumption equivalents. Consumption equivalents are computed for differet groups of household wealth distribution as of the beginning of the ELB episode, assuming that there are no shocks after the end of the sample period. Bars in the positive region indicate welfare gains, while bars in the negative region reflect welfare losses. The sum of the height of the bars in the positive and negative regions show the net welfare gains. Boxes with different colors show the contribution of each variable. Units are percetage points. B0.1 (T0.1), B1 (T1), and B10 (T10) refer to the bottom (top) 0.1%, 1%, and 10% of the wealth distribution, respectively. Q1 to Q5 refer to the first to the fifth quintile.

ployment rate is 8.27%, but the share of unemployed households at the bottom 10% of the wealth distribution is 8.75%. In contrast, the share of unemployed households in the middle quintile is 6.54%. As a higher share of households is unemployed at the bottom of the wealth distribution, welfare gains from the higher job-finding rate are larger for them. In contrast, as profits and equity account for a significant proportion of the top 10% households' income and wealth, higher profits and equity prices lead to higher than average welfare gains for those households.

A noteworthy result 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 the bottom and the top 10% are similar, and the welfare gain is largest for the bottom 0.1%. This result is due to the expected effects of tapering in periods beyond the sample. During the ELB episode, wealthier households enjoy higher consumption gains that mirror higher income gains. However, as the economy enters into the tapering phase, households expect lower equity prices and profits.[81] Lower profits reflect the adverse effects of tapering on banks' net worth, but are not accompanied by equivalently higher

[81] The evolution of key variables during the tapering phase is shown in Figure A18 in the appendix.

unemployment rates. Moreover, as tapering generates downward pressures on the inflation rate, real wages are expected to be higher in the future. Accordingly, welfare gaps between the top 10% and the bottom 90% are smaller than the gaps in the income response during the ELB episode.

To recapitulate, I find that the wealth and income Gini indices are slightly lower during the ELB episode, mainly because QE's positive effects on employment are strong enough to offset its positive effects on profits. However, the Gini index fails to capture the strong income gains for the top 10% households whose income share rises. In terms of welfare gains, all households benefitted from QE, but both ends of the wealth distribution enjoyed higher gains relative to the middle. Overall, the welfare gaps across households are small relative to income gaps, because of the transient effects of QE on future profits and equity prices. I conclude that concerns about QE widening inequality are not supported by the experience of the Great Recession.

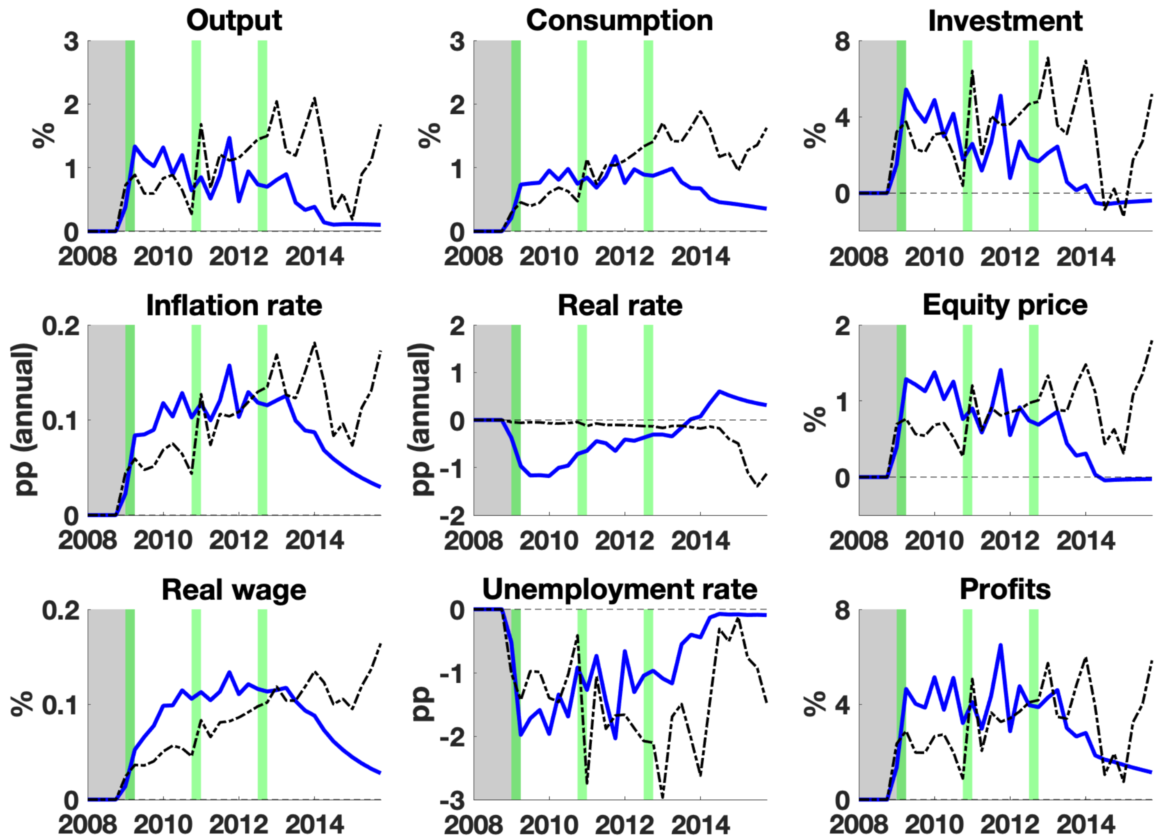
5 QE and Conventional Monetary Policy

The persistent decline in the natural interest rate in recent decades has spurred concerns about increasing incidence of ELB episodes going forward. As a result, the literature has started to discuss increasing the inflation target and thus the steady-state nominal policy rate, thereby securing more room for the operation of conventional monetary policy (CMP).[82] In this section, I compare QE and conventional monetary policy in terms of both aggregate and distributional consequences, to provide a reference for the benefit of avoiding the binding ELB constraint. Specifically, I ask what would have happened 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, ignoring the ELB constraint, but does not conduct any QE.[83] Specifically, the policy rate follows the gray line in Figure 5.

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

[83]In the simulation, the nominal policy rate goes below zero. However, I do not interpret the results presented in this section as the effects of negative interest rates because saving in assets whose nominal rate is negative can be irrational in practice. Instead, I interpret the results as the effects of CMP when the nominal policy rate and the inflation rate are higher by the same amount. In this case, real interest rates are the same as in the baseline case, but the central bank has more room for lowering the nominal policy rate.

Figure 10: Aggregate effects: QE vs CMP



Notes: The solid blue and dashed blak lines show the effects of QE and CMP. Except for the inflation rate, the real interest rate, and the unemployment rate, variables are shown as percentage differences from the corresponding values in the alternative case with no policy interventions. The inflation rate, the real interest rate, and the unemployment rate are shown as the percentage point differences from their corresponding values in the alternative case.

5.1 Aggregate effects

Figure 10 shows that CMP initially has stronger stimulus effects than QE, but in later periods, the effects are smaller than those of QE. As shown in Figure 5, the central bank in the QE case imposes longer expected ELB durations and maintains the policy rate at zero even after the Taylor rule prescribes positive rates. In the case of CMP, these effects are absent as the economy does not stay at the ELB. Thus, CMP (as prescribed by the Taylor rule) would have had weaker stimulus effects than QE in the later periods of the ELB episode.

What is somewhat surprising is the stronger initial effects of CMP than those of QE, especially given the large amount of central bank asset purchases at the beginning of the ELB episode. This result is not due to CMP having particularly strong stimulus effects.[84] Rather, it is due to the weak initial stimulus effects of QE in the model.[85] First, QE directly affects only small fraction of households that hold equity. As these equity holders are mostly rich and have below-average marginal propensities to consume, the direct stimulus effects of QE on household consumption is relatively weak compared to those of CMP.[86] Moreover, QE *crowds out* private investment, especially of banks. An increase in equity prices boosts banks' net worth but decreases the expected gross rate of return on equity, i.e., banks' profitability, which discourages banks' investment. In contrast, CMP does not directly increase banks' net worth, but in the short run it increases the profitability of banks by lowering their financing costs. Thus, CMP *crowds in* banks' investment.[87][88]

Overall, CMP has smaller average stimulus effects than QE because of the larger stimulus effects of QE in later periods. On average, the effects of CMP on equity prices and unemployment rate are about 20% and 30% smaller than those of QE.[89] In contrast, the effects of CMP on profits are only about 3% lower than those of QE. This is because CMP strongly positively affects banks' net worth initially, and such effects propagate through the financial accelerator channel embedded in the model. Finally, the average effect on the real wage is greater in the case of CMP because of the larger initial impact of CMP combined with wage rigidity.[90]

[84]As shown in Figure A6, interest rate shocks have a modest amount of stimulus effect, in line with the findings of the literature.

[85]Note that the central bank's asset purchases in 2009 Q1 are equivalent to about 6.5% of steady state output in the model. Thus, if all of the asset purchases are translated into stimulus on output without any offsetting effects, output should have increased by as much as 6.5%. However, in the model, the initial impact on output is less than 1% of the steady state output, which implies that a substantial proportion of QE's stimulus effects are offset by general equilibrium responses.

[86]Adjustment frictions also contribute to small direct consumption responses out of QE as they cause only a subset of households to adjust their equity holdings.

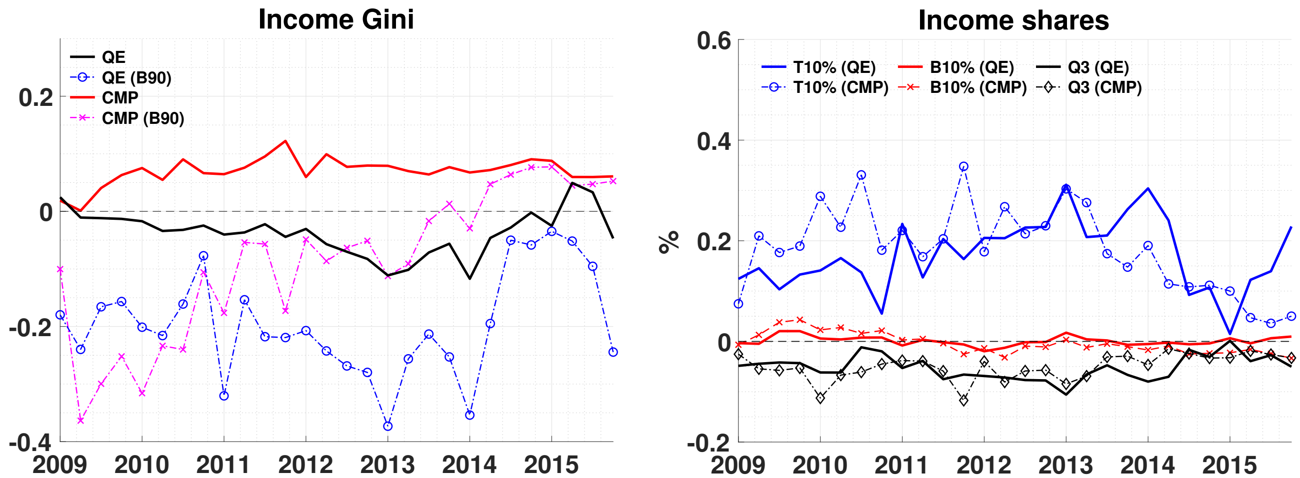
[87]The general equilibrium effects of QE and CMP on equity prices and equity premia are similar. However, higher equity prices and lower equity premia are the consequences of banks' expansion in the case of CMP. In contrast, in the case of QE, a large part of higher equity prices and lower equity premia is due to the central bank's action.

[88]Figure A19 in the appendix shows that CMP has much stronger positive impacts than QE on banks' net worth and investment.

[89]For instance, QE reduces the unemployment rate by 1.4%, while CMP reduces the unemployment rate only by 1% on average during the ELB episode.

[90]Because of high wage rigidity, the increase in the real wage does not quickly dissipate. As the real wage is maintained at a relatively high level, the unemployment rate returns to ints counterfactual level more quickly under CMP than QE.

Figure 11: Distributional effects of QE and CMP: Gini index and income shares



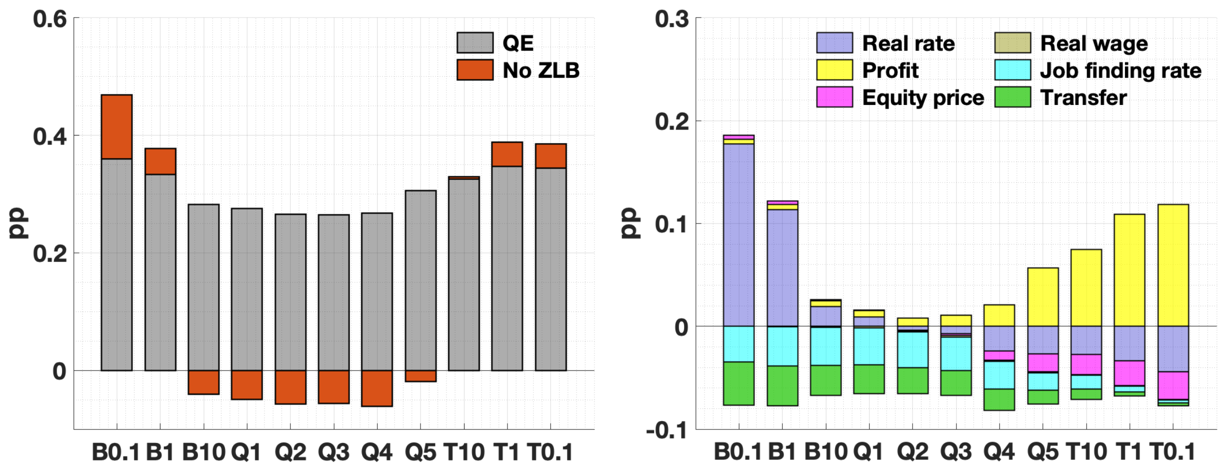
Notes: The left panel shows the differences in income Gini indices between the case of QE or CMP and the counterfactual case of no policies. The black and red solid lines shows income Gini in the case of QE and CMP, respectively. The pink line with crosses and blue line with circles show the Gini indices among the bottom 90% households. The right panel shows the impact of policy on the income share of the top 10%, the bottom 10% and the middle quintile in the case of QE and CMP.

5.2 Distributional effects

Figure 11 compares the distributional effects of QE and CMP in terms of the income Gini index and the income shares across different wealth groups. Note that, for the first two years of the ELB episode, the unemployment rate is lower, and profits and equity prices are higher for the reasons discussed in the previous section. The initial increase in the real wage is also larger under CMP, though still relatively modest. The associated distributional consequences show more contrast between the top 10% and the bottom 90% under CMP than under QE. The top 10% and the bottom 10% are both higher from 2009 to 2011 in the case of CMP than in the case of QE. Accordingly, the overall income Gini is higher under CMP compared to either the counterfactual case witn no policy interventions or QE, but the Gini index among the bottom 90% is lower in the case of CMP. That is, the income distribution becomes more skewed in the case of CMP: while the top 10% moves farther away from other households, the distribution among the rest of households becomes more compressed. Overall, I conclude that CMP has more disequalizing effects than QE: CMP increases the income Gini index, whie QE decreases the income Gini index compared to the counterfactual case with no policy iterventions.[91]

[91] In the case of wealth inequality, both CMP and QE reduces the wealth Gini index. However, the magnitudes of the decrease is smaller under CMP than under QE.

Figure 12: Welfare gain comparison: QE vs CMP



Notes: The figure compares the welfare gains from QE and conventional monetary policy in terms of the consumption equivalent. The gray bars show welfare gains from QE across households. The dark organge bars show additional gains or losses from conventional monetary policy compared to welfare gains from QE. The right panel shows the decomposition of the additional gains or losses from CMP, relative to gains from QE.

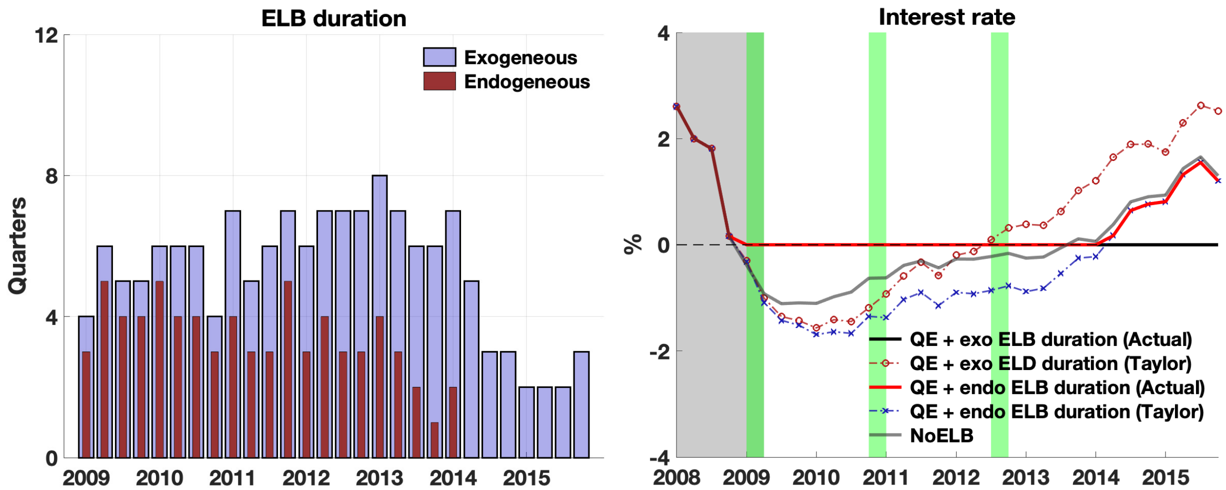
In terms of welfare gains, the non-linear effects of CMP are also stronger than those of QE. Since the average stimulus effects of CMP are smaller, most households experience smaller welfare gains under CMP than under QE. Only the top 1% and the bottom 1% of households enjoy higher gains from CMP than QE, as shown in Figure 12. The magnitudes of welfare losses from CMP relative to QE are largest for the middle 60% and smallest for the fifth quintile among five quintiles, which confirms CMP's more adverse effects on inequality than QE.

Decomposing the effects shows that for poor households at the bottom of the wealth distribution, a lower debt burden provides almost all of the additional gains under CMP, which offsets the losses from (relatively) higher unemployment rates and lower transfers.[92] In contrast, for wealthy households, the response of profits provides all of the additional benefits from CMP.[93] And larger effects on profits, relative to those on other variables, is partly due to the benefits of lower real rates for banks in the case of CMP. That is, there is an additional redistribution towards the financial sector in the case of CMP, which cancels out part of the welfare losses from lower rates for wealthy households.

[92] Since CMP lowers the return on liquid assets and tax revenues are lower in the case of the CMP, lump-sum transfers from the MMMF is smaller.

[93] The average magnitudes of the effects are similar in both cases. However, in the case of the CMP, the initial effects are larger. Since households discount the value of future consumption, larger initial effects on profits lead to welfare gains for wealthy households.

Figure 13: ELB durations and interest rates



Notes: The left panel shows the expected ELB durations in the baseline case with QE and exogenous ELB durations and in the alternative case with QE and endogenous ELB durations. The dark red bars shows the endogenous ELB durations while the light purple bars show the exogenous ELB durations. The right panel shows the dynamics of the actual policy rate and the policy rate prescribed by Taylor rule in the baseline and the counter-factual case. The thick black line and the red dotted line with circles show the actual policy rate and the Taylor rule prescription in the baseline case. The thick red line and the blue dotted line with crosses show the actual and the Taylor rule prescription in the alternative case.

6 The effects of forward guidance

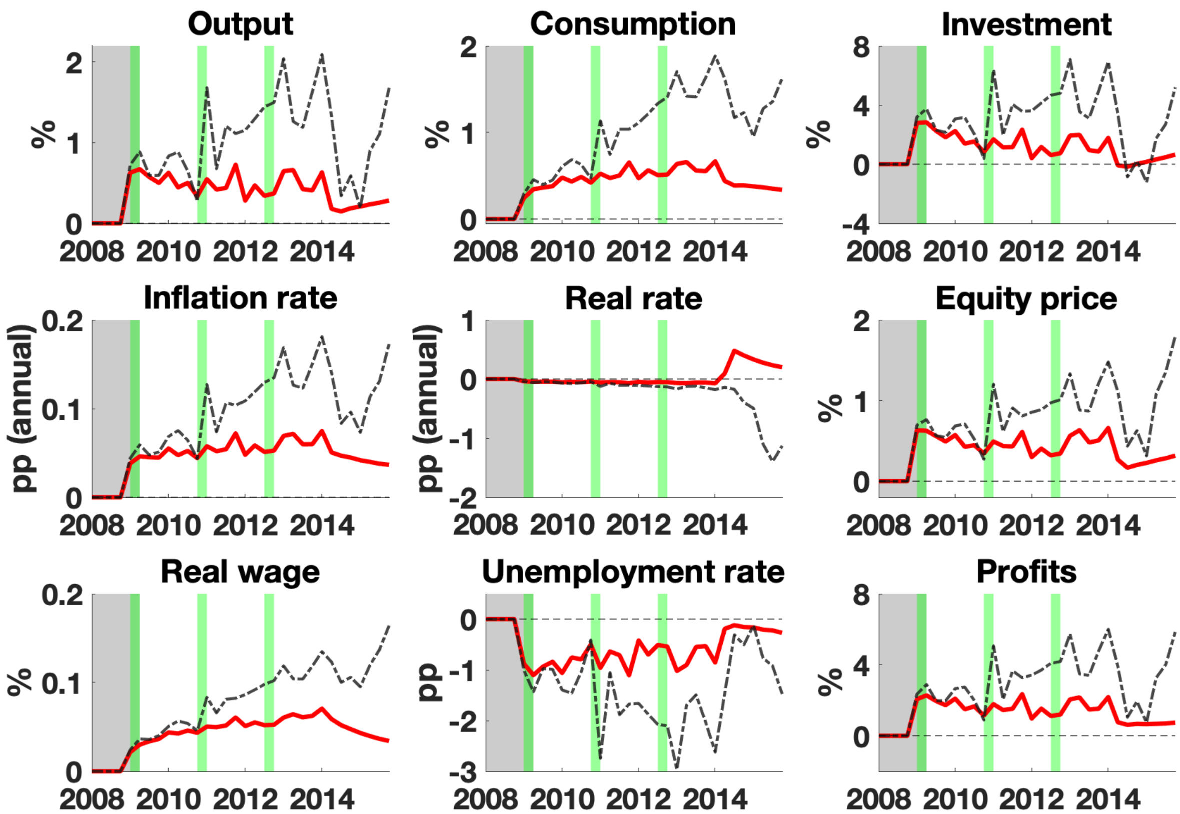
The results presented so far for QE are due not only to the central bank's asset purchases but also to exogenous ELB durations and maintaining the policy rate at zero for longer. In this section, I isolate the effects of asset purchases (pure QE) by simulating the model under the assumption that the expected ELB durations are endogenously determined and that the central bank sets a positive rate as soon as that is prescribed by the Taylor rule.[94]

6.1 Aggregate effects

Figure 13 shows that exogenous durations are longer than endogenous durations during the entire ELB episode, which implies that there was an additional stimulus from forward guidance. The difference between the endogenous and exogenous durations starts to increase in 2011, which is consistent with the finding of Jones (2017). These longer

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

Figure 14: Aggregate effects of forward guidance



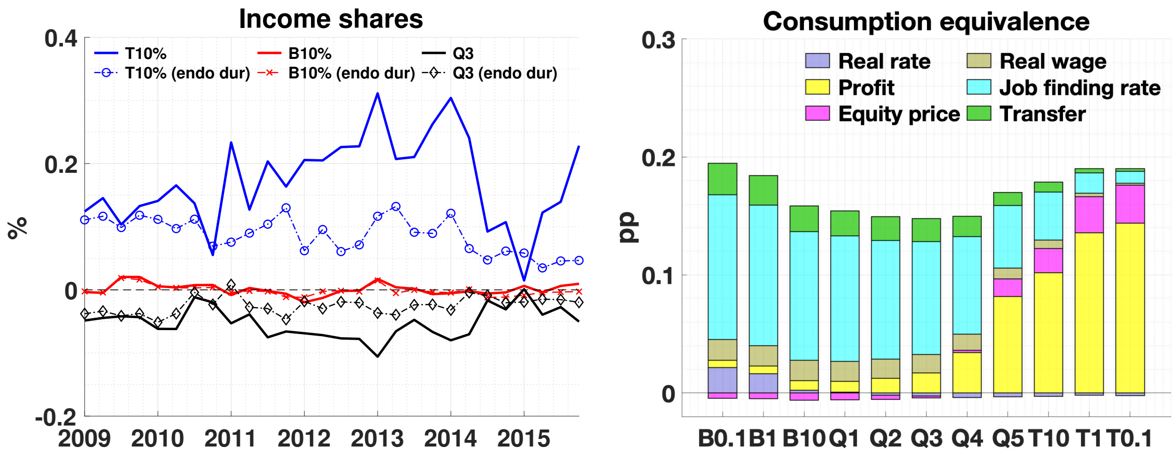
Notes: The solid red and dashed blak lines show the effects of asset purchases with endogenous ELB durations and asset purchases with exogenous durations. Except for inflation, the real interest rate, and the unemployment rate, variables are shown as percentage differences from the corresponding values in the alternative case with no policy interventions. The inflation rate, the real rate, and the unemployment rate are shown as the percentage point differences from their corresponding value in the alternative case.

expected ELB durations work similarly to expected future expansionary monetary policy shocks. Moreover, by maintaining the policy rates at zero, the central bank brings about a further stimulus to the economy around the end of the ELB episode.

When the additional stimulus effects are removed, asset purchases turn out to have much smaller aggregate effects. As discussed in the previous section, this is because asset purchases do not have strong direct stimulus effects on households and crowd out banks' investment.[95] Overall, the central bank's asset purchases account for about 45% of the total aggregate effects of unconventional monetary policies. Unlike the case of CMP, the

[95]Compared to the net effects of QE, conventional monetary policy has stronger overall real effects during the ELB episode.

Figure 15: Distributional effects of forward guidance: income share and CE



Notes: The left panel shows the income share of different wealth groups as the differences, relative to the corresponding shares in the counterfactual case with no policy interventions. The blue, red, and black straight lines show the shares in the case of QE and exogenous ELB durations. The blue, red, and black dashed lines with marks show the shares in the case of asset purchases and endogenous ELB durations. The right panel shows the additional welfare gains from forward guidance and their decomposition. The units are percentage points.

effects of forward guidance on key aggregate variables are very similar to those of the central bank's asset purchases in terms of the relative magnitudes since, in both cases, the policy rate is fixed at the ELB except for a few quarters at the end of the ELB episode.[96]

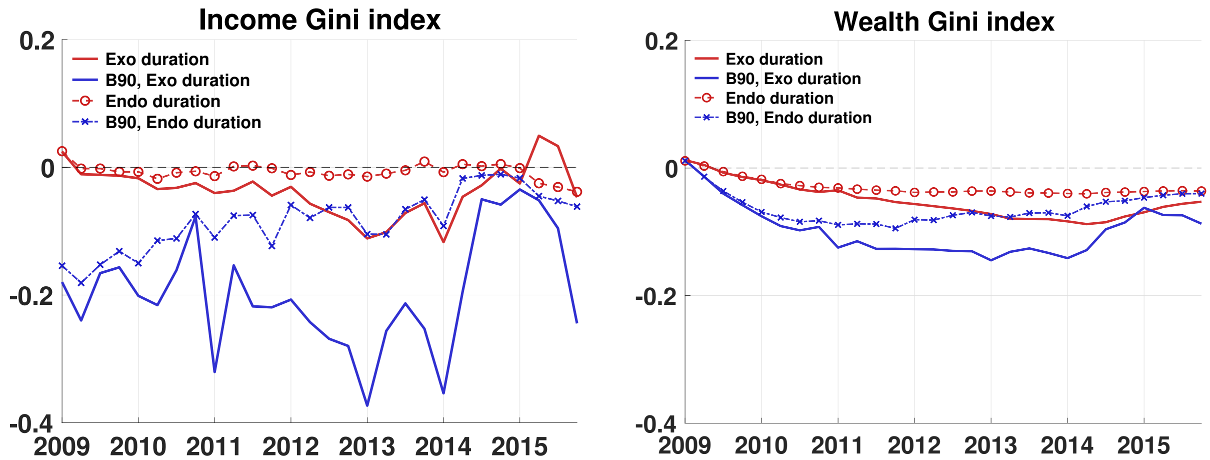
6.2 Distributional effects

Next I examine the effects of the additional stimulus from exogenous ELB durations on inequality and the distribution of welfare. The left panel of Figure 15 shows the differences in the income shares of different wealth groups between the case of full QE versus pure QE, without the additional stimulus coming from FG and lower rates for longer. Full QE amplified the distributional effects of the central bank's asset purchases both in terms of top income shares and in terms of the Gini index. The top 10% income share increased due to forward guidance, especially in later periods. In contrast, the income share of the middle quintile fell, while the income share of the bottom decile remained largely unchanged.

In the case of the Gini index, Figure 16 shows that forward guidance also amplifies the

[96]In the case of CMP, relative magnitudes of the responses of variables are similar to those of QE except for the real interest rate response.

Figure 16: Distributional effects of forward guidance: Gini index



Notes: The figure shows the evolution of the income and wealth Gini index under the different assumptions on the expected ELB durations. The blue and red lines show income and wealth Gini index as differences from their respective level in the counterfactual case with no policy interventions. The blue and red dashed lines with circles show income and consumption Gini index when the ELB durations are endogenous.

dynamics of the Gini indices induced by asset purchases. As the previous results imply, the Gini index is affected more by the dynamics among households in the bottom 90% of the wealth distribution. Since forward guidance further lowers the unemployment rate, it reduces the wealth and income gap between the bottom 10% and the middle. Accordingly, the degree of inequality, measured by the Gini indices, is lower under full QE, augmented by forward guidance, than under pure QE.

Finally, the welfare effect result is consistent with the result of inequality measures. Additional stimulus from forward guidance benefits both ends of the wealth distribution more than the middle. Specifically, forward guidance preserves the relative ranking of welfare gains across wealth groups and amplifies non-linear welfare effects resulting from asset purchases. I conclude that forward guidance amplifies the distributional effects of asset purchases, operating along the same channels.

7 Robustness check

The results presented so far can vary, depending on the parameter values or the specific assumptions on the model structure. In this section, I show that this paper's main results regarding the non-linear distributional effects of QE and sensitivity of the Gini index to dynamics in the lower 90% of the distribution are robust to variations in parametrizations and modeling assumptions.

Table 7: Robustness check 1/2

	Baseline	(1)	(2)	(3)	(4)	(5)	(6)
Aggregate variables							
Equity prices	0.81	0.80	0.82	0.72	0.71	0.77	0.84
Profits	3.04	2.81	2.77	1.93	1.51	3.01	3.09
Real wage	0.08	0.18	0.20	0.08	0.21	0.08	0.09
Unemployment rates	-1.34	-1.26	-1.27	-1.34	-1.25	-1.34	-1.33
Real rate	-0.23	-0.27	-0.28	-0.23	-0.28	-0.23	-0.25
Gini index							
Wealth	-0.05 (-0.09)	-0.07 (-0.13)	-0.08 (-0.14)	-0.07 (-0.11)	-0.10 (-0.18)	-0.10 (-0.19)	0.03 (0.09)
Income	**-0.04 (-0.12)**	**-0.05 (-0.15)**	**-0.06 (-0.16)**	**-0.13 (-0.25)**	**-0.18 (-0.36)**	**-0.25 (-0.54)**	**0.15 (0.26)**
Consumption	0.05 (0.10)	0.04 (0.08)	0.04 (0.08)	0.00 (-0.03)	-0.03 (-0.07)	-0.08 (-0.17)	0.11 (0.18)
Top 10% share							
Wealth	-0.01 (-0.03)	-0.03 (-0.06)	-0.03 (-0.07)	-0.03 (-0.06)	-0.06 (-0.11)	-0.05 (-0.11)	0.05 (0.12)
Income	**0.17 (0.31)**	**0.17 (0.32)**	**0.16 (0.31)**	**0.08 (0.16)**	**0.04 (0.12)**	**0.15 (0.24)**	**0.17 (0.31)**
Consumption	0.06 (0.10)	0.06 (0.10)	0.06 (0.09)	0.02 (0.05)	0.01 (0.03)	0.04 (0.07)	0.07 (0.10)
CE (Δ C during the ELB episode)							
T10	0.32 (1.16)	0.38 (1.35)	0.37 (1.35)	0.22 (0.88)	0.25 (0.99)	0.39 (1.23)	0.25 (1.05)
Q3	0.26 (0.81)	0.33 (0.99)	0.34 (1.01)	0.24 (0.75)	0.32 (0.95)	0.41 (0.96)	0.15 (0.64)
B10	0.28 (0.84)	0.36 (1.03)	0.37 (1.05)	0.27 (0.79)	0.36 (1.00)	0.49 (1.05)	0.15 (0.65)
Average	0.27 (0.90)	0.34 (1.07)	0.67 (1.09)	0.24 (0.78)	0.31 (0.96)	0.42 (1.03)	0.17 (0.74)

Notes: (1): High vacancy posting costs ($\iota = 0.063$), (2): Less rigid wage ($\rho_w = 0.65$), (3): Excluding banks' profit from aggregate profit, (4): (3) + Less rigid wage ($\rho_w = 0.65$), (5): Replacement ratio = 10% ($\upsilon = 0.1$), (6) Replacement ratio = 70% ($\upsilon = 0.7$). The table shows the average effects on aggregate variables, the Gini index, and the top 10% shares in each case. Except for the specified parameter value in each case, all other parameter values are set to values at the posterior mode. In all cases, shocks are re-filtered for a given set of parameters and observables. The average effects are shown as ratios to the effects on output. The Gini index, the top 10% shares, and consumption equivalents are shown as percentage points differences. The values in the parenthesis are the maximum or minimum of the corresponding variable during the ELB episode in the case of the Gini index and the top 10% shares. In the case of CEs, the values in the parenthesis show the amount of consumption increase during the ELB episode as percentage differences relative to the level of consumption in the counterfactual case of no unconventional monetary policies.

Table 7 summarizes a set of robustness test results, including the effects on key aggregate variables, the Gini index and the top 10% income share, and welfare effects on a set of households wealth groups. The aggregate effects are shown as the ratios to the

Table 8: Robustness check 2/2

	Baseline	(7)	(8)	(9)	(10)
Aggregate variables					
Equity prices	0.88	0.88	0.88	0.88	0.88
Profits	3.28	3.28	1.65	1.65	3.28
Real wage	0.09	1.09	0.09	1.09	0.09
Unemployment rates	-1.45	-1.45	-1.45	-1.45	-1.45
Real rate	-0.25	-0.25	-0.25	-0.25	-0.25
Gini index					
Wealth	-0.05 (-0.09)	-0.09 (-0.14)	-0.06 (-0.11)	-0.10 (-0.16)	-0.06 (-0.10)
Income	**-0.04** **(-0.12)**	**-0.12** **(-0.20)**	**-0.16** **(-0.34)**	**-0.25** **(-0.42)**	**-0.05** **(-0.13)**
Consumption	0.05 (+0.10)	-0.06 (-0.12)	-0.02 (-0.05)	-0.12 (-0.16)	0.04 (0.09)
Top 10% share					
Wealth	-0.01 (-0.03)	-0.06 (-0.10)	-0.03 (-0.06)	-0.07 (-0.13)	-0.02 (-0.03)
Income	**0.17** **(0.31)**	**0.06** **(0.20)**	**0.05** **(0.09)**	**-0.06** **(-0.13)**	**0.15** **(0.26)**
Consumption	0.06 (0.10)	-0.02 (-0.06)	0.01 (0.04)	-0.07 (-0.08)	0.06 (-0.1)
CE (Δ C during the ELB episode)					
T10	0.32 (1.16)	0.81 (1.67)	0.23 (0.87)	0.72 (1.37)	0.32 (1.16)
Q3	0.26 (0.81)	1.30 (1.89)	0.25 (0.78)	1.29 (1.86)	0.27 (0.82)
B10	0.28 (0.84)	1.37 (1.98)	0.28 (0.82)	1.37 (1.95)	0.29 (0.86)
Average	0.27 (0.90)	1.23 (1.84)	0.25 (0.80)	1.20 (1.74)	0.28 (0.90)

Notes: (7): 1 percentage point higher real wage, (8) 50% lower profits, (9) 1 percentage point higher real wage + 50% lower profits, (10) Assuming the steady state distribution in 2009 Q1. The table shows the average effects of QE on the Gini index and the top 10% shares during the ELB episode. All the effects are computed from the micro-level simulation only. That is, without estimating the aggregate effects, assumed effects are applied to households' distribution. The Gini index, the top 10% shares, and consumption equivalents are shown as percentage points differences. The values in the parenthesis are the maximum or minimum of the corresponding variable during the ELB episode in the case of the Gini index and the top 10% shares. In the case of CEs, the values in the parenthesis show the amount of consumption increase during the ELB episode as percentage differences relative to the level of consumption in the counterfactual case of no unconventional monetary policies.

effects on output. For instance, 0.81 in the third row and the second column (from the left) implies that, in the baseline case, equity prices are, on average, 0.81% higher when output is, on average, 1% higher during the ELB episode compared to when the central bank did not conduct QE. In the case of unemployment rates and real interest rates, the effects are measured as percentage points. The second column reproduces the baseline resultsfrom Section 4.

In the third column, I double the vacancy posting costs to make extensive margin adjustment of labor more costly. In the fourth column, the degree of wage rigidity is set to 0.65, versus the baseline value of 0.8. In the fifth column, I assume that profits from the financial sector are not distributed to equity holders or business owners. Thus, I remove any direct effects of banks' profits on households' inequality or welfare. In the sixth column, I assume that the wage is less rigid, i.e., $\rho_w = 0.65$, and exclude banks' profits from the aggregate profits. In the last two columns, the replacement ratio for the unemployment benefit is set to 10 and 70%, respectively, while maintaining other parameter values at the posterior mode.

In Table 8, I evaluate the distributional consequences of QE by assuming different paths of profits and real wages. In the third column, I assume that the effects of QE on real wages are one percentage point greater than in the baseline case. In the fourth column, I instead assume that the effects of QE on profits are 50% smaller. In this case, profits are only 1.65% higher than the counterfactual case with no unconventional monetary policies. In the fifth column, I adopt both assumptions. Thus, real wages are, on average, 1.09% higher, while profits are only 1.65% higher than in the counterfactual case. In the last column, I use the steady state households' distribution instead of the model's distribution at the beginning of 2009 Q1 to see if the differences in households' distribution have a significant impact on the main results of the paper.

As the tables show, the main result of this paper is maintained in most cases: QE increases the top 10% income and consumption share while also reducing overall income and wealth inequality, as measured by the Gini index. When vacancy posting costs are higher, or the degree of wage rigidity is lower, the relative magnitudes of the increase in real wages are larger than in the baseline case.[97] As a result, income inequality, as measured by the Gini index, declines by a greater amount than in the baseline case. However, even in these cases, the top 10% income and consumption shares increase because of QE relative to a case without unconventional monetary policies. Similarly, when profits from the financial sector are excluded from aggregate profits, relative increases in profits are smaller. Thus, the income Gini index decreases by a larger amount. When the wage is less rigid, the income Gini falls, on average, by 0.18 percentage points, which is almost four times larger magnitude than that in the baseline case. However, the top 10% income and consumption shares still rise even though the magnitudes are smaller than in the baseline case.

In terms of welfare gains, U-shaped effects are not preserved when banks' profits are

[97] When the vacancy posting cost is high, firms utilize capita more, which increases labor demand by the complementarity between inputs. Thus, the real wage rises by more compare to when the vacancy posting cost is smaller.

excluded from aggregate profits, and thus, not distributed to business owners or equity holders (case (3) & (4)). In those cases, consumption equivalents are decreasing in wealth, unlike in the baseline case. However, even in these cases, the top 10% households experience a higher consumption increase than the middle quintile during the ELB episode. That is, income and consumption gains are U-shaped during the ELB episode, but the expected contractionary effects of tapering mostly offset the welfare gains of wealthy households. Accordingly, the long-run welfare effects become monotonic in these cases. In cases (7) and (9), welfare effects are also monotonically decreasing in wealth. However, in these cases, the assumed wage increases are implausibly high given the degree of wage rigidity that I find via estimation.

Interestingly, I find that fiscal policy, specifically the extent of the unemployment benefit, matters for the distributional consequences of QE. When the replacement rate is only 10%, QE reduces the income Gini index by a much larger amount compared to other cases. Also, welfare gains from QE for the bottom 10% households are much larger than their welfare gains in the baseline case, even though the magnitudes of the unemployment rate and real wage responses are similar. This is because, when the unemployment benefit is relatively smaller, income gain from being employed is much larger. For the same reasoning, if the unemployment benefit is much larger, then income gain from being employed is substantially smaller, and thus, QE mostly increases income and consumption inequality without providing significant benefits to the bottom 90% households.

To recapitulate, the main results that I find in this paper hold across different parametrizations and modeling assumptions unless the unemployment benefit is improbably high or the relative magnitudes of profits and wage responses implausibly similar. Also, the deviations of the households' distribution from the steady state distribution do not have any significant impact on the main results on the distributional consequences of QE.

8 Conclusion

In this paper, I examine the distributional consequences of QE during the ELB episode that followed the Great Recession in the U.S. To this end, I develop a medium-scale HANK model that features portfolio choice, wage rigidity, labor market frictions, banks, and a zero lower bound on the policy rate. I model quantitative easing as central bank private asset purchases, as in Gertler and Karadi (2011), and forward guidance as exogenous expected ELB durations, as in Jones (2017). I parametrize the model to match the micro-data on households' wealth and income composition. Moreover, to discipline the model's parameters associated with the dynamics of key aggregate variables, such as the real wage, the unemployment rate, and profits, I estimate the model with the macro data on the U.S. economy using Bayesian methods.

The estimated model generates empirically plausible dynamics of wages, unemployment, and profits to exogenous shocks. In particular,it generates a procyclical response of profits to an expansionary monetary policy shock, unlike most existing New Keynesian models. Because of this, the model uncovers wealthy households' substantial benefits from expansionary monetary policy that existing New Keynesian models cannot capture.

A counterfactual analysis reveals that QE reduced the wealth and income Gini indices during the ELB episode, mainly via its positive impacts on employment. However, at the same time, QE widens the income gap between the top 10% and the bottom 90% by substantially increasing profits and equity prices.

The results of this paper suggests that both the criticism regarding the adverse effects of QE on inequality and the counterargument based on QE's positive impacts on labor markets can be justified, depending on the focus. If one focuses on the gap between the top 10% and all other households, QE can be seen as increasing inequality. If one focuses on the improvement of welfare at the bottom, QE can be seen as reducing inequality, as it reduces the gap between the bottom 10% and the middle of the wealth distribution. Importantly, the result also implies that if a model fails to capture wealthy households' benefit from monetary policy, an analysis based on it can lead to a misleading or incomplete conclusion on the effects of monetary policy on inequality.

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., 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.

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

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.

Cui, W. and Sterk, V. (2018). Quantitative easing. Discussion Papers 1830, Centre for Macroeconomics (CFM).

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

Del Negro, M., Eggertsson, G., Ferrero, A., and Kiyotaki, N. (2017). The great escape? a quantitative evaluation of the fed's liquidity facilities. *American Economic Review*, 107(3):824–57.

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.

Hohberger, S., Priftis, R., and Vogel, L. (2020). The distributional effects of conventional monetary policy and quantitative easing: Evidence from an estimated DSGE model. *Journal of Banking & Finance*, 113(C).

Holston, K., Laubach, T., and Williams, J. C. (2017). Measuring the natural rate of interest: International trends and determinants. *Journal of International Economics*, 108:S59–S75.

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.

Laubach, T. and Williams, J. C. (2016). Measuring the Natural Rate of Interest Redux. *Business Economics*, 51(2):57–67.

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.

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

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.

Taghizadeh-Hesary, F., Yoshino, N., and Shimizu, S. (2020). The impact of monetary and tax policy on income inequality in japan. *The World Economy*, 43(10):2600–2621.

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

Wallace, N. (1981). A modigliani-miller theorem for open-market operations. *The American Economic Review*, 71(3):267–274.

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.[98] 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.

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

[98] 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.

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

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

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.

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

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.

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

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

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

where Θ_t is the bank's leverage ratio, i.e., the ratio of assets to its net worth.[99] When the constraint binds, I can express the law of motion for net worth as follows.

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

[99]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.

In addition, it follows that

$$z_{t,t+1} = N_{jt+1}/N_{jt} = \left\{(R^a_{t+1} - R_{t+1})\Theta_t + R_{t+1}\right\} \tag{A.15}$$

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

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

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

where A^b_{t+1} 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 $(1-\theta_b)q_{t-1}A^b_t$. Specifically, for each new bank, the equity mutual fund gives $\omega/(1-\theta_b)$ fraction of this value. Then, I have

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

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

$$\Pi^b_t = (1-\theta_b)\{(R^a_t - R_t)\Theta_{t-1} + R_t\}N_{t-1} - \omega q_{t-1}A^b_t \tag{A.19}$$

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).[100] The methodology enables a fast solution that is necessary for Bayesian estimation. However, since the state-space is much larger compared to a representative model even after the reduction, estimating the model by solving the dynamics in full each time during the process is still not feasible.[101] 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.[102] 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

[100] 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.

[101] 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.

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

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 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.[103] 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 house-

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

holds' optimal behaviors. Specifically, we only need to update the following derivatives: $\frac{\partial \mathcal{F}_{3,t}}{\partial X_{2t+1}}$, $\frac{\partial \mathcal{F}_{3,t}}{\partial X_{3t+1}}$, $\frac{\partial \mathcal{F}_{3,t}}{\partial \epsilon_{t+1}}$, $\frac{\partial \mathcal{F}_{3,t}}{\partial Y_{2t+1}}$, $\frac{\partial \mathcal{F}_{3,t}}{\partial Y_{3t+1}}$, $\frac{\partial \mathcal{F}_{3,t}}{\partial X_{2t}}$, $\frac{\partial \mathcal{F}_{3,t}}{\partial X_{3t}}$, $\frac{\partial \mathcal{F}_{3,t}}{\partial \epsilon_{t}}$, $\frac{\partial \mathcal{F}_{3,t}}{\partial Y_{2t}}$, $\frac{\partial \mathcal{F}_{3,t}}{\partial Y_{3t}}$, $\frac{\partial \mathcal{F}_{4,t}}{\partial \epsilon_{t}}$, $\frac{\partial \mathcal{F}_{7,t}}{\partial X_{2t+1}}$, $\frac{\partial \mathcal{F}_{7,t}}{\partial X_{3t+1}}$, $\frac{\partial \mathcal{F}_{7,t}}{\partial \epsilon_{t+1}}$, $\frac{\partial \mathcal{F}_{7,t}}{\partial Y_{2t+1}}$, $\frac{\partial \mathcal{F}_{7,t}}{\partial Y_{3t+1}}$. Then, the number of equations that we need to evaluate is close to the number of equations in a representative model with the same features. Thus, estimating the model using Bayesian method is possible.

B.2 Inversion filter

In this paper, I use an inversion filter to back out the structural shocks, following Guerrieri and Iacoviello (2017) and Cuba-Borda et al. (2019). Let $\mathbb{Y}_{\{1:T\}} = \{Y_1, Y_2, ..., Y_T\}$ denote the set of observables, where Y_j is the $n_y \times 1$ vector that contains the data on n_y observables in period j for $j = 1, ...$ T. Also, denote the set of all the endogenous variables of the model in period t with the $n_x \times 1$ vector X_t. Similarly, 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.[104] 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.

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

[104] 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.

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 corresponding structural shocks are conditional on the duration T of the ZLB episodes. Once I find the series of shocks using the filter, I compute the likelihood of the model given the data as follows.

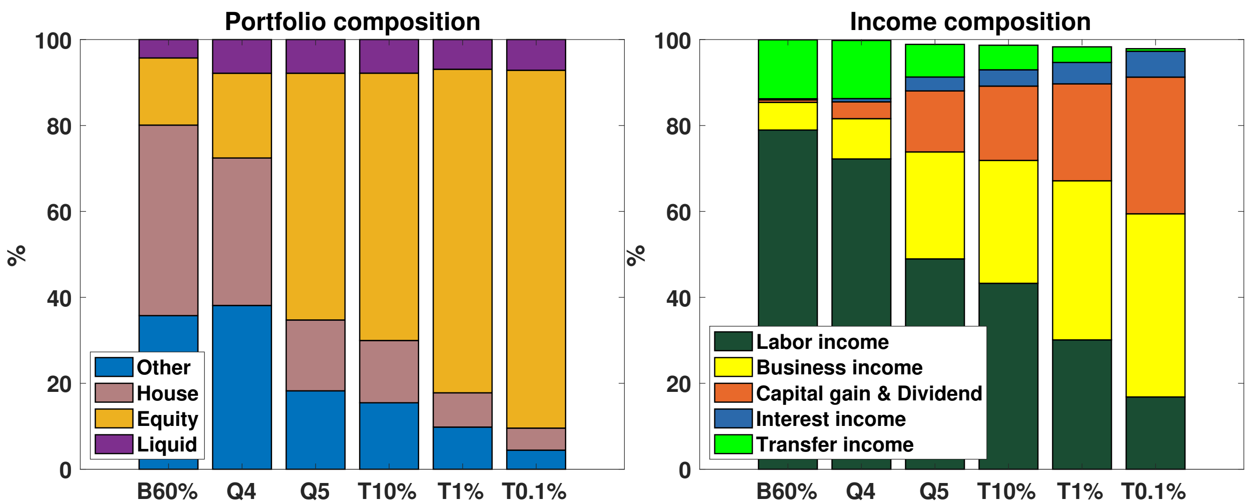
$$\log p\Big(\mathbb{Y}_{\{1:T\}}\Big) = -\frac{Tn_y}{2}\log(2\pi) - \frac{T}{2}\log(\det(\Sigma)) - \frac{1}{2}\sum_{t=1}^{T}\epsilon_t'\Sigma^{-1}\epsilon_t + \sum_{t=1}^{T}\log\left(|\det\frac{\partial\epsilon_t}{\partial Y_t}|\right) \tag{A.28}$$

where $\frac{\partial\epsilon_t}{\partial Y_t} = \Big\{H_t Q_t\Big\}^{-1}$.[105]

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

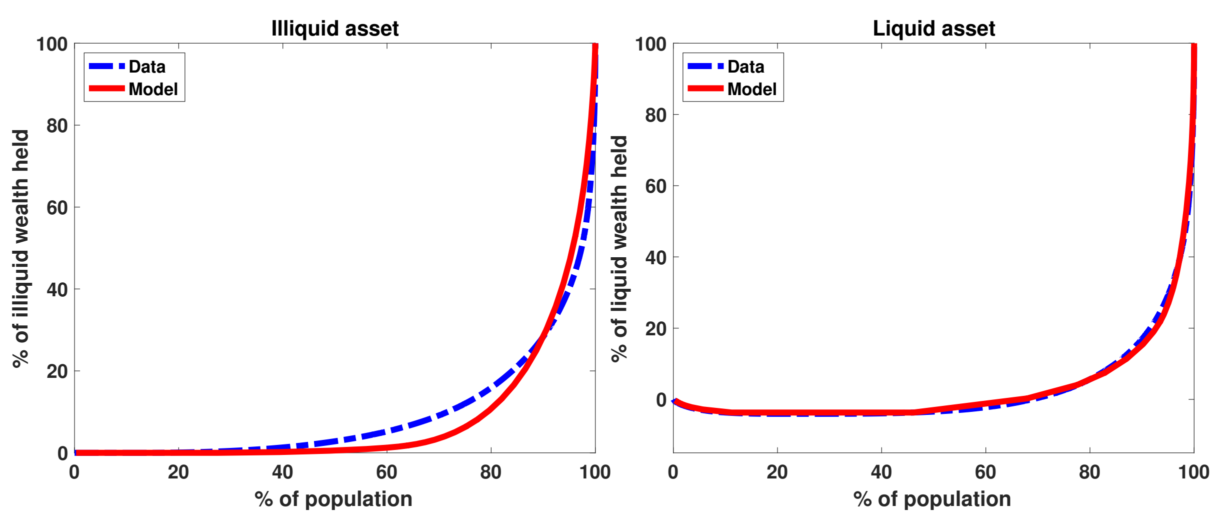
C Fit of the model

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.

D Further details on the estimation

D.1 Observables and a mapping between the data and the model

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

1. Output
   - Model : $\tilde{Y}_t^{\text{obs}} = \log\left(\frac{Y_t}{Y_{t-1}}\right)$
   - Data : Nominal GDP (FRED, GDP), divided by GDP deflator (FRED, GDPDEF) and civilian non-institutionalized population (FRED, CNP16OV), log-transformed, first-differenced and de-meaned.
2. Consumption
   - Model : $\tilde{C}_t^{\text{obs}} = \log\left(\frac{C_t}{C_{t-1}}\right)$
   - Data : The sum of PCE on non-durable goods and services (BEA NIPA Table 2.3.5, item 8 & 13), divided by GDP deflator (FRED, GDPDEF) and civilian non-institutionalized population (FRED, CNP16OV), log-transformed, first-differenced and de-meaned.
3. Investment
   - Model : $\tilde{I}_t^{\text{obs}} = \log\left(\frac{I_t}{I_{t-1}}\right)$
   - Data : The sum of private fixed investment (BEA NIPA Table 5.3.5, all types) and PCE on durable goods (BEA NIPA Table 2.3.5, item 3), divided by GDP deflator (FRED, GDPDEF) and civilian non-institutionalized population (FRED, CNP16OV), log-transformed, first-differenced and de-meaned.
4. Inflation rate
   - Model : $\tilde{\pi}_t^{\text{obs}} = \log\left(\frac{\pi_t}{\pi}\right)$
   - Data : Log difference of GDP Implicit Price Deflator (FRED, GDPDEF) minus 0.5 percentage point.
5. Interest rate
   - Model : $\tilde{i}_t^{\text{obs}} = \log\left(\frac{R_t}{R}\right)$

- Data : Effective Federal Funds Rate, divided by 400 to express in quarterly units minus logarithm of the model's steady state nominal rate.

6. Real wage
   - Model : $\tilde{w}_t^{\text{obs}} = \log\left(\frac{w_t}{w_{t-1}}\right)$
   - Data : Average hourly earnings of production and non-supervisory employees in total private sector (FRED, AHETPI), divided by GDP deflator (FRED, GDPDEF), log-transformed, first-differenced and de-meaned.
7. Unemployment rate
   - Model : $\tilde{u}_t^{\text{obs}} = \log\left(\frac{u_t}{u}\right)$
   - Data : Unemployment as the number of unemployed as a percentage of the labor force (FRED, UNRATE) minus minus 5 percent divided by 100.
8. Lump-sum transfer
   - Model : $\tilde{T}_t^{\text{obs}} = \log\left(\frac{T_t^g}{T_{t-1}^g}\right)$
   - Data : The sum of government's current transfer payment (BEA NIPA table 3.2, item 26), capital transfer payments (item 22), net of current transfer receipts (item 19), capital transfer receipts (item 42), and unemployment benefit (NIPA underlying table 3.12U, item 7), divided by GDP deflator (FRED, GDPDEF) and civilian non-institutionalized population (FRED, CNP16OV), log-transformed, first-differenced and de-meaned.
9. Profits
   - Model : $\tilde{\Pi}_t^{\text{obs}} = \log\left(\frac{\Pi_t}{\Pi_{t-1}}\right)$
   - Data : Corporate profits after tax with inventory valuation adjustment and capital consumption adjustment (BEA account code: A551RC), divided by GDP deflator (FRED, GDPDEF), and civilian non-institutionalized population (FRED, CNP16OV), log-transformed, first-differenced and de-meaned.
10. Central bank's assets
   - Model : $\tilde{A}_{t+1}^{\text{CB,obs}} = \log\left(\frac{A_{t+1}^{\text{CB}}}{A_{2007}^{\text{CB}}}\right)$
   - Data : All Federal Bank's assets (FRED, WALCL), divided by GDP deflator (GDP deflator), civilian non-institutionalized population (CNP16OV), and its end of 2007 level. Log-transformed

D.2 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

$$\log(\Psi^k_t) = \rho_k \log(\Psi^k_{t-1}) + \epsilon_{k,t} \ \ , \sim \ N(0, \sigma^2_k) \tag{A.34}$$

5. Liquidity preference shock

$$\log(\Psi^l_t) = \rho_l \log(\Psi^l_{t-1}) + \epsilon_{l,t} \ , \ \epsilon_{l,t} \ \sim \ N(0, \sigma^2_l) \tag{A.35}$$

6. Wage shock

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

(A.37)

7. Lump-sum transfer shock

$$T^g_t = \left(1 - \frac{1}{\Psi^g_t}\right) Y \tag{A.38}$$

$$\log\left(\frac{\Psi^g_t}{\Psi^g}\right) = \rho_g \log\left(\frac{\Psi^g_{t-1}}{\Psi^g}\right) + \epsilon_{g,t} \ , \epsilon_{g,t} \ \sim \ N(0, \sigma^2_g) \tag{A.39}$$

8. Monetary policy shock

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

$$i_{t+1} = \min\{0, \hat{i}_{t+1}\} \quad \text{(A.41)}$$

9. Fixed cost shock

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

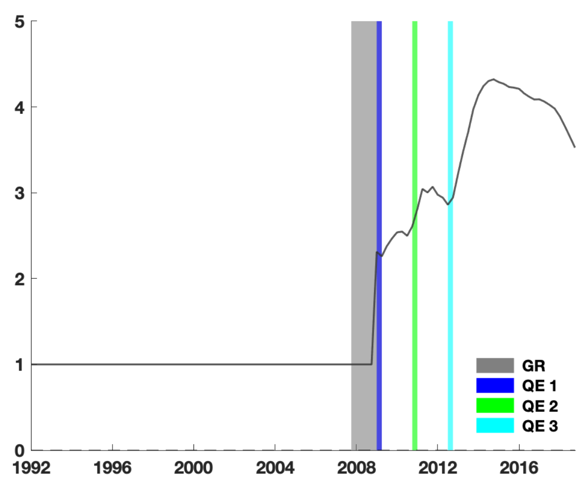
$$\quad \text{(A.43)}$$

10. QE shock

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

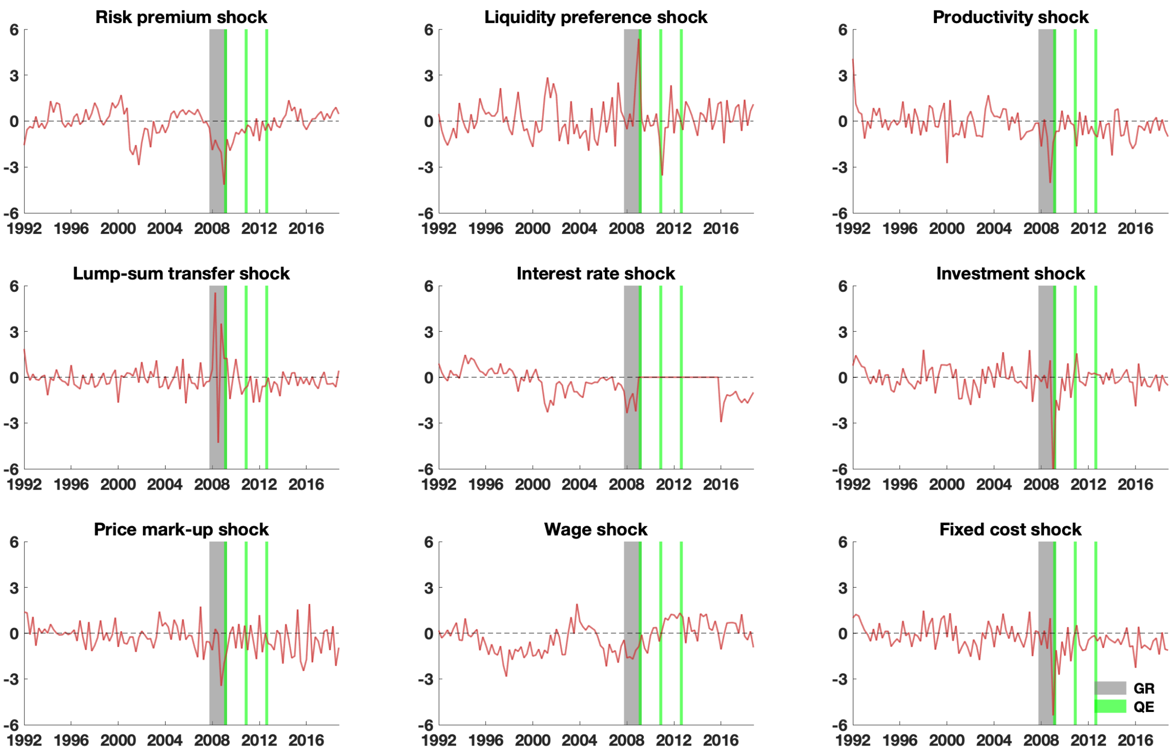
D.3 Additional figures and tables

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



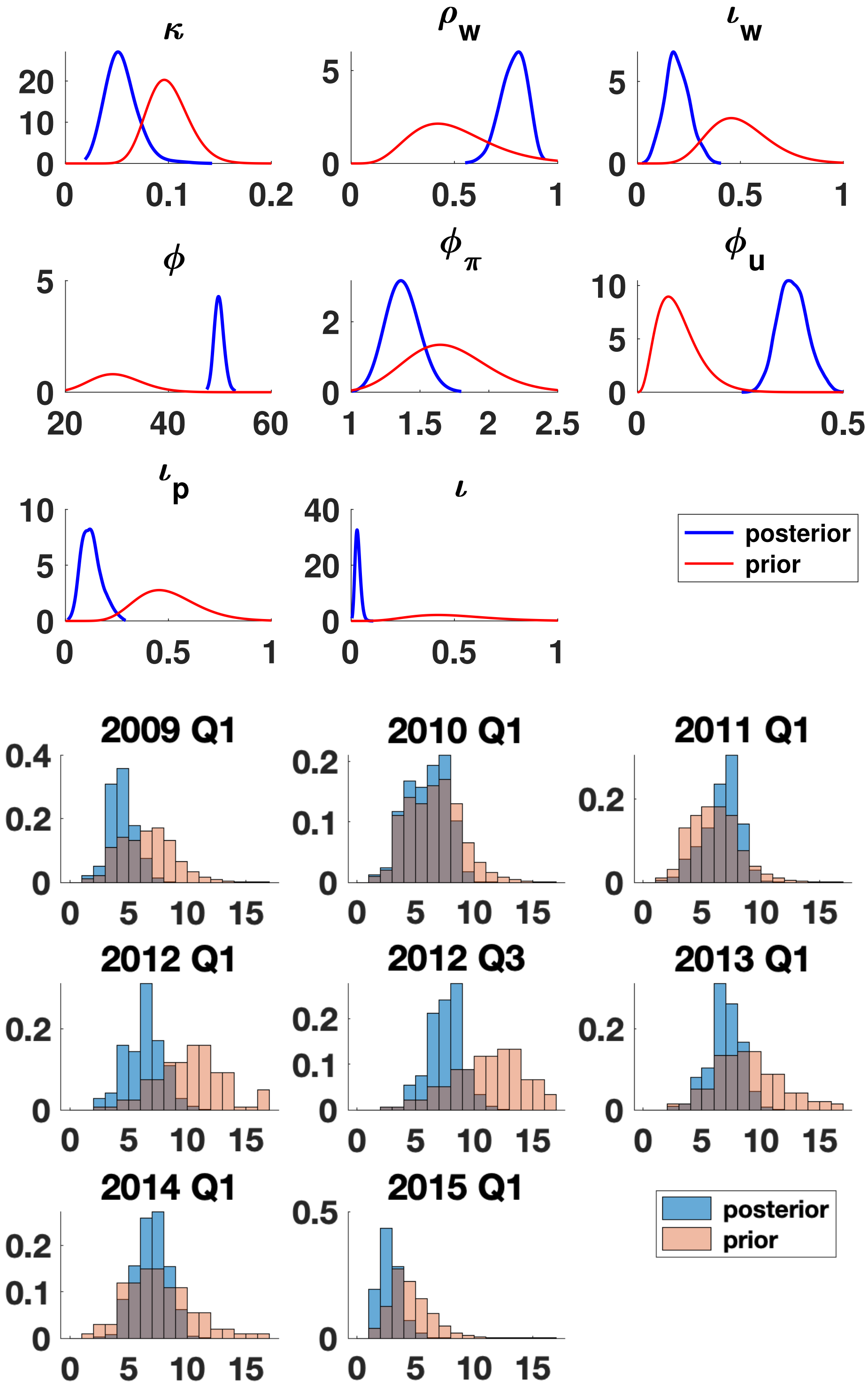
Notes: The figure shows the time series of the filtered shocks during the sample periods as a ratio to its standard deviation. The shaded gray area represents the periods of the Great Recession. The transparent green bars represent the quarters in which QE 1, 2, and 3 are announced or implemented.

Table A1: Prior and posterior distributions of expected ELB durations

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

Notes: The unit is one quarter.

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



E Model Dynamics

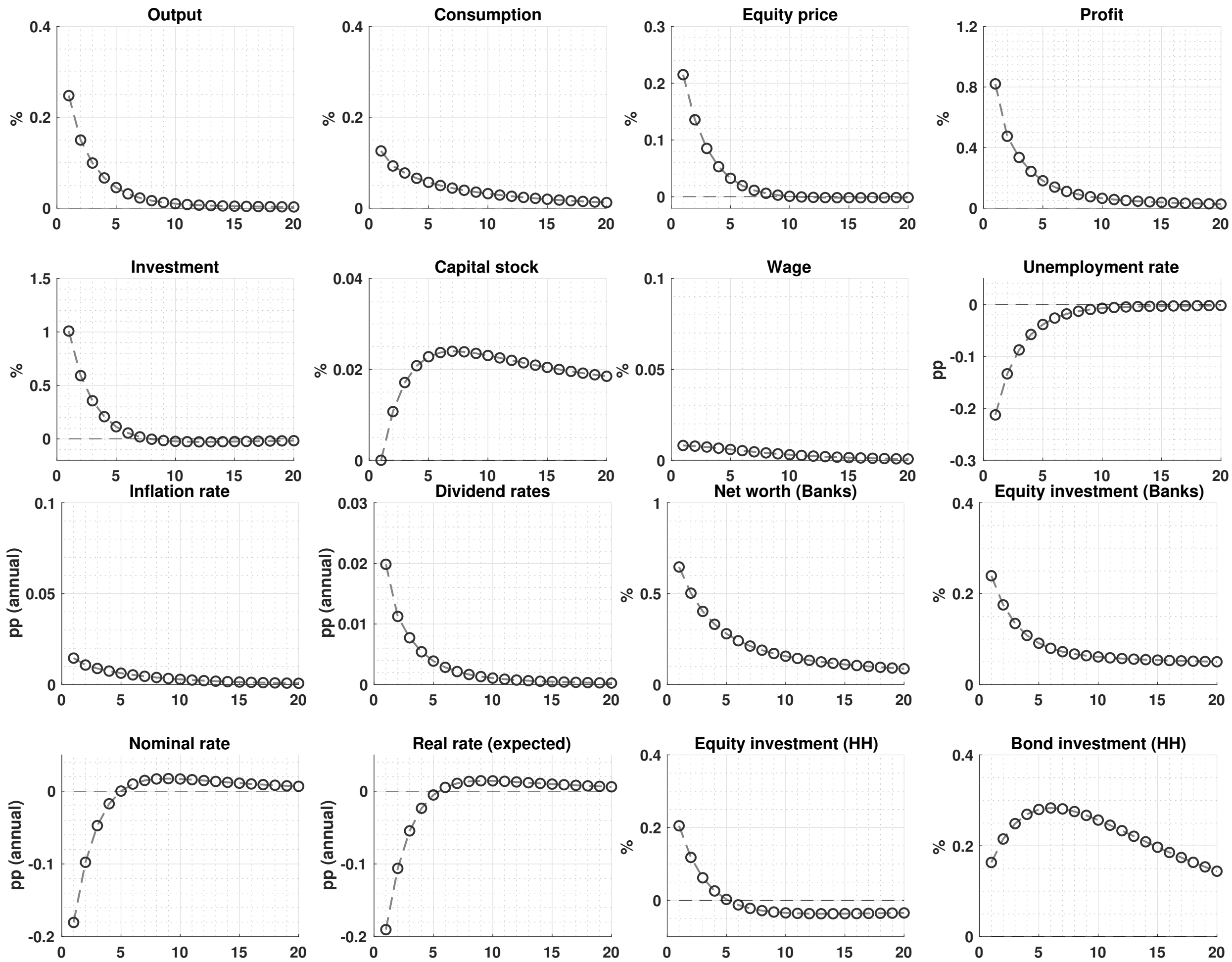
A countercyclical response of profits to demand shocks is a common feature of New Keynesian models. Since the factor prices are relatively flexible while the price is assumed to be rigid, a markup of the price over marginal cost is countercyclical in New Keynesian models when demand shocks, such as monetary policy and government spending shocks, occur. Consequently, profits fall after an increase in aggregate demand.[106] 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.

Figure A6: Impulse responses to an expansionary monetary policy shock



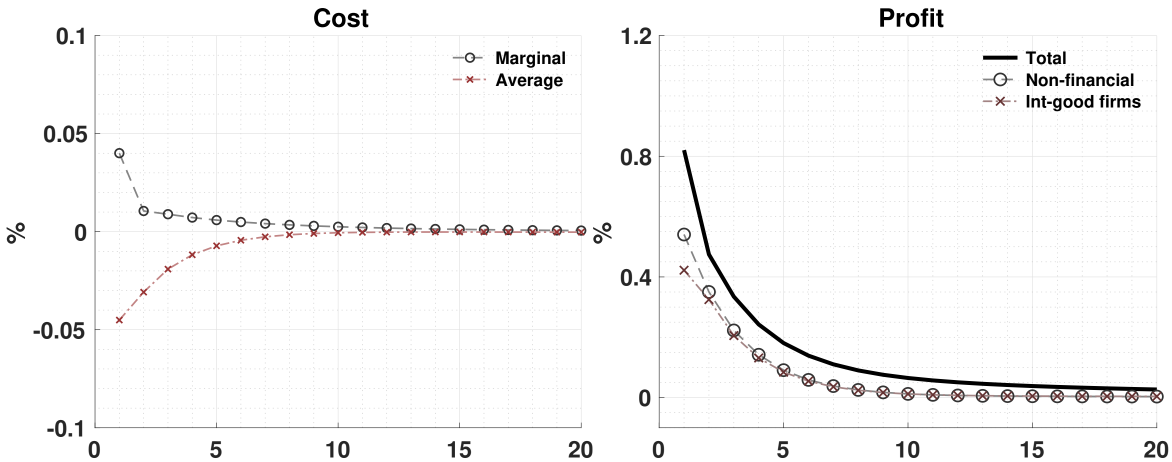
Notes: The figure shows the model's impulse responses to a negative 25 basis points (annualized) interest rate shock. All variables are shown as the percentage deviations from their respective steady state values except for the nominal rate, the inflation rate, the dividend rates and the unemployment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

E.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 negative interest rate shock occurs, profits substantially increase in the model.

[106] 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 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.

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.[107]

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.[108] However, in the model, the

[107]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.

[108]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

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. [109]

Besides, based on a recent finding of Anderson et al. (2018), I assume that the fixed cost accounts for a significant proportion of the total production cost.[110] 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.[111] 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.[112] When the interest rate falls and investment increases, the equity price increases, and thus the

literature.

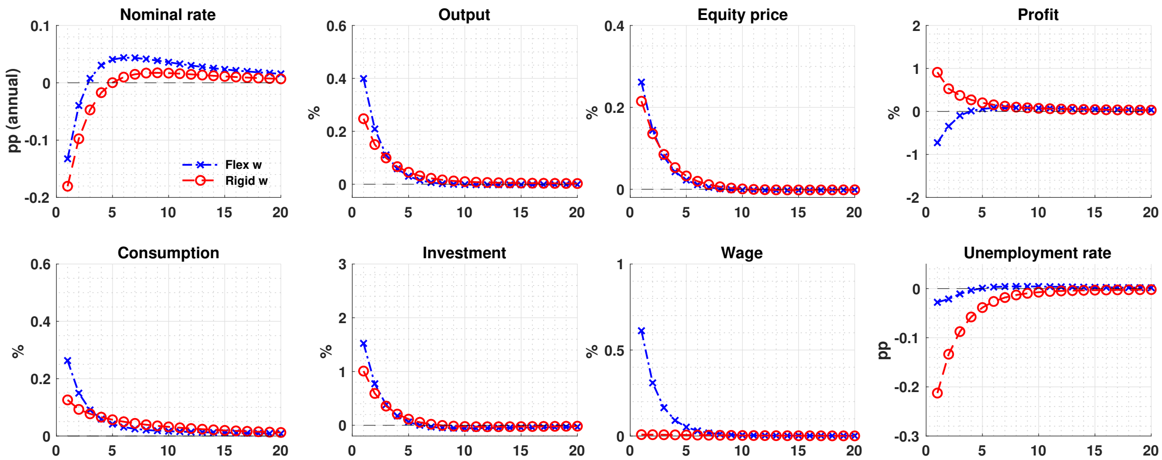
[109]Note that the marginal cost for intermediate good firms is determined by the capital and labor rental rate, and I do not impose any rigidity on the labor rental rate. However, wage rigidity and labor supply via labor agencies effectively increase the elasticity of labor supply with respect to changes 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.

[110]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.

[111]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.

[112]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.

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.

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.[113] In the process, banks also lead to strong investment responses. Thus, even though consumption response is relatively small due to a weak redistribution and the wage rigidity, the overall demand of goods can increase significantly because of banks' investment demand.

E.2 Comparison with a model with the flexible wage

Figure 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 A2 shows the values of key parameters at the posterior mode.

Two things are noticeable in the figure. First, depending on the assumption of wage rigidity, the response of profits is entirely different. When the wage is assumed to be

[113] 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 A2: 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

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.[114]

[114]The role of the wage rigidity recently regained attention in the literature. Broer et al. (2019) advocate

F Structural VAR analysis

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

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

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

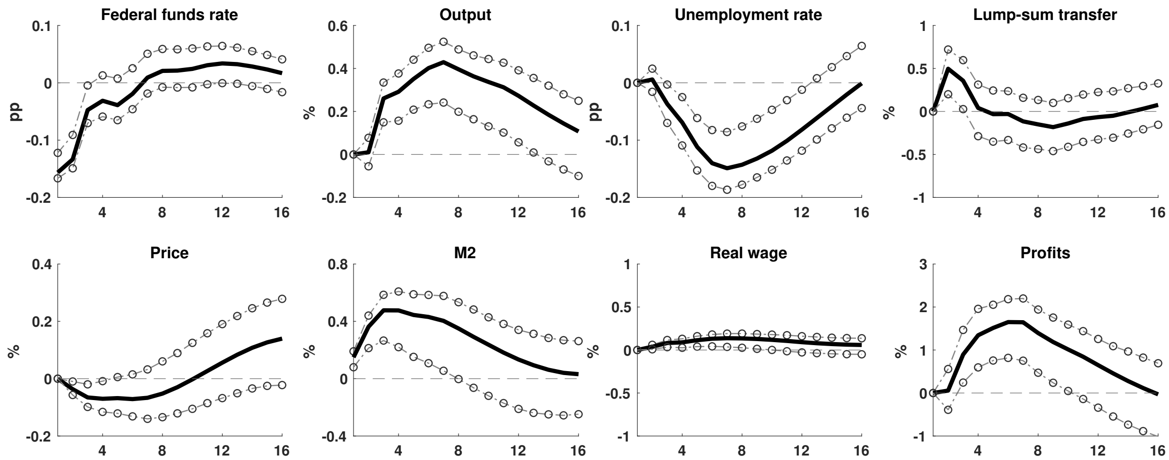
where f is the feedback rule, Ω_t is the information set available to the central bank in period t, and $\epsilon_{r,t}$ is an exogenous shock to the policy decision. Let $\mathcal{Y}_t$ denote the vector of the variables included in the VAR model.

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

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-

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.

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.[115] 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

[115]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.

periods after the shock, but the responses are mostly statistically insignificant. The corresponding variables in the model exhibit similar dynamics except for the lack of hump-shaped responses, which is a common limitation of the most of existing HANK models in the literature.[116] Most of variables in the SVAR model peaks between 4th and 8th quarters after the shock. In contrast, in the model, responses are immediate.

G Further details on the results

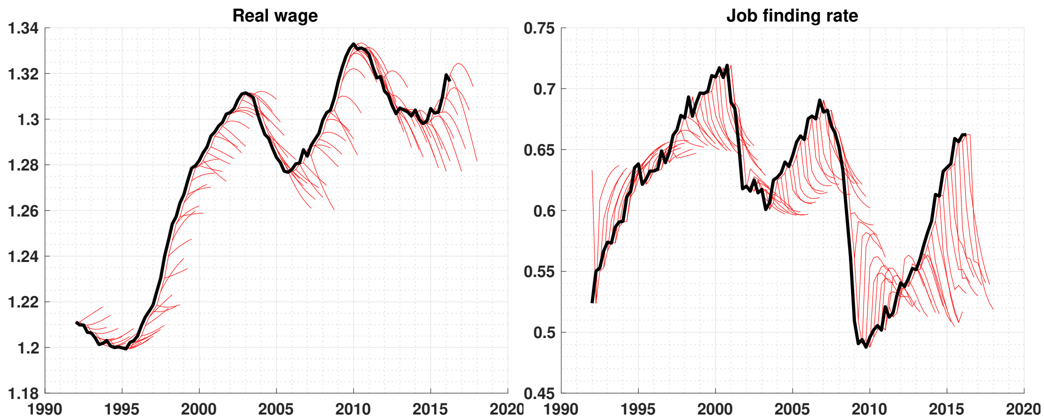
G.1 The decomposition method

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 .

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

[116]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.

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 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.[117] 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.[118] Figure A11

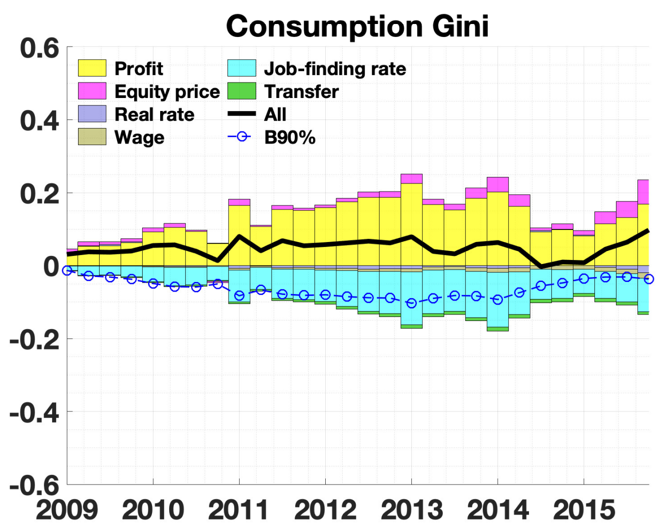
[117]The number of the expected paths is equal to the number of periods in the sample multiplied by the number of variables. The starting value of each path, i.e., $x_{t,t}$, coincides with the realized value as it is observed, but all the future expectations are not necessarily correct because of unexpected shocks in the future. That is, $x_{t,t+1}$ is, in general, different from $x_{t+1,t+1}$.

[118]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

shows the realized path of the real wage and the job-finding rate along with each period's household expectations on it.

G.2 Additional figures

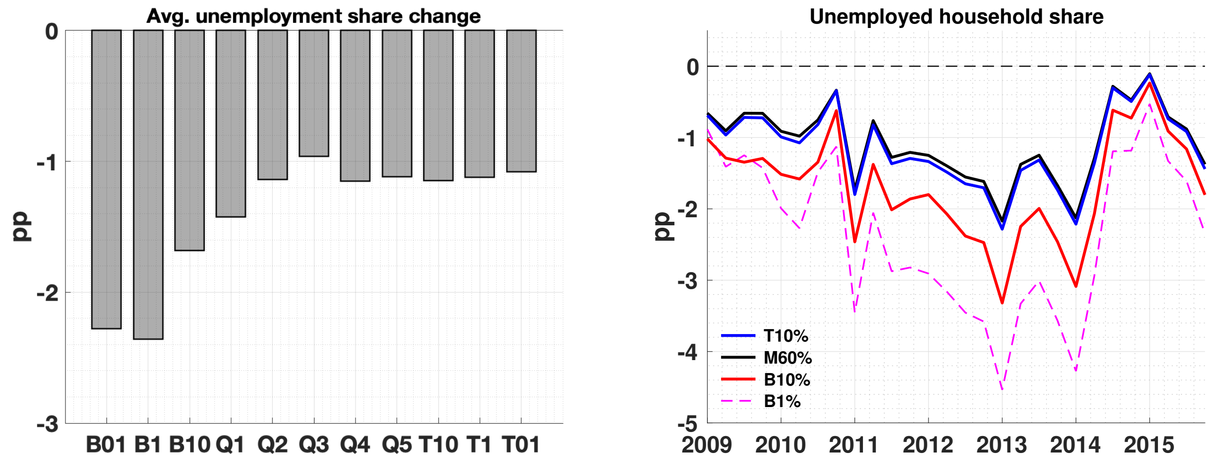
Figure A12: Distributional effects of QE: Gini index



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 QE, 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.

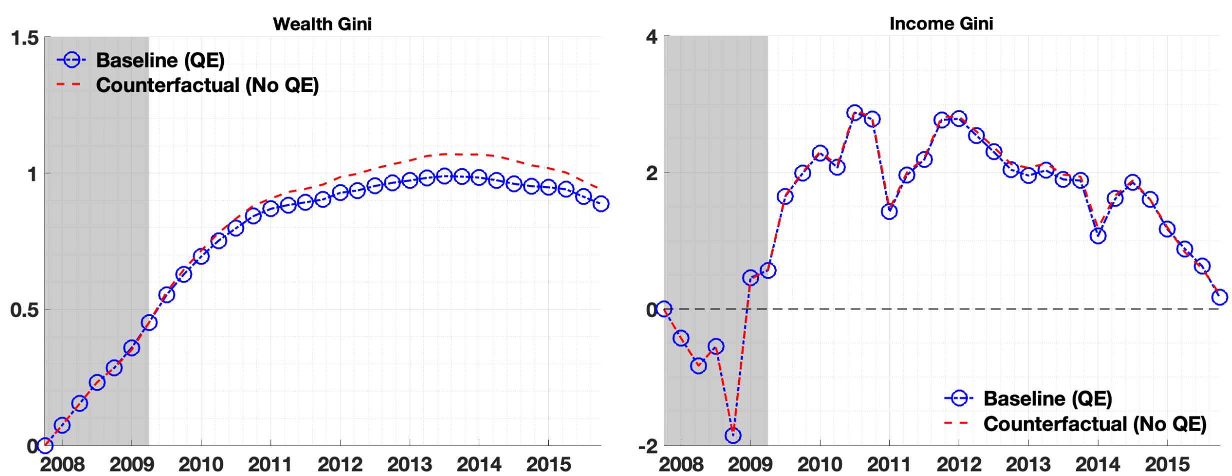
combination, only the job-finding rate follows the baseline path. In the seventh combination, only the equity price follows the baseline path. Finally, in the eighth combination, only the total transfer follows the baseline path, and all others follow the paths in the counter-factual case.

Figure A13: Unemployed household shares across wealth groups



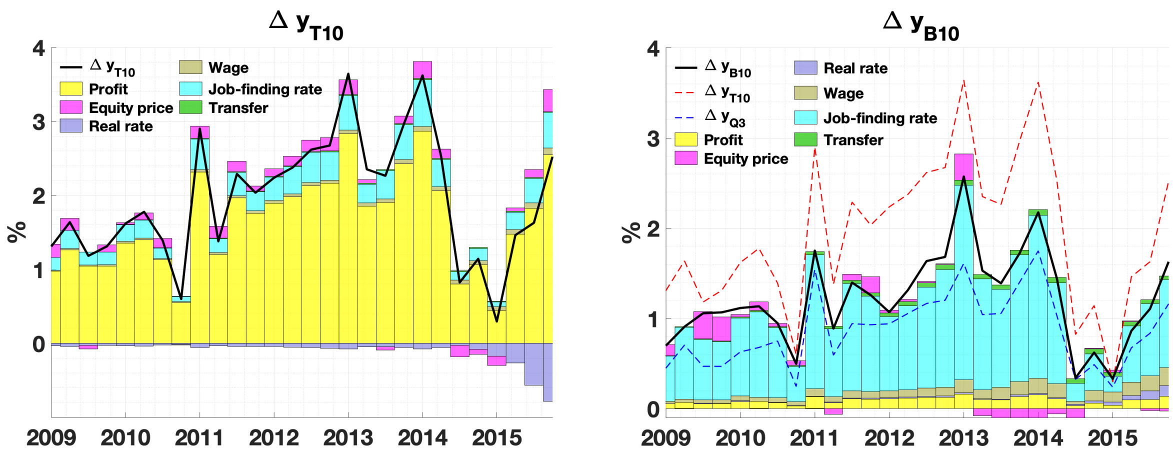
Notes: The left panel shows average changes in the share of unemployed households induced by QE across wealth groups. The right panel shows the evolution of unemployed household shares during the ELB episode, as percentage point difference from the corresponding values in the counterfactual case with 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



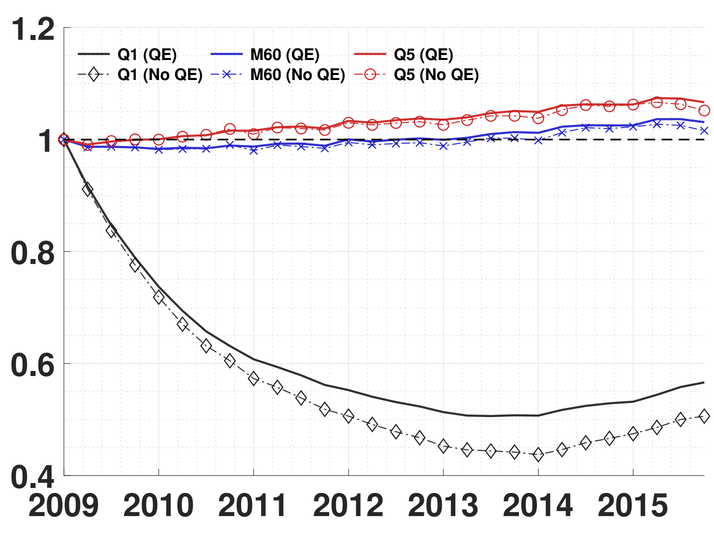
Notes: The figure shows the evolution of the wealth and income Gini indices during the ELB episode, as differnces of the index relative to its 2007 Q4 level. The blue lines with circles show the Gini indices in the baseline case. The dashed red lines show the Gini indices in the counterfactual case.

Figure A15: Top 10% vs Bottom 10%: Income growth decomposition



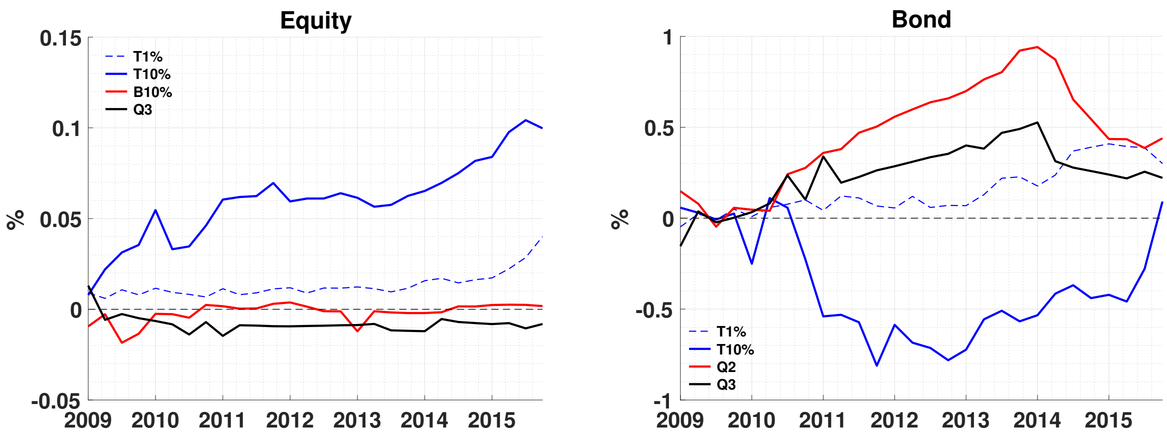
Notes: The figure shows the decomposition of income growth due to QE, for the top 10% and the bottom 10% of the wealth distribution. The black line in each panel shows the growth rate of total income. Each bar shows the contribution of each income component to the total income growth attributable to QE during the ELB episode. In the right panel, the red and the blue dotted line show the income growth rates of the top 10% and the middle quintile, respectively.

Figure A16: Effects of QE on households' wealth



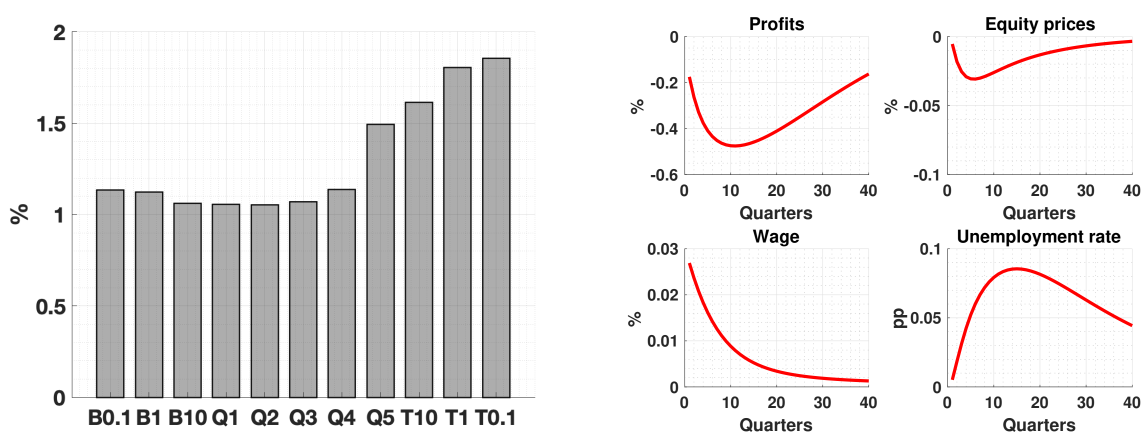
Notes: The black, blue, and red straight lines show the ratio of households' wealth during the ELB episode, relative to their 2009 Q1 level, in the baseline case with QE operation. The black, blue, and red dashed lines with diamond, crosses, and circles show the ratio in the counterfactual case with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

Figure A17: Effects of QE on equity and bond shares



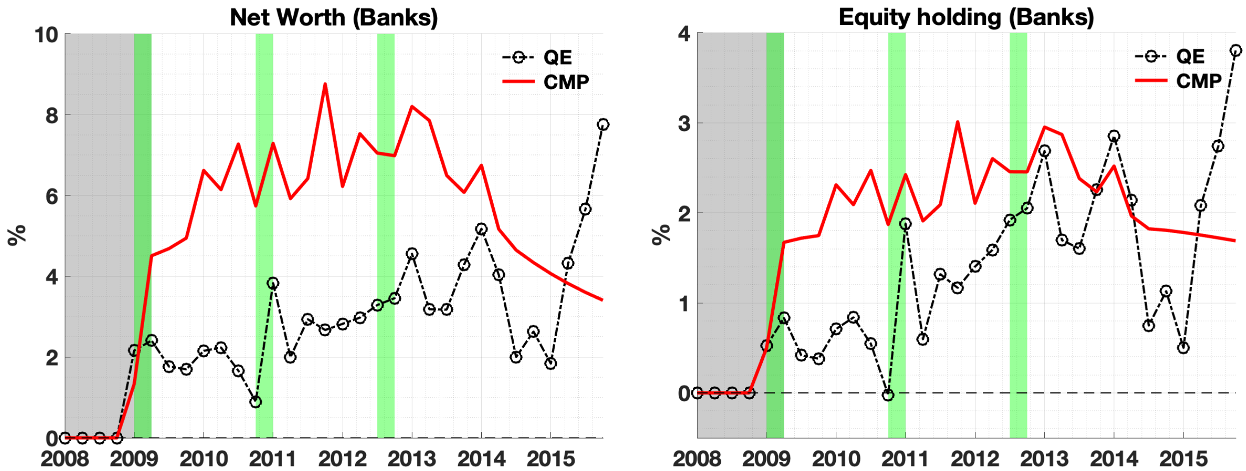
Notes: The figure shows different wealth groups' equity and bond shares during the ELB episode. T1%, T10%, B10%, Q2 and Q3 refers to the top 1% and 10%, the bottom 10%, the second and the middle quintile, respectively. The unit is the difference in the share of equity and bond between the baseline and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

Figure A18: Average consumption gain and households' expectations beyond the sample



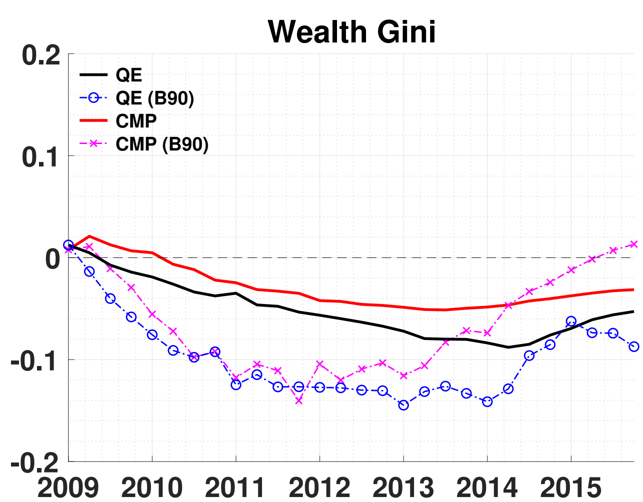
Notes: The left panel shows relative levels of consumption in the baseline case of QE during the ELB episode, relative to the corresponding consumption levels in the counterfactual case of no QE across households' wealth groups. The right panel shows households' expectations on profits, equity prices,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

Figure A19: Effects of monetary policy on banks: QE vs CMP



Notes: The red lines show banks' net worth and equity holdings in the case of CMP, relative to their respective level in the counterfactual case with no policy interventions. The black dotted lines with circles show the corresponding values in the case of QE. QE refers to quantitative easing. CMP refers to conventional monetary policy.

Figure A20: Distributional effects of QE and CMP: the wealth Gini index



Notes: The figure shows the differences in wealth Gini indices between the case of QE or CMP and the counterfactual case of no policies. The black and red solid lines shows the wealth Gini index in the case of QE and CMP, respectively. The pink line with crosses and blue line with circles show the Gini indices among the bottom 90% households.