Shocks, Frictions, and Inequality in US Business Cycles*

Christian Bayer, Benjamin Born, and Ralph Luetticke

November 4, 2022

Abstract

We show how a heterogeneous-agent New-Keynesian (HANK) model with incomplete markets and portfolio choice can be estimated in state space using a Bayesian approach. To render estimation feasible, the structure of the economy can be exploited and the dimensionality of the model automatically reduced based on the Bayesian priors. We apply this approach to analyze how much inequality matters for the business cycle and vice versa. Even when the model is estimated on aggregate data alone and with a set of shocks and frictions designed to match aggregate data, it broadly reproduces observed US inequality dynamics.

---

1 Introduction

A new generation of monetary business cycle models has become popular, featuring heterogeneous agents and incomplete markets (known as HANK models). This new class of models implies new transmission channels of monetary[1] and fiscal[2] policy, as well as new sources of business cycle fluctuations working through household portfolio decisions.[3] Much of this literature so far has focused on specific channels of transmission, shocks, or puzzles. In contrast, the present paper takes a more encompassing approach and shows how a discrete-time HANK model can be estimated using its state-space representation via Bayesian methods.

Following Reiter (2009), we linearize the model, but then leverage two aspects for the estimation: First, we exploit the fact that the economy can be written in a modular way, where the elements reflecting household heterogeneity are free of the estimated parameters. This implies that re-*linearizing* the HANK model after a parameter change is as little numerical burdensome as it is for the representative agent model (RANK). The modularity also allows us to provide a toolbox for the estimation of HANK models where the aggregate part can be readily customized. Second, we develop a novel model reduction approach that naturally emerges from the Bayesian setting and speeds up *solving* the linearized economy drastically. We show that speed and precision of the proposed solution method is comparable to sequence-space approaches suggested by Auclert et al. (2021b) and Boppart et al. (2018).

Concretely, we extend the method of Bayer and Luetticke (2020) with a more flexible and better-informed treatment of the value and distribution functions. This yields an arbitrarily precise approximation of the full model using information from the stationary equilibrium. Our novel Bayesian reduction technique then leverages the fact that we have prior information for model reduction, which allows us to go beyond what the literature has suggested. Specifically, we solve the model once under the prior parameterization and use this solution to derive a factor representation of the heterogeneous-agent part of the model. We show that this factor representation can be used for a very strong model reduction even when parameters change. It reduces the system of difference equations from originally more than 660,000 variables and equations to 3,500 in a first step based on the model's stationary equilibrium and finally, using the factor representation of the model's dynamics under the prior, to under 500. Key here is the same insight that renders sequence-space methods feasible: Households' decisions are driven by a small set of prices and their limited and tractable dynamics.

---

[1]Auclert (2019) analyzes the redistributive effects of monetary policy, Kaplan et al. (2018) show the importance of indirect income effects, and Luetticke (2021) analyzes the portfolio rebalancing channel of monetary policy. McKay et al. (2016) study the effectiveness of forward guidance.

[2]Auclert et al. (2018), Bayer et al. (2022), and Hagedorn et al. (2019) discuss fiscal multipliers, McKay and Reis (2016, 2021) discuss the role of automatic stabilizers.

[3]Bayer et al. (2019) quantify the importance of shocks to idiosyncratic income risk, and Guerrieri and Lorenzoni (2017) look at the effects of shocks to the borrowing limit.

As an application, we study a business cycle model in the spirit of Smets and Wouters (2007), fusing this with the New-Keynesian incomplete markets model of Bayer et al. (2019). This fused model then features capacity utilization, a frictional labor market with sticky wages, and progressive taxation, as well as the battery of shocks that drive business cycle fluctuations in estimated New-Keynesian models: aggregate and investment-specific productivity shocks, wage- and price-markup shocks, monetary- and fiscal-policy shocks, risk-premium shocks. To this battery of standard RANK model business cycle shocks, we add two additional incomplete-market-specific ones: shocks to the progressivity of taxes and shocks to idiosyncratic productivity risk. With this model at hand, we tackle three questions: First, to what extent does the inclusion of incomplete markets change our view of US business cycles? Second, can the model also capture US inequality dynamics? Third, if so, which business cycle shocks and policies are important drivers?

In our model, precautionary motives play an important role for consumption-savings decisions. Since individual income is subject to idiosyncratic risk that cannot be directly insured and borrowing is constrained, households structure their savings decisions and portfolio allocations to optimally self-insure and achieve consumption smoothing. In particular, we assume that households can either hold liquid nominal bonds or invest in illiquid physical capital. Capital is illiquid because its market is segmented and households participate only from time to time. This portfolio-choice component, which gives rise to an endogenous liquidity premium, and the presence of occasional hand-to-mouth consumers lead the HANK model to have rich distributional dynamics in response to aggregate shocks.

To answer the first question whether incomplete markets change our view of US business cycles, we estimate the HANK model using the same set of aggregate shocks and observables as in Smets and Wouters (2007), covering the time period of 1954 to 2019, and compare it to the representative household analogue (RANK). We find that both models tell a similar story about the business cycle, but there are some differences because the incomplete markets structure intimately links aggregate consumption to the distributional consequences of shocks. In particular, technology shocks – at the expense of markup shocks – become more important for consumption growth in HANK because a sizable fraction of households has few assets and a high marginal propensity to consume out of wage income.

These distributional consequences are also important for our second question. Here, we find that the HANK model can simultaneously account for the dynamics of the US business cycle and inequality between 1954 and 2019. Our model translates the business cycle shocks estimated from aggregate data into persistent movements in wealth and income inequality that are broadly consistent with the U-shaped evolution of the shares of wealth and income held by the top 10 percent of US households.[4]

---

[4]We focus on the top 10 shares taken from the World Inequality Database because these measures are

Based on this finding, we then answer our third question, what drives US inequality, by re-estimating the model. We now include the top 10 shares as observables, and allow two additional, and also observed, variables to directly affect the distribution of income: the progressivity of taxes and idiosyncratic income risk. We find that income risk partly replaces risk-premium shocks in explaining aggregate consumption growth. Shocks to the progressivity of taxes have some importance for wealth inequality as they persistently change the net income distribution and self-insurance incentives.

These findings are also reflected in historical decompositions of US inequality. We find that wealth inequality—measured by the share held by the top 10 percent—is largely driven by two factors: shocks to investment technology and shocks to price markups. Shocks to technology have strong effects on asset prices and returns and through them persistently affect the wealth distribution (as empirically shown in Kuhn et al., 2020). Key for this are portfolio differences between the wealthy and the poor. The former hold their wealth in liquid low return assets, the latter in illiquid form. Price markup shocks work through the income distribution and we estimate that the persistent increase in income inequality since the 1980s is linked to higher price markups. However, there is some tension between the high volatility of markups in the model and the low cyclical volatility of the top income share.

To our knowledge, our paper is one of the first to provide an encompassing estimation of shocks and frictions using a HANK model with portfolio choice. Most of the literature on monetary heterogeneous-agent models has used a calibration approach.[5] Auclert et al. (2020), Hagedorn et al. (2018), and Bayer et al. (2022) go beyond calibration but use a limited information approach based on impulse response function matching. Auclert et al. (2021b), go beyond such limited information approaches and propose a technique for estimating heterogeneous agent models that solves the model in sequence space and uses the resulting MA-∞ representation for estimation. In contrast, our proposed technique, building on a state-space representation, connects more directly to established ways of estimating dynamic stochastic general equilibrium models and the vast toolset that has developed around it: for example, obtaining variance decompositions at business cycle frequencies (Uhlig, 2001), using different sampling and filtering techniques (Acharya et al., 2021; Herbst and Schorfheide, 2014, 2015), or dealing with mixed-frequency data.[6]

---

available from the 1950s onward and are most similar across alternative data sources such as the Survey of Consumer Finances; see Kopczuk (2015) and Bricker et al. (2016).

[5]See, for example, Auclert et al. (2018); Ahn et al. (2018); Bayer et al. (2019); Broer et al. (2019); Challe and Ragot (2015); Den Haan et al. (2017); Ferriere and Navarro (2018); Gornemann et al. (2012); Guerrieri and Lorenzoni (2017); McKay et al. (2016); McKay and Reis (2016); Ravn and Sterk (2017); Sterk and Tenreyro (2018); Wong (2019).

[6]Of course, sequence-space solutions can be translated into a quasi state-space representation and Auclert et al. (2021b) show how this can be done. The impulse-response representation, however, remains the natural form of a sequence-space solution.

The inclusion of distributional data in the estimation is novel and potentially informative for identifying the sources of business cycle variations. Our paper is related to Chang et al. (2021) in the sense that it estimates a state-space model of both distributional (cross-sectional) data and aggregates. Chang et al. (2021) find that, in an SVAR sense, shocks to the cross-sectional distribution of income only have a mild impact on aggregate time series. Our finding of structural estimates being relatively robust to the inclusion or exclusion of cross-sectional information resembles their results.[7] The estimated muted importance of cross-sectional shocks is also in line with the finding of Berger et al. (2022) who use a business-cycle accounting approach. Bilbiie et al. (2022) estimate a tractable heterogeneous agent model in state-space form (two types of agents with stochastic transitions between types) using a full-information approach and data on the cross-sectional dispersion of labor earnings and income. They find amplification of aggregate shocks with heterogeneity.

Our findings add some insights to the literature on the drivers of inequality.[8] Kaymak and Poschke (2016) and Hubmer et al. (2020) use quantitative models to study permanent changes in the US tax and transfer system and the variance of income. In terms of methods, these papers solve for steady-state transitions of calibrated models, while we estimate our model on US macro and micro time-series data. They find that tax and transfer changes can explain a significant part of the recent increase in wealth inequality. Ours is the first paper to quantify the distributional consequences of all standard business cycle shocks and estimate their importance in explaining US inequality. Alongside business cycle shocks, our model features changes in the US tax system and income risk and allows us to compare their relative importance for the evolution of income and wealth inequality. We find business cycle shocks, through their effect on asset prices and returns, to be important for wealth inequality.

The remainder of this paper is organized as follows: Section 2 describes our model economy, its sources of fluctuations and frictions. Section 3 provides details on the numerical solution method and estimation technique. Section 4 presents our model variants, the parameters that we calibrate to match steady-state targets, prior and posterior distributions for the remaining parameters that we estimate, and an assessment of our solution approach based on the estimated posterior distribution. It also gives an overview of the data we employ in our estimation. Section 5 discusses the estimated shocks and frictions driving the US business cycle and inequality dynamics. Section 6 concludes. An Appendix follows.

---

[7]Our approach is different to and simpler than the method suggested by Liu and Plagborg-Møller (2022), which includes full cross-sectional information in the estimation of a heterogeneous-agent DSGE model. By contrast, we only use the model to fit certain generalized cross-sectional moments.

[8]There is a growing literature on inequality dynamics: on the theory side, e.g., Gabaix et al. (2016) and on the empirical side, e.g., Heathcote et al. (2010), Piketty and Saez (2003) or Saez and Zucman (2016).

2   Model

We model an economy composed of a firm sector, a household sector, and a government sector. Of these three sectors only the household sector deviates from the standard New Keynesian DSGE model structure as in Smets and Wouters (2007) or Christiano et al. (2005). In detail, the firm sector comprises (a) final goods producers who bundle the intermediate goods; (b) intermediate goods producers who rent out labor services and capital from perfectly competitive markets, but face monopolistic competition in the goods market as they produce differentiated goods and set prices; (c) producers of capital goods that turn final goods into capital subject to adjustment costs; (d) labor packers that produce labor services combining differentiated labor from unions that differentiate raw labor rented out from households. Price setting for the intermediate goods as well as wage setting by unions is subject to a pricing friction à la Calvo (1983).

Households earn income from supplying (raw) labor and capital and from owning the firm sector, absorbing all its rents that stem from the market power of unions and intermediate goods producers, and decreasing returns to scale in capital goods production. They face idiosyncratic income risk against which they can self-insure by trading liquid and illiquid assets giving rise to endogenous heterogeneity. We also consider a representative-agent variant with full insurance.

The government sector runs both a fiscal authority and a monetary authority. The fiscal authority levies progressive taxes on labor incomes and profits, issues government bonds, and adjusts expenditures to stabilize debt in the long run and aggregate demand in the short run. The monetary authority sets the nominal interest rate on government bonds according to a Taylor rule.

2.1   Firm sector

Since the firm sector involves dynamic decisions, we need to make an assumption about the discount factor used in these decisions. With household heterogeneity, stochastic discount factors across households might differ. For this reason, we make the simplifying assumption that the firm sector is run by managers that are risk neutral, have no asset market access, but have the same time preferences as households.[9] Managers are a mass-zero group in the economy so that their consumption does not show up in any resource constraint and as a result all profits of the firm sector go to households.

---

[9]Since we solve the model by a first-order perturbation in aggregate shocks, fluctuations in stochastic discount factors are irrelevant.

2.1.1 Final goods producers

Final goods producers bundle varieties j of differentiated goods according to the Dixit-Stiglitz aggregator

$$Y_t = \left(\int y_{jt}^{\frac{\eta_t-1}{\eta_t}} dj \right)^{\frac{\eta_t}{\eta_t-1}}, \tag{1}$$

with elasticity of substitution η_t. Each of these differentiated goods is offered at price p_{jt}, so that the aggregate price level is given by $P_t = \left(\int p_{jt}^{1-\eta_t} dj \right)^{\frac{1}{1-\eta_t}}$ and the demand for each of the varieties is

$$y_{jt} = \left(\frac{p_{jt}}{P_t} \right)^{-\eta_t} Y_t. \tag{2}$$

2.1.2 Intermediate goods producers

Intermediate goods are produced with a constant returns to scale production function

$$Y_{jt} = Z_t N_{jt}^{\alpha} (u_{jt} K_{jt})^{(1-\alpha)}, \tag{3}$$

where α is the labor share in production, Z_t is total factor productivity that follows an autoregressive process in logs, N_{jt} is the labor bundle firm j hires at time t, and $u_{jt} K_{jt}$ are capital services taking into account utilization u_{jt}, i.e., the intensity with which the capital stock K_{jt} is used. An intensity higher than normal results in increased depreciation of capital according to $\delta(u_{jt}) = \delta_0 + \delta_1(u_{jt}-1) + \delta_2/2(u_{jt}-1)^2$, which, assuming $\delta_1, \delta_2 > 0$, is an increasing and convex function of utilization. Without loss of generality, capital utilization in steady state is normalized to 1, so that δ_0 denotes the steady-state depreciation rate of capital goods.

Given demand, the producer minimizes costs, $w_t^F N_t - [r_t + q_t \delta(u_{jt})] K_t$, where r_t and q_t are the rental rate and the (producer) price of capital goods, respectively, and w_t^F is the real wage the firm faces. Factor markets are perfectly competitive. Hence, the first-order conditions for labor and effective capital read

$$w_t^F = \alpha mc_{jt} Z_t \left(\frac{u_{jt} K_{jt}}{N_{jt}} \right)^{1-\alpha} \tag{4}$$

$$\text{and} \qquad r_t + q_t \delta(u_{jt}) = u_{jt}(1-\alpha) mc_{jt} Z_t \left(\frac{N_{jt}}{u_{jt} K_{jt}} \right)^{\alpha}, \tag{5}$$

where mc_{jt} is the marginal cost of firm j. The optimal utilization is given by

$$q_t \left[\delta_1 + \delta_2(u_{jt}-1) \right] = (1-\alpha) mc_{jt} Z_t \left(\frac{N_{jt}}{u_{jt} K_{jt}} \right)^{\alpha}. \tag{6}$$

Combining the three first-order conditions, it is easy to see that given the constant returns to scale production function, marginal costs are constant across producers $mc_{jt} = mc_t$.

We assume that intermediate goods producers face price adjustment frictions à la Calvo (1983); and the firms' managers maximize the present value of real profits subject to this price adjustment friction and the demand curve (2). They hence maximize

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \lambda_Y^t (1 - \tau_t^L) Y_t^{1-\tau_t^P} \left\{ \left(\frac{p_{jt}\bar{\pi}_Y^t}{P_t} - mc_t \right) \left(\frac{p_{jt}\bar{\pi}^t}{P_t} \right)^{-\eta_t} \right\}^{1-\tau_t^P}, \tag{7}$$

with a time-constant discount factor β. Prices are indexed to the steady-state inflation rate $\bar{\pi}$ and can be discretionally adjusted price with probability $1 - \lambda_Y$. The parameters τ_t^P and τ_t^L characterize the progressivity and level of the tax schedule, which we discuss in more detail, when describing the household sector.

The corresponding first-order condition for price setting implies a Phillips curve

$$\log \left(\frac{\pi_t}{\bar{\pi}} \right) = \beta \mathbb{E}_t \log \left(\frac{\pi_{t+1}}{\bar{\pi}} \right) + \kappa_Y \left(mc_t - \frac{1}{\mu_t^Y} \right), \tag{8}$$

where we dropped all terms irrelevant for a first-order approximation and defined $\kappa_Y = \frac{(1-\lambda_Y)(1-\lambda_Y\beta)}{\lambda_Y}$. Here, π_t is the gross inflation rate of final goods, $\pi_t := \frac{P_t}{P_{t-1}}$, $mc_t := \frac{MC_t}{P_t}$ is the real marginal costs, and $\mu_t^Y = \frac{\eta_t}{\eta_t - 1}$ is the target markup. This target fluctuates in response to markup shocks, $\epsilon_t^{\mu Y}$, and follows a log AR(1) process.

2.1.3 Capital goods producers

Capital goods producers take the relative price of capital goods, q_t, as given in deciding about their output, i.e., they maximize

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t I_t \left\{ \Psi_t q_t \left[1 - \frac{\phi}{2} \left(\log \frac{I_t}{I_{t-1}} \right)^2 \right] - 1 \right\}, \tag{9}$$

where ϕ controls the strength of the quadratic investment adjustment costs and Ψ_t governs the marginal efficiency of investment à la Justiniano et al. (2011), which follows an AR(1) process in logs and is subject to shocks ϵ_t^{Ψ}.[10]

---

[10]This shock has to be distinguished from a shock to the relative price of investment, which has been shown in the literature (Justiniano et al., 2011; Schmitt-Grohé and Uribe, 2012) to not be an important driver of business cycles as soon as one includes the relative price of investment as an observable. We therefore focus on the MEI shock.

Optimality of the capital goods production requires (again dropping all terms irrelevant up to first order)

$$\Psi_t q_t \left[1 - \phi \log \frac{I_t}{I_{t-1}}\right] = 1 - \beta \mathbb{E}_t \left[\Psi_{t+1} q_{t+1} \phi \log \left(\frac{I_{t+1}}{I_t}\right)\right], \tag{10}$$

and each capital goods producer will adjust its production until (10) is fulfilled.

Since all capital goods producers are symmetric, we obtain as the law of motion for aggregate capital:

$$K_t - (1 - \delta(u_t)) K_{t-1} = \Psi_t \left[1 - \frac{\phi}{2} \left(\log \frac{I_t}{I_{t-1}}\right)^2\right] I_t . \tag{11}$$

The functional form assumption implies that investment adjustment costs are minimized and equal to 0 in the steady state.

2.1.4 Labor packers and unions

Workers sell their labor services to a mass-one continuum of unions indexed by j, each of whom offers a different variety of labor to labor packers who then provide labor services to intermediate goods producers. Labor packers produce final labor services according to the production function

$$N_t = \left(\int \hat{n}_{jt}^{\frac{\zeta_t - 1}{\zeta_t}} dj\right)^{\frac{\zeta_t}{\zeta_t - 1}}, \tag{12}$$

out of labor varieties $\hat{n}_{jt}$ with elasticity of substitution ζ_t. Cost minimization by labor packers implies that each variety of labor, each union j, faces a downward-sloping demand curve

$$\hat{n}_{jt} = \left(\frac{W_{jt}}{W_t^F}\right)^{-\zeta_t} N_t , \tag{13}$$

where W_{jt} is the *nominal* wage set by union j and W_t^F is the *nominal* wage at which labor packers sell labor services to intermediate goods producers.

Since unions have market power, they pay the households a wage lower than the price at which they sell labor to labor packers. Given the nominal wage W_t at which they buy labor from households and given the nominal wage index W_t^F, unions seek to maximize their discounted stream of profits. However, they face a Calvo (1983)-type adjustment friction with indexation, where λ_w is the probability to keep wages constant. They therefore maximize

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \lambda_w^t \frac{W_t^F}{P_t} N_t \left\{\left(\frac{W_{jt}\bar{\pi}_W^t}{W_t^F} - \frac{W_t}{W_t^F}\right) \left(\frac{W_{jt}\bar{\pi}_W^t}{W_t^F}\right)^{-\zeta_t}\right\}, \tag{14}$$

by setting W_{jt} in period t and keeping it constant except for indexation to $\bar{\pi}_W$, the steady-state wage inflation rate.

Since all unions are symmetric, we focus on a symmetric equilibrium and obtain the linearized wage Phillips curve from the corresponding first-order condition as follows, leaving out all terms irrelevant at a first-order approximation around the stationary equilibrium

$$\log\left(\frac{\pi_t^W}{\bar{\pi}_W}\right) = \beta\mathbb{E}_t\log\left(\frac{\pi_{t+1}^W}{\bar{\pi}_W}\right) + \kappa_w\left(mc_t^w - \frac{1}{\mu_t^W}\right), \tag{15}$$

with $\pi_t^W := \frac{W_t^F}{W_{t-1}^F} = \frac{w_t^F}{w_{t-1}^F}\pi_t^Y$ being wage inflation, w_t and w_t^F being the respective *real* wages for households and firms, $mc_t^w = \frac{w_t}{w_t^F}$ is the actual and $\frac{1}{\mu_t^W} = \frac{\zeta_t-1}{\zeta_t}$ being the target mark-down of wages the unions pay to households, W_t, relative to the wages charged to firms, W_t^F and $\kappa_w = \frac{(1-\lambda_w)(1-\lambda_w\beta)}{\lambda_w}$. This target fluctuates in response to markup shocks, $\epsilon_t^{\mu W}$, and follows a log AR(1) process.

2.2 Households

There is a continuum of ex-ante identical households of measure one, indexed by i. Households are infinitely lived, have time-separable preferences with time-discount factor β, and derive felicity from consumption c_{it} and leisure. They obtain income from supplying labor, n_{it}, from renting out capital, k_{it}, and from earning interest on bonds, b_{it}. What is more, they receive profits of firms, $\Pi_t^Y = (1 - mc_t)Y_t$, and unions, $\Pi_t^U = (w_t^F - w_t)N_t$. Households pay taxes on labor and profit income. Our baseline model features household heterogeneity. Households differ in their productivity and in whether the obtain profit income. They face incomplete markets in this baseline, and capital as an asset is illiquid while bonds are liquid. For comparison, we also consider a representative-agent variant.

2.2.1 Preferences

With respect to leisure and consumption, households have Greenwood et al. (1988) (GHH) preferences and maximize the discounted sum of felicity

$$\mathbb{E}_0 \max_{\{c_{it},n_{it}\}} \sum_{t=0}^{\infty} \beta^t u\left[c_{it} - G(h_{it}, n_{it})\right]. \tag{16}$$

The maximization is subject to the budget constraints described further below. The felicity function u exhibits a constant relative risk aversion (CRRA) with risk aversion parameter ξ,

$$u(x_{it}) = \frac{x_{it}^{1-\xi} - 1}{1 - \xi}, \tag{17}$$

9

where $x_{it} = c_{it} - G(h_{it}, n_{it})$ is household i's composite demand for goods consumption c_{it} and leisure and G measures the disutility from work. While n_{it} denotes a household's labor supply, h_{it} is the household's labor productivity.[11]

Assuming a (progressive) income-tax schedule (which we borrow from Benabou, 2002; Heathcote et al., 2017), a household's net labor income, y_{it}, is given by

$$y_{it} = (1 - \tau_t^L)(w_t h_{it} n_{it})^{1-\tau_t^P} \ , \tag{18}$$

where w_t is the aggregate real wage rate and τ_t^L and τ_t^P determine the level and progressivity of the tax code. Given net labor income, the first-order condition for labor supply is

$$\frac{\partial G(h_{it}, n_{it})}{\partial n_{it}} = (1 - \tau_t^P)(1 - \tau_t^L)(w_t h_{it})^{1-\tau_t^P} n_{it}^{-\tau_t^P} = (1 - \tau_t^P)\frac{y_{it}}{n_{it}} \ . \tag{19}$$

Assuming that G has a constant elasticity w.r.t. n, $\frac{\partial G(h_{it}, n_{it})}{\partial n_{it}} = (1 + \gamma)\frac{G(h_{it}, n_{it})}{n_{it}}$ with $\gamma > 0$, we can simplify the expression for the composite consumption good, x_{it}, making use of this first-order condition (19), and substitute $G(h_{it}, n_{it})$ out of the individual planning problem

$$x_{it} = c_{it} - G(h_{it}, n_{it}) = c_{it} - \frac{1 - \tau_t^P}{1 + \gamma} y_{it} \ . \tag{20}$$

When the Frisch elasticity of labor supply is constant and the tax schedule has the form (18), the disutility of labor is always a fraction of labor income and constant across households. Therefore, in both the household's budget constraint and felicity function, only after-tax income enters and neither hours worked nor productivity appear separately.

---

[11] The assumption of GHH preferences is mainly motivated by the fact that many estimated DSGE models of business cycles find small aggregate wealth effects in the labor supply; see, e.g., Schmitt-Grohé and Uribe (2012); Born and Pfeifer (2014). Unfortunately, it is not feasible to estimate the flexible form of preference of Jaimovich and Rebelo (2009), which also encompasses King et al. (1988) (KPR) preferences. This would require solving the stationary equilibrium in every likelihood evaluation, which is substantially more time consuming than solving for the dynamics around this equilibrium. However, we estimate a version with KPR preferences; see Section 5.3 and also Appendix B. According to the marginal data density, the data clearly prefer the GHH specification over the KPR specification. What is more, the KPR version of the HANK model has more difficulty matching business cycle and inequality dynamics simultaneously. The GHH assumption has been criticized by Auclert et al. (2021a) on the basis of producing "too high" multipliers. In a companion paper (Bayer et al., 2022), we show that our model produces multipliers of reasonable size both in the short and in the long run. The reason for this lies in the combination of model elements only briefly discussed or even absent in the stylized Auclert et al. (2021a) economy: sticky wages, distortionary taxes, capacity utilization, and a Taylor rule. Capacity utilization allows for output adjustment without adjusting hours; additional wage stickiness translates increasing labor demand into higher wage markups instead of hours and consumption; distortionary taxes absorb an additional fraction of income; and the Taylor rule translates the fiscal shock into to a real interest rate increase. The back-of-the envelope calculation of the multiplier based on formula (15) in Auclert et al. (2021a), counter-factually assuming fixed real rates and ignoring capacity utilization, would be: $(1 - (1-\tau)(\eta-1)/\eta(\zeta-1)/\zeta)^{-1} \approx 2.5$. The true multiplier in the model with capacity utilization and interest rate response is, in line with the data, much smaller.

Aggregate effective labor supply depends on the distribution of h_{it} and tax progressivity. Without loss of generality, we assume $G(h_{it}, n_{it}) = h_{it}^{1-\bar{\tau}^P} \frac{n_{it}^{1+\gamma}}{1+\gamma}$, where $\bar{\tau}^P$ is the stationary equilibrium level of progressivity of the tax code. This functional form simplifies the household problem as h_{it} drops out from the first-order condition as long as tax progressivity is constant. Then, all households supply the same number of hours $n_{it} = N(w_t)$. Total effective labor input, $\int n_{it} h_{it} di$, is then hence also equal to $N(w_t)$ because we normalize $\int h_{it} di = 1$. Individual after tax labor income is then

$$y_{it} = (1 - \tau_t^L)(w_t h_{it} n_{it})^{1-\bar{\tau}^P} = (1-\tau_t^L)^{\frac{1+\gamma}{\gamma+\bar{\tau}^P}} (1-\bar{\tau}^P)^{\frac{1-\bar{\tau}^P}{\gamma+\bar{\tau}^P}} w_t^{\frac{1+\gamma}{\gamma+\bar{\tau}^P}(1-\bar{\tau}^P)} h_{it}^{1-\bar{\tau}^P} \ . \qquad (21)$$

In an extension, we allow for tax progressivity to vary over time because it can directly affect the distribution of net incomes and thereby wealth.[12]

2.2.2 Baseline: heterogeneous households and incomplete markets

In our baseline, the household sector is subdivided into two types of agents: workers and entrepreneurs. The transition between both types is stochastic. On top, workers face idiosyncratic labor productivity risk. Both, workers and entrepreneurs, rent out physical capital, but only workers supply labor. Entrepreneurs do not work, but earn all pure rents in our economy except for the rents of unions which are equally distributed across workers. All

---

[12]In this case, the parameter τ_t^P governing the progressivity of the tax schedule evolves according to

$$\frac{\tau_t^P}{\bar{\tau}^P} = \left(\frac{\tau_{t-1}^P}{\bar{\tau}^P}\right)^{\rho_P} \epsilon_t^P \ , \qquad (33a)$$

where ϵ_t^P are shocks to tax progressivity. When tax progressivity does not coincide with its stationary equilibrium value, individual hours worked differ across agents and are given by

$$n_{it} = \left[(1-\tau_t^P)(1-\tau_t^L)\right]^{\frac{1}{\gamma+\tau_t^P}} h_{it}^{\frac{\bar{\tau}^P-\tau_t^P}{\gamma+\tau_t^P}} w_t^{\frac{1-\tau_t^P}{\gamma+\tau_t^P}} \ , \qquad (19a)$$

such that aggregate effective hours are given by

$$N_t = \int n_{it} h_{it} = \left[(1-\tau_t^P)(1-\tau_t^L)\right]^{\frac{1}{\gamma+\tau_t^P}} w_t^{\frac{1-\tau_t^P}{\gamma+\tau_t^P}} \underbrace{\int h_{it}^{\frac{\gamma+\bar{\tau}^P}{\gamma+\tau_t^P}}}_{:=H_t} \ . \qquad (19b)$$

Here H_t measures how the tax progressivity influences the (hours-weighted) average labor productivity. Scaling of the disutility of labor by $h_{it}^{1-\bar{\tau}^P}$ is thus a normalization of H_t to one in the stationary equilibrium. Household after-tax labor income, plugging in the optimal supply of hours, is then

$$y_{it} = (1-\tau_t^L)(w_t h_{it} n_{it})^{1-\tau_t^P} = (1-\tau_t^L)^{\frac{1+\gamma}{\gamma+\tau_t^P}} (1-\tau_t^P)^{\frac{1-\tau_t^P}{\gamma+\tau_t^P}} w_t^{\frac{1+\gamma}{\gamma+\tau_t^P}(1-\tau_t^P)} h_{it}^{\frac{\gamma+\bar{\tau}^P}{\gamma+\tau_t^P}(1-\tau_t^P)} \ . \qquad (22)$$

11

households self-insure against the income risks they face by saving in a liquid nominal asset (bonds) and a less liquid asset (capital). Trading these illiquid assets is subject to random participation in the capital market.

We assume that productivity evolves according to a log-AR(1) process and a fixed probability of transition between the worker and the entrepreneur state:

$$
\tilde{h}_{it} = \begin{cases} \exp\left(\rho_h \log \tilde{h}_{it-1} + \epsilon_{it}^h\right) & \text{with probability } 1 - \zeta \text{ if } h_{it-1} \neq 0, \\ 1 & \text{with probability } \iota \text{ if } h_{it-1} = 0, \\ 0 & \text{else,} \end{cases} \tag{23}
$$

with individual productivity $h_{it} = \frac{\tilde{h}_{it}}{\int \tilde{h}_{it} di}$ such that $\tilde{h}_{it}$ is scaled by its cross-sectional average, $\int \tilde{h}_{it} di$, to make sure that average worker productivity is constant. The shocks ϵ_{it}^h to productivity are normally distributed with variance $\bar{\sigma}_h^2$.[13] With probability ζ households become entrepreneurs ($h = 0$). With probability ι an entrepreneur returns to the labor force with median productivity. Besides their labor income, workers receive a share in union rents, Π_t^U, which are distributed lump sum, leading to labor-income compression. For tractability, we assume union profits to be taxed at a fixed rate independent of the recipient's labor income.

For the distribution of firm profits (aside union profits), we assume that they primarily go to entrepreneurs. However, entrepreneurs as a group can sell claims to a fraction ω^Π of their profits as shares. These claims have stochastic maturity and are liquid. This stochastic maturity ensure finite prices for profit claims even at zero interest rates of liquid assets. Each period ι^Π claims mature. When a claim matures, it loses value and is replaced by a new issuance by the entrepreneurs. We assume a unit mass of profit shares which then trade at price q_t^Π. The entrepreneurs then receive in each period the sum of the profits they have not sold plus the value of the new shares they sell: $\Pi_t^E = (1 - \omega^\Pi)\Pi_t^F + \iota^\Pi q_t^\Pi$.[14]

This modeling strategy allows us to match the income and wealth distribution following the idea by Castaneda et al. (1998) while limiting the impact of profits on investment behavior and asset markets.

---

[13]In our baseline, we treat the variance as time fixed. We consider an extension where income risk follows a log-AR(1) process with endogenous feedback to aggregate output growth:

$$
\begin{aligned}
\sigma_{h,t}^2 &= \bar{\sigma}_h^2 \exp \widehat{s}_t, \\
\widehat{s}_{t+1} &= \rho_s \widehat{s}_t + \Sigma_Y \frac{Y_{t+1}}{Y_t} + \epsilon_t^\sigma \,,
\end{aligned} \tag{23a}
$$

i.e., at time t households observe a change in the variance of shocks that drive the next period's productivity.

[14]Boar and Midrigan (2019) use a similar structure, where entrepreneurs retain a fraction of firm profits and thus the size of markups has an impact on inequality in the economy.

Given incomes, households optimize intertemporally subject to their budget constraint

$$c_{it} + b_{it+1} + q_t k_{it+1} = b_{it} \frac{R_{it}}{\pi_t} + (q_t + r_t)k_{it} + y_{it} \tag{24}$$
$$+ \mathbb{I}_{h_{it} \neq 0}(1 - \tau_t)\Pi_t^U + \mathbb{I}_{h_{it}=0}(1 - \tau_t^L)(\Pi_t^E)^{1-\bar{\tau}^P}, \qquad k_{it+1} \geq 0, \quad b_{it+1} \geq \underline{B},$$

where Π_t^U is union profits taxed at the average tax rate τ_t, Π_t^E is profit payouts to entrepreneurs, b_{it} is real liquid assets, k_{it} is the amount of illiquid assets, q_t is the price of these assets, r_t is their dividend, $\pi_t = \frac{P_t}{P_{t-1}}$ is realized inflation, and R_{it} is the nominal interest rate on liquid assets, which depends on whether the household is a borrower or lender, the efficiency of intermediation, the returns on profit shares, and the central bank's interest rate R_t^b, which is set one period before. All households that do not participate in the capital market ($k_{it+1} = k_{it}$) still obtain dividends and can adjust their liquid asset holdings. Depreciated capital has to be replaced for maintenance, such that the dividend, r_t, is the net return on capital. Holdings of bonds have to be above an exogenous debt limit $\underline{B}$, and holdings of capital have to be non-negative.

Substituting the expression $c_{it} = x_{it} + \frac{1-\bar{\tau}^P}{1+\gamma}y_{it}$ for consumption, we obtain the budget constraint for the composite leisure-consumption good,

$$x_{it} + b_{it+1} + q_t k_{it+1} = b_{it} \frac{R_{it}}{\pi_t} + (q_t + r_t)k_{it} + \frac{\bar{\tau}^P + \gamma}{1+\gamma}y_{it} \tag{25}$$
$$+ \mathbb{I}_{h_{it} \neq 0}(1 - \tau_t)\Pi_t^U + \mathbb{I}_{h_{it}=0}(1 - \tau_t^L)(\Pi_t^E)^{1-\bar{\tau}^P}, \qquad k_{it+1} \geq 0, \quad b_{it+1} \geq \underline{B}.$$

Households make their savings and portfolio choice between liquid bonds and illiquid capital in light of a capital market friction that renders capital illiquid because participation in the capital market is random and i.i.d. in the sense that only a fraction, λ, of households are selected to be able to adjust their capital holdings in a given period. Ex-post returns R_{it} on the liquid asset are given by the average return of the liquid asset portfolio, composed of government bonds B_t and profit shares with a value of q_t^Π, i.e.,

$$R_{it} = \begin{cases} A_t \frac{R_t^b B_t + \pi_t[(1-\iota^\Pi)q_t^\Pi + \omega^\Pi \Pi_t^F]}{B_t + q_{t-1}^\Pi} & \text{if } b_{it} \geq 0 \\ A_t \frac{R_t^b B_t + \pi_t[(1-\iota^\Pi)q_t^\Pi + \omega^\Pi \Pi_t^F]}{B_t + q_{t-1}^\Pi} + \overline{R} & \text{if } b_{it} < 0 \end{cases}. \tag{26}$$

The shifter A_t reflects a "risk-premium shock" (as in, for example, Smets and Wouters, 2007) and is technically modeled as an intermediation efficiency here. The first part of the sum in the numerator is the interest payments on government bonds issued and bought in the previous period, the second part is the returns from selling the non-matured profit claims and the share of profits that is paid out to shareholders. The denominator is the sum of the

value of bonds and profit shares bought in the previous period.

Since a household's saving decision—(b_a', k') for the case of adjustment and (b_n', k) for non-adjustment—will be some non-linear function of that household's wealth and productivity, inflation and all other prices will be functions of the joint distribution, Θ_t, of (b, k, h) in t. This makes Θ a state variable of the household's planning problem and this distribution evolves as a result of the economy's reaction to aggregate shocks. For simplicity, we summarize all effects of aggregate state variables, including the distribution of wealth and income, by writing the dynamic planning problem with time-dependent continuation values.

This leaves us with three functions that characterize the household's problem: value function V^a for the case where the household adjusts its capital holdings, the function V^n for the case in which it does not adjust, and the expected continuation value, $\mathbb{W}$, over both,

$$
\begin{aligned}
V_t^a(b, k, h) =& \max_{b_a', k'} u[x(b, b_a', k, k', h)] + \beta \mathbb{E}_t \mathbb{W}_{t+1}(b_a', k', h') \ , \\
V_t^n(b, k, h) =& \max_{b_n'} u[x(b, b_n', k, k, h)] + \beta \mathbb{E}_t \mathbb{W}_{t+1}(b_n', k, h') \ , \\
\mathbb{W}_{t+1}(b', k', h') =& \lambda V_{t+1}^a(b', k', h') + (1 - \lambda) V_{t+1}^n(b', k', h') \ .
\end{aligned}
\tag{27}
$$

Expectations about the continuation value are taken with respect to all stochastic processes conditional on the current states. Maximization is subject to the corresponding budget constraint. The distribution Θ_t then evolves according to

$$
\begin{aligned}
\Theta_{t+1}(b', k', h') =& \lambda \int_{b' = b_{a,t}^*(b,k,h), k' = k_t^*(b,k,h)} \Phi(h, h') d\Theta_t(b, k, h) \\
& + (1 - \lambda) \int_{b' = b_{n,t}^*(b,k,h), k' = k} \Phi(h, h') d\Theta_t(b, k, h) \ ,
\end{aligned}
\tag{28}
$$

where $\Phi(\cdot)$ is the transition probability for h and $b_{a/n,t}^*$ and k_t^* are the time-t optimal policies.

Importantly, following Reiter (2009), one can view the discretized version of (27) and (28) as a set of equations that pins down the dynamics of the value functions and optimal policy for each $b \times k \times h$ node as well as the transition of the mass of households at each of the nodes.

2.2.3 Model variant with a representative household and complete markets

With complete markets, when all households are homogeneous with equal and constant labor productivity $h_i = 1$ and equally obtain all profit incomes, the planning problem is described by the consumption Euler equation for bonds instead of the above mentioned set of equations.

For an optimal consumption-savings policy,

$$u_c(x_t) = \beta \mathbb{E}_t \frac{A_t R_t^b}{\pi_{t+1}} u_c(x_{t+1}) \qquad (29)$$

needs to hold, replacing (27). Again, $x_t = c_t - G(n_t)$ is the composite consumption-leisure good.

The law of motion for the distribution (28) is replaced by the wealth accumulation equation given by the budget constraint

$$q_t K_{t+1} + B_{t+1} = \frac{R_t^b}{\pi_t} B_t + (q_t + r_t) K_t + (1 - \tau_t) \left[\frac{\tau_t^P + \gamma}{1+\gamma} (w_t N_t)^{(1-\tau_t^P)} + \Pi_t^U + \Pi_t^F \right] - x_t , \quad (30)$$

and the consumption Euler equation for capital

$$u_c(x_t) = \beta \mathbb{E}_t \frac{q_{t+1} + r_{t+1}}{q_t} u_c(x_{t+1}), \qquad (31)$$

which then yields the optimal portfolio combination of K and B given return expectations.

2.3 Government

This leaves us with the government sector. The government operates a monetary and a fiscal authority. The monetary authority controls the nominal interest rate on liquid assets, while the fiscal authority issues government bonds to finance deficits, chooses both the average tax rate in the economy and the tax progressivity, and makes expenditures.

We assume that monetary policy sets the nominal interest rate following a Taylor-type (1993) rule with interest rate smoothing:

$$\frac{R_{t+1}^b}{\bar{R}^b} = \left(\frac{R_t^b}{\bar{R}^b} \right)^{\rho_R} \left(\frac{\pi_t}{\bar{\pi}} \right)^{(1-\rho_R)\theta_\pi} \left(\frac{Y_t}{Y_{t-1}} \right)^{(1-\rho_R)\theta_Y} \epsilon_t^R . \qquad (32)$$

The coefficient $\bar{R}^b \geq 0$ determines the nominal interest rate in the steady state. The coefficients $\theta_\pi, \theta_Y \geq 0$ govern the extent to which the central bank attempts to stabilize inflation and output growth, $\frac{Y_t}{Y_{t-1}}$. The parameter $\rho_R \geq 0$ captures interest rate smoothing.

We assume that government debt evolves according to the rule (c.f. Woodford, 1995)

$$\frac{B_{t+1}}{B_t} = \left(\frac{B_t}{\bar{B}} \right)^{-\gamma_B} \left(\frac{\pi_t}{\bar{\pi}} \right)^{\gamma_\pi} \left(\frac{Y_t}{Y_{t-1}} \right)^{\gamma_Y} D_t , \qquad D_t = D_{t-1}^{\rho_D} \epsilon_t^D , \qquad (33)$$

where D_t is a persistent shock to the government's structural deficit. Besides issuing bonds, the government uses tax revenues T_t, defined below, to finance government consumption, G_t,

and interest on debt. The parameters γ_B, γ_π, and γ_Y measure, respectively, how the deficit reacts to outstanding debt, inflation, and the output gap.

The government sets the average tax rate in the economy according to a similar rule

$$\frac{\tau_t}{\bar{\tau}} = \left(\frac{\tau_{t-1}}{\bar{\tau}}\right)^{\rho_\tau} \left(\frac{B_t}{B_{t-1}}\right)^{(1-\rho_\tau)\gamma_B^\tau} \left(\frac{Y_t}{Y_{t-1}}\right)^{(1-\rho_\tau)\gamma_Y^\tau}. \tag{34}$$

The level parameter of the tax code τ_t^L adjusts such that the average tax rate on income equals this target level, i.e.,

$$\tau_t = \frac{\mathbb{E}_t \left(w_t n_{it} h_{it} + \mathbb{I}_{h_{it}=0}\Pi_t^E\right) - \tau_t^L \mathbb{E}_t \left(w_t n_{it} h_{it} + \mathbb{I}_{h_{it}=0}\Pi_t^E\right)^{\bar{\tau}^P}}{\mathbb{E}_t \left(w_t n_{it} h_{it} + \mathbb{I}_{h_{it}=0}\Pi_t^E\right)}, \tag{35}$$

where $\mathbb{E}_t$ is the expectation operator, which here gives the cross-sectional average. Total government tax revenues T_t are then $T_t = \tau_t \left(w_t n_{it} h_{it} + \mathbb{I}_{h_{it}\neq 0}\Pi_t^U + \mathbb{I}_{h_{it}=0}\Pi_t^E\right)$ and the government budget constraint determines government spending residually: $G_t = B_{t+1} + T_t - R_t^b/\pi_t B_t$.

There are thus two shocks to government rules: monetary policy shocks, ϵ_t^R, and structural deficit shocks, ϵ_t^D. We assume these shocks to be log normally distributed with mean zero.

2.4 Goods, asset, and labor market clearing

The labor market clears at the competitive wage given in (4). The liquid asset market clears whenever the following equation holds:

$$\begin{aligned}
B_{t+1} + q_t^\Pi &= B^d(A_t, w_t, w_t^F, \Pi_t^E, \Pi_t^U, q_t, r_t, q_t^\Pi, q_{t-1}^\Pi, R_t^b, \pi_t, \pi_t^W, \tau_t, \Theta_t, \mathbb{W}_{t+1}) \\
&:= \mathbb{E}_t \left[\lambda b_{a,t}^* + (1-\lambda)b_{n,t}^*\right], \tag{36}
\end{aligned}$$

where $b_{a,t}^*, b_{n,t}^*$ are functions of the states (b, k, h), and depend on how households value asset holdings in the future, $\mathbb{W}_{t+1}(b, k, h)$, and the current set of prices (and tax rates) $(A_t, w_t, w_t^F, \Pi_t^E, \Pi_t^U, q_t, r_t, q_t^\Pi, q_{t-1}^\Pi, R_t^b, \pi_t, \pi_t^W, \tau_t, \Theta_t, \mathbb{W}_{t+1})$. Future prices do not show up because we can express the value functions such that they summarize all relevant information on the expected future price paths. Expectations in the right-hand-side expression are taken w.r.t. the distribution $\Theta_t(b, k, h)$. Equilibrium requires the total *net* amount of bonds the household sector demands, B^d, to equal the supply of government bonds plus the value of profit shares. In gross terms there are more liquid assets in circulation; some households borrow up to $\underline{B}$.

The value of profit shares is, given the linearized solution, determined by a no-arbitrage condition between bonds and profit shares. Both need to have the same expected return:

$$q_t^\Pi R_t^b = \mathbb{E}_t \pi_{t+1} \left[(1 - \iota^\Pi) q_{t+1}^\Pi + \omega^\Pi \Pi_{t+1}^F \right]. \tag{37}$$

Last, the market for capital has to clear, i.e.,

$$\begin{aligned}
K_{t+1} = K^d(A_t, w_t, w_t^F, \Pi_t^E, \Pi_t^U, q_t, r_t, q_t^\Pi, q_{t-1}^\Pi, R_t^b, \pi_t, \pi_t^W, \tau_t, \Theta_t, \mathbb{W}_{t+1}) \\
:= \mathbb{E}_t[\lambda k_t^* + (1 - \lambda)k] , \tag{38}
\end{aligned}$$

where K^d defines the aggregate supply of funds from households—both those that trade capital, λk_t^*, and those that do not, $(1 - \lambda)k$. Again k_t^* is a function of the current prices and continuation values. The goods market then clears due to Walras' law, whenever labor, bonds, and capital markets clear.

When we consider the representative household model, we can think of K^d and B^d as simply given by (30) and (31). In other words, the representative household model only changes equilibrium conditions in replacing the Bellman equation and the capital and bonds demand equations, but leaves the entire other model structure unchanged.

2.5 Equilibrium

A *sequential equilibrium with recursive planning* in our model is a sequence of policy functions $\{x_{a,t}^*, x_{n,t}^*, b_{a,t}^*, b_{n,t}^*, k_t^*\}$, a sequence of value functions $\{\mathbb{W}_t\}$, a sequence of prices $\{w_t, w_t^F, \Pi_t^E, \Pi_t^U, q_t, r_t, q_t^\Pi, R_t^b, \pi_t, \pi_t^W, \tau_t\}$, a sequence of stochastic states $\{A_t, Z_t, \Psi_t, \mu_t^Y, \mu_t^W, D_t\}$ and shocks $\{\epsilon_t^A, \epsilon_t^Z, \epsilon_t^\Psi, \epsilon_t^{\mu Y}, \epsilon_t^{\mu W}, \epsilon_t^D, \epsilon_t^R\}$, aggregate capital and labor supplies $\{K_t, N_t\}$, distributions Θ_t over individual asset holdings and productivity, and expectations for the distribution of future prices, such that

1. Given the functionals $\mathbb{E}_t \mathbb{W}_{t+1}$ for the continuation value and period-t prices, policy functions $\{x_{a,t}^*, x_{n,t}^*, b_{a,t}^*, b_{n,t}^*, k_t^*\}$ solve the households' planning problem; and given the policy functions $\{x_{a,t}^*, x_{n,t}^*, b_{a,t}^*, b_{n,t}^*, k_t^*\}$ and prices, the value functions $\{\mathbb{W}_t\}$ are a solution to the Bellman equation (27).

2. Distributions of wealth and income evolve according to households' policy functions.

3. The labor, the final goods, the bond, the capital, and the intermediate goods markets clear in every period, interest rates on bonds are set according to the central bank's Taylor rule, fiscal policies are set according to the fiscal rules, and stochastic processes evolve according to their law of motion.

4. Expectations are model consistent.

3 Numerical solution and estimation technique

Solving for the sequential equilibrium of the model is a challenging task, in particular in an estimation setting. In the following, we develop our new Bayesian reduction technique starting from well-known principles in solving for state-space solutions of difference equations. First, we summarize how one can write a heterogeneous-agent model in form of a system of difference equations, following Reiter (2009). Then, we explain how our method achieves the necessary complexity reduction that renders the solution algorithm fast enough to repeatedly obtain a state-space solution within the estimation procedure, even when the household problem is high dimensional. Third, we propose a metric to assess the quality of our solution relative to alternatives. Finally, we discuss how the proposed state-space solution is integrated in a Bayesian estimation.

3.1 From representative-agent to heterogeneous-agent solution

To understand the solution procedure we propose, it is useful to start by thinking about the representative agent twin of our model as in Section 2.2.3. Following Klein (2000) and Schmitt-Grohé and Uribe (2004), we represent the sequential equilibrium as the solution of a non-linear difference equation

$$\mathbb{E}_t F(X_t^*, X_{t+1}^*) = 0, \tag{39}$$

where X_t^* are combined state and control variables. While the partitioning of X_t^* into states and controls matters for the solution of (39), it is not central to our argument. Instead, in moving from the representative-agent to the heterogeneous-agent model, it is useful to partition the variables in X_t^* in terms of whether they are household choices or not. This means, we separate out composite consumption, x_t, capital demand, K_t^d, and bond demand, B_t^d, and capture by X_t all other aggregate variables. Both B_t^d and K_t^d are state variables and x_t is a control. The corresponding equilibrium conditions are the consumption Euler equations and the budget constraint of the household, equations (29) – (31). In equilibrium, the capital demanded by firms, K_t, and the bonds offered by the government, B_t, need to align with the household plans, i.e., (36) and (38) need to hold. This means, we can write the representative household equilibrium as

$$\mathbb{E}_t F(\{x_t, K_t^d, B_t^d\}, X_t, \{x_{t+1}, K_{t+1}^d, B_{t+1}^d\}, X_{t+1}) = 0. \tag{40}$$

Moving to the heterogeneous agent case we simply replace the household variables by a large vector of variables and the corresponding optimality and equilibrium conditions (see Reiter, 2009), discretizing the idiosyncratic state space. Specifically, the Bellman equation

18

(27) pins down the optimal policy at each node (b, k, h) and the sequence of continuation values, $\mathbb{W}_t(b, k, h)$.[15] Similarly, for each node, (28) describes how the mass of households evolves over time. Taken together, (27) and (28) define a set of equilibrium conditions for the evolution of continuation values and distribution masses—two conditions for each node (b, k, h). These "replace" (29), (30), and (31) of the representative agent model. We summarize these "idiosyncratic variables", continuation values and masses, by f_t in the following and obtain as the equilibrium condition—with a slight abuse of notation:

$$\mathbb{E}_t F(f_t, X_t, f_{t+1}, X_{t+1}) = 0. \tag{41}$$

In principle, a first-order approximation of (41) can be solved in the usual way: Calculating the Jacobian of F and, e.g., running a QZ-decomposition as proposed by Klein (2000).

3.2 Reducing complexity

In practice, however, the direct attack to linearizing and solving (41) becomes infeasible, given the size of the system in our application. This is due to the high number of nodes of the liquid/illiquid asset-income grid ($100 \times 100 \times 22$). Reiter (2009), Bayer and Luetticke (2020), and Ahn et al. (2018) suggest methods that tackle the issue based on reducing the number of state and control variables. The three approaches do so, respectively, before solving for the stationary equilibrium, after solving for this equilibrium but before knowledge of the dynamics, or after linearizing the system of differential equations. These approaches render a *single solution* of the economic model at a good quality of approximation numerically feasible. For *estimation*, we develop a novel Bayesian reduction technique for HANK models that is more aggressive and also better informed. Below we provide an overview, further technical details can be found in Appendix C.

Using the partitioning of variables and parameters to simplify the linearization

The first improvement is based on the observation that the dimensionality of the heterogeneous-agent part, f_t, is much larger than that of the aggregate part, X_t. Keeping this partitioning, we write the linearized version of the difference equation (41) as

$$\underbrace{\begin{bmatrix} A_{ff} & A_{fX} \\ A_{Xf} & A_{XX} \end{bmatrix}}_{=A} \begin{bmatrix} f_t \\ X_t \end{bmatrix} = -\mathbb{E}_t \underbrace{\begin{bmatrix} B_{ff} & B_{fX} \\ B_{Xf} & B_{XX} \end{bmatrix}}_{=B} \begin{bmatrix} f_{t+1} \\ X_{t+1} \end{bmatrix}, \tag{42}$$

---

[15] In practice our algorithm is based on the endogenous grid method of Carroll (2006) and Hintermaier and Koeniger (2010) and thus works with marginal value functions instead of value functions.

where we ordered, in line with the ordering of variables, the set of "idiosyncratic" equations—(27) and (28)—first and all other equations last.

This partitioning is useful not only conceptually but also very practically. The idiosyncratic equations contain only a small subset of the parameters of the model. In other words, A_{ff}, B_{ff}, A_{fX}, and B_{fX} only depend on this small subset. What is more, A_{Xf} and B_{Xf} can be made parameter-free if we introduce the aggregate capital/bond holdings of households as auxiliary variables such that the distribution no longer enters the aggregate model block directly. Therefore, we only need to update A_{XX} and B_{XX} when changing parameters during the estimation that do not directly appear in the household problem. For example, changing the fiscal-policy or Taylor-rule coefficients, or nominal rigidities only affects A_{XX} and B_{XX}. This means that the number of derivatives to be calculated in the second step, i.e., when estimating the business cycle model with heterogeneous agents, is the same as the number of derivatives to be updated during the estimation of its representative agent "twin".

Importantly, those parameters directly affecting the other "idiosyncratic" blocks also affect the stationary equilibrium. This means they can be identified (calibrated) in an incomplete markets model using time-series averages of aggregate and cross-sectional data, approaching the estimation effectively in two steps. In our case these include: the discount factor, the liquidity of assets, borrowing constraints, and the average income risk.

Dimensionality reduction prior to knowledge of the dynamics of the economy

Naturally the partitioning itself does not solve the problem of A and B being very large matrices because of the high dimensional idiosyncratic grid. To tackle this problem, we proceed in two steps. First, we reduce the dimensionality without knowledge of the dynamics of the economy (where we extend Bayer and Luetticke, 2020). Second, we leverage that, after solving the model once, we can reduce the model further based on the dynamics of the model.

The first step is to think of f_t as parameters of the deviation of value functions and distributions from their stationary equilibrium counterparts. This view of f_t as parameters of functions immediately implies that the dimensionality of f_t no longer needs to be tightly connected to the size of the asset-income grid. We construct these difference functions as a mix of Chebychev polynomials and linear interpolations. The Chebychev polynomials are used to construct the nodal values of the marginal value functions $\frac{\partial \mathbb{W}_t}{\partial b}$ and $\frac{\partial \mathbb{W}_t}{\partial k}$ used in the first-order conditions of the optimization problem (27). When perturbing the problem (27), we perturb the coefficients instead of the function values themselves. To reduce complexity, we treat those polynomial coefficients as fixed (and do not perturb them) that are numerically small (in absolute value) in the stationary equilibrium or do not affect the partial derivative of the marginal value functions with respect to prices. The latter aspect improves on Bayer

and Luetticke (2020); see Appendix C.1. The discrete cosine transform (DCT) and its inverse can be used to efficiently transform nodal values to coefficients and reverse. We also use the DCT to project the equilibrium conditions to the space of the perturbed coefficients.

The distribution function Θ_t, see (28), we treat similarly to the marginal value functions, improving on Bayer and Luetticke (2020). Following their approach, we represent the distribution by marginals, F_t^b, F_t^k, F_t^h and a copula, $C_t(\cdot)$.[16] Different from their approach, we write the copula $C_t(F_t^b, F_t^k, F_t^h)$ at time t as the sum of the linear interpolants generated from the steady-state copula $\bar{C}(\cdot)$ and a perturbation term, $\hat{C}_t(\cdot)$. This perturbation term is new compared to Bayer and Luetticke (2020) and is a linear interpolant, too. We allow the two components to have different nodal grids. Again, we write the nodal values of the perturbation term in the form of Chebychev polynomials (by means of a DCT).[17] The marginal distributions, F_t^b, F_t^k, F_t^h, enter directly in f_t. More details can be found in Appendix C.1.

Dimensionality reduction based on knowledge of the economy's dynamics

This first reduction step allows us to solve the model, but if we want the solution to be sufficiently precise, the number of variables in the difference equation remains high and thus the computational time for the QZ-decomposition long. One way to address this issue is further model reduction. Ahn et al. (2018) give an overview of model reduction techniques for difference equations. Model reduction in our case means finding an orthonormal basis $\mathcal{P} \in \mathbb{R}^{n \times m}$ with $m << n$ such that we can write $f_t \approx \mathcal{P}Y_t$ and replace the system (42) by a system with factors Y_t,

$$\underbrace{\begin{bmatrix} \mathcal{P}'A_{ff}\mathcal{P} & \mathcal{P}'A_{fX} \\ A_{Xf}\mathcal{P} & A_{XX} \end{bmatrix}}_{=A'} \begin{bmatrix} Y_t \\ X_t \end{bmatrix} = -\mathbb{E}_t \underbrace{\begin{bmatrix} \mathcal{P}'B_{ff}\mathcal{P} & \mathcal{P}'B_{fX} \\ B_{Xf}\mathcal{P} & B_{XX} \end{bmatrix}}_{=B'} \begin{bmatrix} Y_{t+1} \\ X_{t+1} \end{bmatrix}. \tag{43}$$

The solution of this reduced model should give us an arbitrarily close approximation to the solution of the original problem (42).

For a given set of parameters, this can be achieved by solving first the original system (42) given a prior parameterization that includes the shock processes. Based on this solution, we calculate the variance-covariance matrix, Σ_f, of each sub-group of elements in f (the value functions and the copula separately) and perform a Jordan eigenvalue decomposition thereof.

---

[16]There is a numerical advantage of writing the distribution in the form of a copula and marginals. The copula reflects cross-terms, which can be expected to be of lesser importance for aggregate dynamics. The separation then allows us to calibrate separate degrees of precision in line with the differences in importance.

[17]The fact that C is a copula poses some parameter restrictions on the histogram of $\hat{C}$. Its marginals have to be zero. These restrictions translate naturally into zero-coefficients of some Chebychev polynomials because higher-order polynomials on the roots grid sum to zero.

This gives us a factorization of f, where factors associated with small eigenvalues, λ_{2f}, are approximately constant and thus irrelevant for the model dynamics and in particular the estimation:

$$\Sigma_f = \begin{bmatrix} \mathcal{Q}'_{1f} & \mathcal{Q}'_{2f} \end{bmatrix} \begin{bmatrix} \lambda_{1f} & 0 \\ 0 & \lambda_{2f} \end{bmatrix} \begin{bmatrix} \mathcal{Q}_{1f} \\ \mathcal{Q}_{2f} \end{bmatrix}. \tag{44}$$

This means, we can choose $\mathcal{P} = \mathcal{Q}'_{1f}$ as basis for the model reduction. The reduced model then, by construction, has a tiny approximation error relative to the full model at the parameterization used to construct $\mathcal{P}$. Furthermore, one can expect that this basis $\mathcal{P}$ remains a reasonably good basis in the vicinity of the parameters.[18] Therefore, our algorithm updates $\mathcal{P}$ only infrequently. We show, for our application, that, even when parameters change but $\mathcal{P}$ does not, the solution of the reduced model is a precise approximation of the full model solution. In Appendix C.2 and C.3, we argue in more detail, why one can expect a strong model reduction possibility and why this reduction can be expected to be stable to parameter variations in the estimation.

In practice, we generate $\mathcal{P}$ initially based on the model's priors, and then update it once during mode-finding and once after finding a tentative mode of the parameter distributions but before running the Markov-Chain Monte-Carlo algorithm. In our application, the model reduction leaves us with 376 state variables and 77 controls. This includes 21 aggregate state variables and 44 aggregate controls.

3.3 Assessment of quality

We check the precision of our second model-reduction step by comparing the likelihood and the impulse response functions (IRF) across model solutions. The first solution we compare to is the one in (42) without second model-reduction step. The second solution we compare to is the sequence-space method proposed by Boppart et al. (2018) and Auclert et al. (2021b).

We suggest to evaluate the distance between two model solutions based on the IRFs they produce for the data used in estimation. Concretely, we define the distance between the reference model solution $S1$ and the alternative solution $S2$ as

$$D_{S1,S2}(H, x|\vartheta) = 1 - \frac{\sum_{s \in S} \sum_{h=1}^{H} [IRF_{S1}(x, s, h|\vartheta) - IRF_{S2}(x, s, h|\vartheta)]^2}{\sum_{s \in S} \sum_{h=1}^{H} [IRF_{S1}(x, s, h|\vartheta)]^2}, \tag{45}$$

where H is the horizon up to which we evaluate the response of variable x to a one-standard-deviation shock ϵ_s, summing over all aggregate shocks, $s \in S$. ϑ is the parameter vector. The metric is akin to a forecast error variance decomposition: the distance informs us about the

---

[18]Ahn et al. (2018) argue that small variations compared to the ideal minimal basis can be compensated by retaining a somewhat too high dimensional basis.

fraction of the forecast error variance for x under the reference solution $S1$ that is "explained" by the alternative solution $S2$. We show in Section 4.4 below that our solution is practically identical to both reference solutions.

3.4 Estimation

As explained, it is useful to group the model parameters ϑ into those that affect the stationary equilibrium and those that do not when thinking about estimating the model. The representative agent literature knows this split, too, and sometimes calibrates parameters like steady-state markups or average government expenditure shares (e.g. Smets and Wouters, 2007; Christiano et al., 2005). In the heterogeneous-agent setting, the set of parameters that affect the stationary equilibrium is larger. Thus, more parameters can be calibrated. At the same time, parameters that affect the steady state are more costly to estimate from the time-series properties of data because they require recalculating the stationary equilibrium. For this reason, we calibrate those parameters to cross-sectional moments and time-series averages of the data. This means that in our application, compared to the typical representative-agent estimation that tries to calibrate as little as possible (e.g., Justiniano et al., 2011), we additionally include the discount factor, the risk aversion, the Frisch elasticity of labor supply, and the capital share in the set of calibrated parameters. Of course, we also calibrate the forcing processes of heterogeneity itself. We come back to this in more detail in Section 4.

Those parameters that do not affect the stationary equilibrium can be estimated from the time-series evolution of the data (cross-sectional and aggregate). For this estimation, we use an off-the-shelve Bayesian approach as described in An and Schorfheide (2007) and Fernández-Villaverde (2010). In particular, we use the Kalman filter to obtain the likelihood from the state-space representation of the model solution and employ a standard random walk Metropolis-Hastings (RWMH) algorithm to generate draws from the posterior likelihood after an extensive mode finding. Smoothed estimates of the states at the posterior mean of the parameters are obtained via a Kalman smoother of the type described in Koopman and Durbin (2000) and Durbin and Koopman (2012). One likelihood evaluation takes ca. 550 ms on a desktop computer (Intel i7-10700K, code written in Julia), 80% of the time is needed for the model solution, the remainder for the Kalman filter.[19] The full grid for income and wealth has 220,000 nodes and the model after the first step of dimensionality reduction which is used to calculate $\mathcal{P}$ has 3484 variables. Finding the stationary equilibrium takes roughly 7 minutes, the first linearized solution that does not yet use the second model reduction takes

---

[19]We do not leverage the fact that the model solution can be obtained without repeating the QZ-decomposition when changing only the parameters of the forcing processes. When only parameters of the exogeneous shock processes are estimated, the model solution can, as usual, be obtained in virtually no time.

around 15 minutes to compute.

We deliberately choose the most commonly used approach in estimating DSGE models, but our state-space based solution approach also lends itself directly to more sophisticated estimation techniques. One particular example is the Sequential Monte Carlo approach (Herbst and Schorfheide, 2014, 2015) that Acharya et al. (2021) advocate for the estimation of large heterogeneous-agent business cycle models.

4   Parameterization, priors, posteriors, and quality

We estimate several variants of the model using the procedure outlined above. In the following we focus on two main variants and compare them to their representative agent twin; details on all other variants can be found in the appendix. First, we estimate the model on aggregate data alone and allow only for standard aggregate shocks (HANK). Second, we include in the estimation distributional time-series data and allow for shocks to income risk and tax progressivity, for which we also add observables (HANK-X).[20]

4.1   Calibrated parameters

We fix some parameters that affect the stationary equilibrium targeting average data ratios; see Table 1 (all at the quarterly frequency of the model).[21] In addition, we directly take some parameter estimates from the literature. In particular, we take the idiosyncratic income process from Storesletten et al. (2004), which gives us $\rho_h = 0.98$ and $\bar{\sigma}_h = 0.12$. Guvenen et al. (2014) provide the probability that a household will fall out of the top 1 percent of the income distribution in a given year, which we take as the transition probability from entrepreneur to worker, $\iota = 6.25\%$. We set the relative risk aversion, ξ, to 4, which is common in the incomplete markets literature; see Kaplan and Violante (2014).[22] We set the Frisch elasticity to 0.5; see Chetty et al. (2011). The steady-state price and wage markups are both fixed to 10%, following Born and Pfeifer (2014).

All other calibrated parameters are closely tied to time-series averages of data moments. While they are calibrated jointly, we present them as if they are only informed by the statistic that is most informative for a given parameter: The transition probability to become an entrepreneur, $\zeta = 0.0002$, pins down the per capita profits of entrepreneurs and therefore the top 10 share in wealth. The discount factor, $\beta = 0.984$, and the liquidity of assets, $\lambda = 0.058$,

---

[20]See Footnotes 12 and 13 for the adjustments to the baseline model.

[21]Appendix A.2 provides the table of steady-state parameters for the recalibrated representative agent analogue of our model.

[22]We also estimate our model with a relative risk aversion of 2; see Section 5.3 and Appendix B.

Table 1: Calibration (quarterly frequency)

Par.	Value	Description	Target	Data Source
Households: Income process				
ρ_h	0.980	Persistence labor income	Storesletten et al. (2004)	
σ_h	0.120	Standard dev. labor income	Storesletten et al. (2004)	
ι	0.063	Trans. prob. from E. to W.	Guvenen et al. (2014)	
ζ	2.0E-5	Trans. prob. from W. to E.	Top 10 wealth share: 67%	WID 1954-2019
Households: Financial frictions				
λ	0.058	Portfolio adj. prob.	Liquid to illiquid, $\frac{B+\bar{q}^\Pi}{K}=0.25$	SCF 1950-2016
$\bar{R}$	0.018	Borrowing penalty	Share of borrowers: 16%	SCF 1983-2016
$\bar{q}^\Pi/Y$	1.140	Value of profit shares	Gov. debt to output, $\frac{B}{Y}=1.72$	NIPA 1954-2019
Households: Preferences				
β	0.984	Discount factor	Capital to output, $\frac{K}{Y}=11.44$	NIPA 1954-2019
ξ	4.000	Relative risk aversion	Kaplan and Violante (2014)	
γ	2.000	Inverse of Frisch elasticity	Chetty et al. (2011)	
Firms				
α	0.680	Share of labor	Average labor income share	BLS 1954-2019
δ_0	0.018	Depreciation rate	7.0% p.a.	NIPA 1954-2019
$\bar{\eta}$	11.000	Elasticity of substitution	Born and Pfeifer (2014)	
$\bar{\zeta}$	11.000	Elasticity of substitution	Born and Pfeifer (2014)	
Government				
$\bar{\tau}^L$	0.175	Tax rate level	Gov. consumption, $\frac{G}{Y}=0.2$	NIPA 1954-2019
$\bar{\tau}^P$	0.120	Tax progressivity	Average progressivity	SOI 1954-2019
$\bar{R}^b$	1.000	Gross nominal rate	Growth $\approx$ interest rate	
$\bar{\pi}$	1.000	Gross inflation	Indexation, w.l.o.g.	

Notes: Calibration targets are the sample averages when a data source is given. Otherwise the parameter is fixed to the value in the cited literature. BLS: Bureau of Labor Statistics. NIPA: National Income and Product Accounts. SCF: Survey of Consumer Finances. SOI: Statistics of Income. WID: World Inequality Database. Details on the data can be found in Appendix A.

pin down the capital to output ratio and the share of liquid assets in household portfolios. The borrowing penalty $\bar{R} = 0.018$ determines how many households are indebted. The difference between total liquidity and government bonds pins down the value of profit shares relative to output, $\bar{q}^\Pi/Y = 1.14$, which determines the ratio of ω^Π and ι^Π as $\omega^\Pi/\iota^\Pi = \eta\bar{q}^\Pi/Y$ through the steady-state version of (37). We set steady-state inflation to zero as we have assumed indexation to the steady-state inflation rate in the Phillips curves. We set the steady-state net interest rate on bonds also to zero, in order to broadly capture the average federal funds rate in real terms minus output growth over 1954 – 2019.

The share of labor in production, $\alpha = 0.68$, is pinned down by the average labor income share (given η). The average quarterly depreciation of $\delta_0 = 0.0175$ can be read off the average depreciation rates on US capital (including buildings). We set the average taxation level, $\bar{\tau}^L = 0.175$, such that the budget balances for the observed level of government consumption. Finally, we follow Ferriere and Navarro (2018) in constructing a direct estimate for tax progressivity, extending their estimates until 2017; see Appendix A.1.2. This approach uses the our assumed non-linear tax schedule, which approximates the progressivity of the US tax system well (see Heathcote et al., 2017), to measure progressivity based on Mertens and Montiel Olea (2018)'s estimates of average marginal tax rates. The tax progressivity exponent, $\bar{\tau}^P = 0.120$, matches the time-series average of this statistic.

4.2   Time-series data used for estimation

For the estimation, we use quarterly US data from 1954Q3 to 2019Q4 and include the following seven observable time series: the growth rates of per capita GDP, private consumption, investment, and wages, all in real terms; the logarithm of the level of per capita hours worked; the log difference of the GDP deflator; and the (shadow) federal funds rate. Our model is stationary so all growth rates are demeaned; see Appendix A.1.2 for a formal depiction of the vector of observables. These data are standard in the estimation of DSGE models.

In the HANK-X extension, we add more data with shorter and/or non-quarterly availability: First, cross-sectional information on wealth and income shares of the top 10 percent. These are available at an annual frequency from 1954 to 2019 from the World Inequality Database.[23] The reason we focus on the top 10 wealth and income shares is that these measures are most similar across alternative, but less frequently available, data sources such as the Survey of Consumer Finances (SCF); see Kopczuk (2015). Second, we use the time series of the tax progressivity estimates that we construct in Appendix A.1.2. Third, we add income risk estimates, available at a quarterly frequency from 1983Q1 to 2013Q1, from Bayer et al. (2019) based on panel data in the Survey of Income and Program Participation (SIPP). In this extended estimation we also allow for shocks to progressivity and income risk, which directly affect the distribution of income. We allow for measurement error on the cross-sectional data (the top 10 shares) to avoid stochastic singularity. For simplicity, we treat the measurement error on the top 10 shares as classical (normal and i.i.d.) despite the fact that the shares are non-linear functions of the sampled micro data. In other words, we assume that measurement errors reflect more than just sampling uncertainty.

---

[23]This database draws on work by Piketty, Saez, and Zucman; see, e.g., Piketty and Saez (2003) or Saez and Zucman (2016).

4.3 Priors and posteriors

Columns 1-4 of Table 2 present the parameters we estimate, their assumed prior distributions and their posteriors. The priors and posteriors for the distributions of the shock processes are listed in Appendix A.3. Where available, we use prior values that are standard in the literature and independent of the underlying data.

Priors

Following Justiniano et al. (2011), we impose a gamma distribution with prior mean of 5.0 and standard deviation of 2.0 for δ_2/δ_1, the elasticity of marginal depreciation with respect to capacity utilization, and a gamma prior with mean 4.0 and standard deviation of 2.0 for the parameter controlling investment adjustment costs, ϕ. For the slopes of price and wage Phillips curves, κ_Y and κ_w, we assume gamma priors with mean 0.10 and standard deviation 0.03. This corresponds to price and wage contracts having an average length of four quarters at the prior mode. Regarding the profit-shares parameters ι^Π and ω^Π, we assume that ι^Π follows a shifted beta distribution with mean 0.5 and standard deviation 0.25. We set the extrema of the shifted distribution such that the expected duration of the profit shares, $\frac{1}{\iota^\Pi}$, is at least ten years and at most 200 years. The duration at the prior mode is 20 years. The value for ω^Π then follows from keeping the stationary equilibrium value of $\bar{q}^\Pi$ constant.

For monetary policy, we estimate feedback parameters in the Taylor rule for inflation and output growth, θ_π and θ_Y. We impose normal distributions with prior means of 1.7 and 0.13, respectively. In addition, we allow for interest rate smoothing with parameter ρ_R. Here we assume a beta distribution with parameters $(0.5, 0.2)$.

In the bond rule, the debt-feedback parameter γ_B is assumed to follow a gamma distribution with mean 0.10 and standard deviation 0.08. This centers the prior for the auto-correlation of debt around 0.9 and implies a half-life of between one and eight years for a deviation in debt. The parameters governing the feedback to inflation and output growth, γ_π and γ_Y, follow standard normal distributions. Similarly, the autoregressive parameters, in the tax rules, ρ_P and ρ_τ, are assumed to follow beta distributions (with mean 0.5 and standard deviation 0.2). The feedback parameters for average tax rates, γ_Y^τ and γ_B^τ, follow standard normal distributions.

Following Smets and Wouters (2007), the autoregressive parameters of the shock processes are assumed to follow a beta distribution with mean 0.5 and standard deviation 0.2. The standard deviations of the shocks follow inverse-gamma distributions with prior mean 0.1% and standard deviation 2%.[24] In our baseline we do not include measurement errors, but allow for these when including estimates of top income and wealth shares as additional data.

---

[24]In the HANK-X extension, we use a higher prior mean for income risk shocks, s, given the evidence in Bayer et al. (2019).

Posteriors

In Table 2, columns 5-7 report the posterior distributions across the three main estimation variants: RANK, HANK, and HANK-X. Here, we focus on the frictions and policy parameters. The estimated parameters of the exogenous driving processes can be found in Appendix A.3. Checks on the convergence of the estimator are provided in Appendix A.8.[25] The parameter estimates for HANK and HANK-X are typically close to the RANK estimates, but there are few notable differences.

In particular, the estimated investment friction and the price rigidities in RANK are a bit larger. For the investment adjustment costs, this reflects, in part, that the portfolio adjustment costs at the household level already generate inertia in aggregate investment. Our estimates of the wage- and price-setting frictions imply that wages adjust roughly every two to three quarters in all models and prices adjust every third quarter in HANK and every fourth quarter in RANK.

The estimated policy rules are even more similar across models. There is substantial interest rate inertia. All variants estimate a coefficient of 0.8 for interest smoothing. The Taylor rule coefficient on inflation is between 2.0 and 2.2; the one on output growth is between 0.2 and 0.3. The fiscal rule that governs deficits and hence government spending exhibits a countercyclical response to inflation with elasticities between, -2.9 and -2.2. The elasticities with respect to output growth are between -0.8 and -0.4. Deficits feature a high degree of persistence as well. The tax rule that governs average taxation has much less inertia. This implies, given the transitory nature of output growth variations, that tax rates respond mostly to the level of government debt. More debt implies higher taxes.

When we estimate the income risk process in HANK-X, we find income risks to be procyclical in the sense that they go up when other shocks drive up output growth. Given the size of shocks that we estimate, however, the feedback is small. A one percent output growth increase, leads to a 0.55 percent higher income risk; a one standard deviation shock to income risk leads to a 70 percent higher level of risk.

The posterior for ι^Π implies a duration of profit shares of 27 years on average in the HANK specification and 14 years in HANK-X. Using the top 10 income data, leads to a lower duration and therefore higher payout share ω^Π (13% vs 23%). The higher payout share implies a smoother income inequality series, but amplifies the volatility of aggregate consumption. Looking at the credible intervals, one sees that adding the cross-sectional data substantially increases the precision of the estimate.

---

[25]We estimate each model using a single RWMH chain after an extensive mode search. After a long burn-in, 400,000 draws from the posterior are used to compute the posterior statistics. The acceptance rates across chains are between 20% and 30%. Appendix A.8 provides Geweke (1992) convergence statistics as well as traceplots of individual parameters.

Table 2: Prior and posterior distributions of estimated parameters

Parameter	Distribution	Prior Mean	Prior Std. Dev.	Posterior RANK	Posterior HANK	Posterior HANK-X
				Frictions		
δ_s	Gamma	5.00	2.00	0.723	0.987	0.706
				(0.418, 1.098)	(0.766, 1.225)	(0.536, 0.897)
ϕ	Gamma	4.00	2.00	7.918	3.561	1.941
				(6.039, 10.074)	(2.936, 4.246)	(1.811, 2.071)
κ	Gamma	0.10	0.03	0.119	0.173	0.146
				(0.087, 0.155)	(0.134, 0.216)	(0.113, 0.182)
κ_w	Gamma	0.10	0.03	0.279	0.279	0.239
				(0.203, 0.363)	(0.217, 0.344)	(0.184, 0.299)
ι^Π	Beta*	0.50	0.25	—	0.401	0.700
				(—, —)	(0.075, 0.810)	(0.372, 0.948)
				Monetary policy		
ρ_R	Beta	0.50	0.20	0.795	0.784	0.803
				(0.769, 0.819)	(0.757, 0.810)	(0.778, 0.826)
σ_R	Inv.-Gamma	0.10	2.00	0.237	0.235	0.231
				(0.218, 0.258)	(0.216, 0.256)	(0.212, 0.251)
θ_π	Normal	1.70	0.30	2.165	1.953	2.078
				(1.983, 2.362)	(1.750, 2.169)	(1.858, 2.315)
θ_Y	Normal	0.13	0.05	0.254	0.201	0.219
				(0.184, 0.324)	(0.133, 0.269)	(0.150, 0.288)
				Fiscal policy: deficit		
ρ_D	Beta	0.50	0.20	0.959	0.960	0.968
				(0.924, 0.983)	(0.923, 0.990)	(0.936, 0.992)
γ_B	Gamma	0.10	0.08	0.086	0.047	0.020
				(0.035, 0.136)	(0.012, 0.088)	(0.004, 0.043)
γ_π	Normal	0.00	1.00	-2.730	-2.858	-2.174
				(-3.136, -2.346)	(-3.20, -2.534)	(-2.375, -1.986)
γ_Y	Normal	0.00	1.00	-0.718	-0.784	-0.436
				(-0.877, -0.568)	(-0.88, -0.693)	(-0.494, -0.383)
				Fiscal policy: taxes		
ρ_τ	Beta	0.50	0.20	0.480	0.414	0.493
				(0.357, 0.589)	(0.247, 0.557)	(0.393, 0.591)
γ_B^τ	Normal	0.00	1.00	3.568	2.944	3.293
				(2.481, 4.704)	(1.936, 4.031)	(3.254, 3.334)
γ_Y^τ	Normal	0.00	1.00	-1.746	1.571	-0.921
				(-3.108, -0.385)	(1.118, 1.979)	(-0.943, -0.90)
				Income risk		
ρ_s	Beta	0.70	0.20	—	—	0.545
				(—, —)	(—, —)	(0.476, 0.607)
Σ_Y	Normal	0.00	100.00	—	—	28.88
				(—, —)	(—, —)	(28.864, 28.895)
Log marginal data density (only aggregate data)				6581	6586	
Log marginal data density (with cross-sectional data)						6622

Notes: The table displays the estimated parameters, their priors and posterior means across three model variants: HANK, HANK-X, and RANK. The parameters of the shock processes are shown in Appendix A.3. The 90% credible intervals are shown in parentheses. Posteriors are obtained by an MCMC method. The standard deviations have been multiplied by 100 for better readability. *: The parameter actually estimated and displayed is $\frac{800\iota_\Pi - 1}{19}$ to ensure an expected duration of the profit shares between 10 and 200 years.

Table 2 also reports the marginal data densities for the three model estimates. When estimating the models with aggregate data alone, HANK is more strongly supported by that data. What is more, the HANK models predict certain correlations of cross-sectional and aggregate data that RANK, by construction, cannot.[26]

4.4 Quality of the model reduction

We assess the quality of our model reduction based on the estimated posterior distribution. First, we evaluate the quality of the second-stage of the model reduction. For this purpose, we draw 1000 parameter vectors from the posterior distribution and, for each parameter draw, we solve the model once with, and once without, the second stage reduction. The second-stage reduction matrix $\mathcal{P}$ is kept constant throughout the experiments. This gives us a sampled distribution of distances between the two model solutions based on the impulse response functions (IRF). We find that the distance between the two solutions is basically zero. Table 3 reports the 1st percentile of the sampled quality measure (i.e., the largest distance). Note that the first-stage reduction is, by construction, invariant to the estimated parameters. This means that, at least in the vicinity of the estimated parameters, there is virtually no loss from using the second-stage model reduction even when $\mathcal{P}$ is not re-optimized.

Still, one might be concerned that the necessary first-stage reduction introduces approximation errors. To assess this, we evaluate the solution under our method against a sequence-space Jacobian solution. Auclert et al. (2021b) argue that the approximation quality of such method is very good if the horizon is chosen sufficiently long. Also here, we find that the distance between the solution techniques in terms of IRFs is extremely small. Given that we have shown the invariance of the model reduction in our first experiment, we only calculate the distance at the posterior mean.

Concretely, we calculate for both sets of experiments our metric for all observable variables used in the estimation. The metric is calculated based on all shocks using their estimated variances. Columns 2 and 4 show the statistics for the HANK specification, columns 3 and 5 for HANK-X. Here, we report the largest distance over the first 32 quarters. The distances are, as mentioned before, minimal.[27] In Appendix C.4, we also compare graphically the IRFs conditional on a specific shock between sequence-space and state-space solution. All IRFs are extremely close across both techniques also beyond the business cycle horizon.

---

[26]We use this cross-sectional data in the estimation of HANK-X, which also means we cannot directly compare its marginal data density with the two other estimates.

[27]We calculate the sequence-space solution based on a 300 period transition as in Auclert et al. (2021b). We use the IRFs from the state-space solution to obtain the endpoint of that transition in order to eliminate the effect of not being exactly at the stationary equilibrium after 300 periods. We do so because after some of the shocks, the economy has even after 300 periods not fully returned to the stationary equilibrium.

Table 3: IRF variation captured by state-space solution with second-stage model reduction

Observable	relative to no second stage		relative to sequence space	
	HANK	HANK-X	HANK	HANK-X
Output growth	100.00	100.00	100.0	100.00
Investment growth	100.00	100.00	100.00	99.99
Consumption growth	100.00	100.00	99.96	99.97
Hours worked	100.00	100.00	100.00	99.99
Wage growth	100.00	100.00	100.00	100.00
Policy rate	100.00	100.00	99.98	99.99
Inflation	100.00	100.00	99.99	99.99
Top 10 wealth share	–	100.00	–	99.85
Top 10 income share	–	100.00	–	100.00
Tax progressivity	–	100.00	–	100.00
Income risk	–	100.00	–	100.00

Notes: The table displays (in percent) variation of IRFs based on a baseline solution that is captured by our state-space solution with second-stage model reduction. This statistic measures the distance between impulse responses generated by solving the model with different solution techniques and is based on (45). Columns 2 and 3 compare the solution with and without second-stage model reduction, columns 4 and 5 our solution technique to a sequence-space Jacobian solution. Columns 2 and 3 refer to the 1st percentile obtained by drawing 1000 parameter vectors from the posterior distribution. Columns 4 and 5 refer to the posterior mean. All columns report the minimum (i.e., largest distance) over the first 32 quarters. A value of exactly 100 means that impulse responses are identical.

5 US business cycles and inequality

We apply our estimated model to study the shocks that drive the US business cycle and to understand what these shocks and their propagation imply for the dynamics of inequality. We do so in terms of variance decompositions at business cycle frequency (between 6 and 32 quarters, based on the frequency domain decompositions in Uhlig, 2001) and in terms of historical decompositions.[28]

5.1 Variance decompositions

Figure 1 shows the variance decompositions for output and consumption growth (top panels) for the representative agent model (RANK) as well as for the two hetereogeneous agent model estimates (HANK and HANK-X). Further variance decompositions of observables and the credible intervals can be found in Appendices A.4 and A.5. Despite the very similar

---

[28]Appendix A.7 provides impulse response functions for all shocks and observables.

parameter estimates across models, there are some notable differences in the importance of shocks for the business cycle, but also many similarities. In all models, technology shocks are the most important drivers of output. This is in line with what other authors have found for the RANK model (Smets and Wouters, 2007; Justiniano et al., 2010, 2011).

In the HANK model, investment-specific technology shocks are even more important than in RANK. The reason behind this is that these shocks have a stronger impact on consumption. This is because positive investment-specific technology shocks move asset prices down and expected wage incomes up; see the impulse response functions in Appendix A.7. This drives consumption up because wage-earners have a larger (intertemporal) marginal propensity to consume than capital holders. More specifically, within our two-asset framework, there is a sizable share of households with little illiquid assets but sufficient liquidity such that they can increase their consumption on impact in the expectations of higher future wages. By contrast, illiquid-asset-rich households expect lower asset income in the future but they smooth this change over a longer period. As a result, consumption goes up on impact. In RANK, this difference of intertemporal MPCs between wage and capital-income earners is absent, and consumption and investment negatively comove on impact after an investment-specific technology shock. For this reason, investment specific technology shocks add little to consumption movements in RANK.

Comparing the variance decompositions of the HANK and HANK-X estimates, we see that the inclusion of inequality data, tax-progressivity, and income risk also has some impact on the relative contribution of shocks to aggregate consumption (and to a lesser degree on output). For consumption, demand side shocks become somewhat more important. In particular, we see that uncertainty shocks contribute significantly. However, given the relatively low degree of price stickiness and relatively strong monetary stabilization we estimate, the contribution of all demand shocks to output growth is limited.

The HANK models, different to the RANK setting, make it possible to study the effect of business cycle shocks on inequality; see the middle panels of Figure 1. In fact, the HANK-X model also exploits the inequality time-series data for the estimation. Again, we find that technology shocks have an important impact on inequality. However, if we actually use the inequality data for the estimation, the effect of investment-specific technology shocks in income inequality is tuned down and markup shocks become more important. For wealth inequality, there is little difference in the effect of technology shocks between the two models. What is more, as the HANK-X variant also allows tax-progressivity shocks to impact on the economy, they gain a visible influence on wealth inequality.

One advantage of the HANK model is that it allows us to think about drivers of consumption throughout the income and wealth distribution. The bottom panels of Figure 1 give an

Figure 1: Variance decompositions: Output growth, consumption growth, and inequality
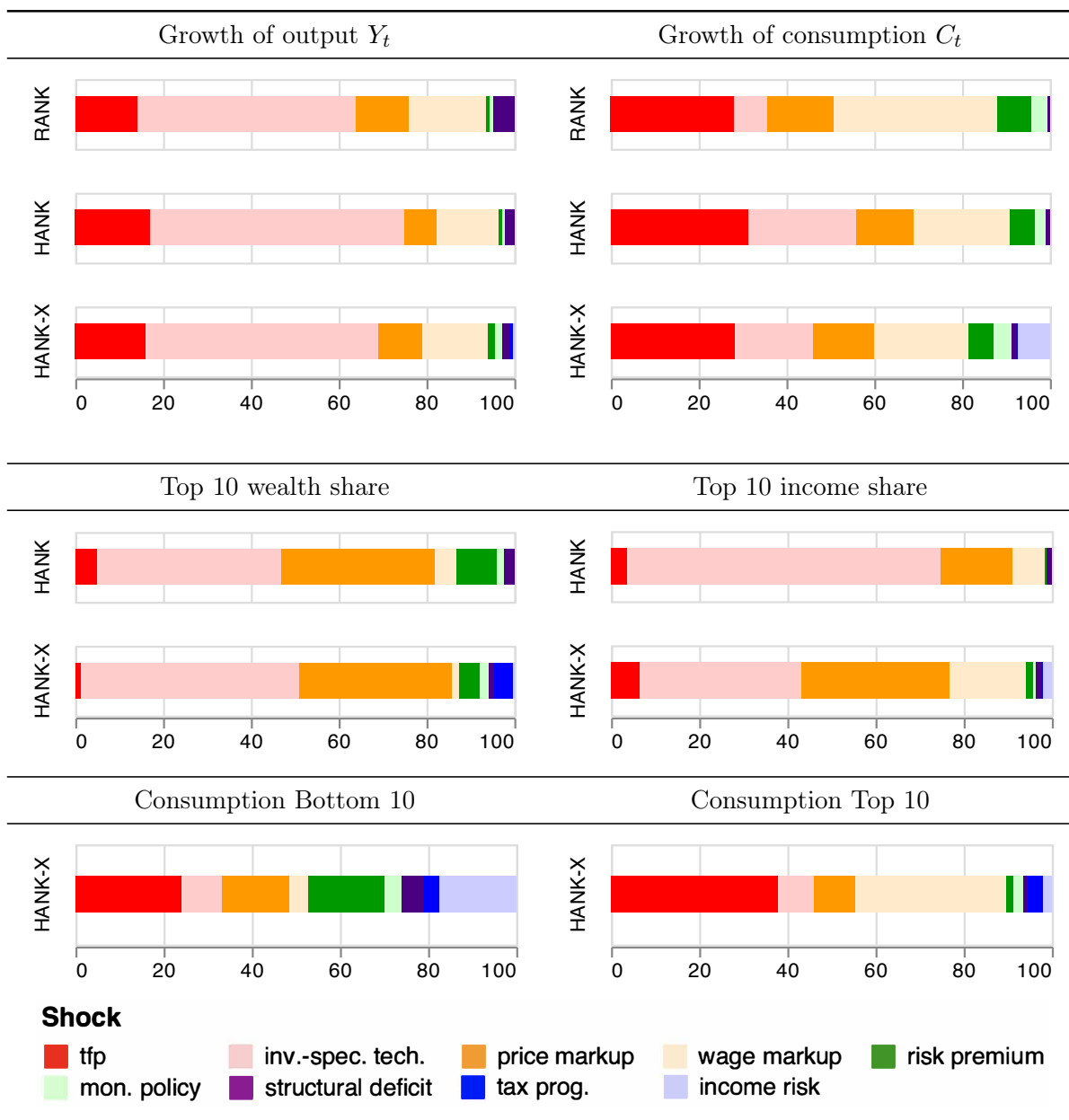
Notes: Conditional variance decompositions at business cycle frequencies (6-32 quarter forecast horizon) for the estimated RANK, HANK, and HANK-X models.

example for this and show the variance decomposition of the average consumption of the 10 percent consumption poorest and consumption richest households. The differences of which shocks are important for these two groups are stark. Consumption of the consumption poor is much more strongly influenced by demand side shocks than consumption of the rich. This suggests that more detailed data on the time series of the consumption distribution might be helpful in identifying business cycle shocks in future work.

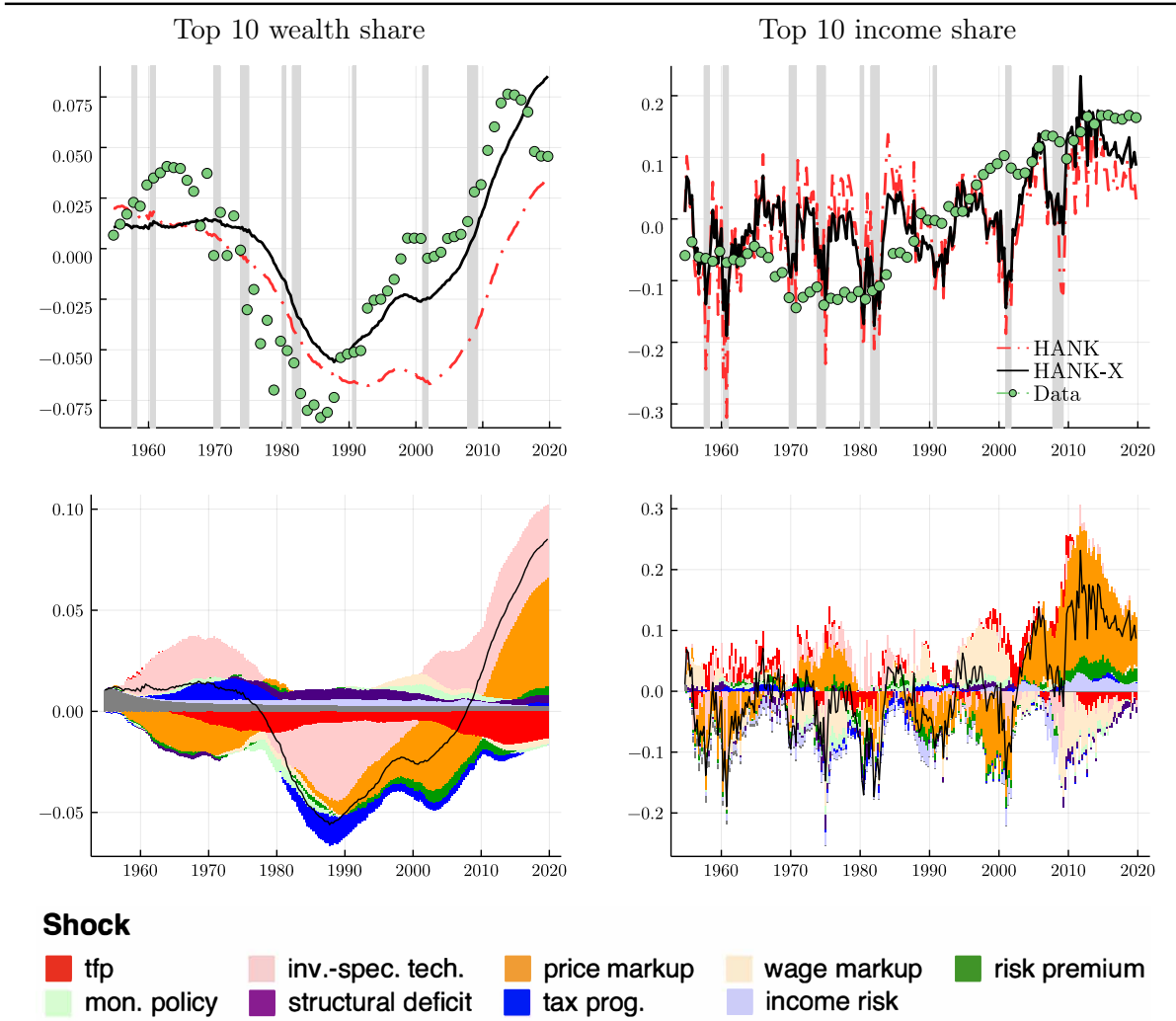## 5.2 Historical decompositions of inequality

Figure 2 adds to these findings by providing historical decompositions for US inequality dynamics. Recall that we allow for measurement error on inequality when estimating the HANK-X model. This implies that neither the HANK nor the HANK-X variant need to match the inequality data perfectly. The top panels of the figure compare the model predicted movements of the top 10 wealth and income shares with the actual data. Perhaps surprisingly, already the HANK estimate implies a falling wealth inequality during the 1970s which recovers (and overshoots) in the 21st century. The HANK-X model follows the observed wealth inequality closely. The model rather changes the shocks and frictions that drive the business cycle slightly instead of giving up on wealth inequality and "explaining" the data by measurement error.[29] In other words, the cross-sectional data is somewhat informative for the business cycle model, but there is no strong tension.

For income inequality, the situation is different. The HANK models imply a fluctuation of the income share of the top 10 percent that is too volatile. This high volatility is driven by large and pro-cyclical swings in profits. Yet, such strong fluctuations are absent in the income inequality data. Surprisingly, however, the model does broadly capture the lower frequency movements in income inequality with its increase since the 1980s. The figure further suggests that persistent increases in price markup targets are behind those low frequency movements in income inequality. This finding resembles the evidence by De Loecker and Eeckhout (2020) on the evolution of markups in the US.[30] Since the cyclical properties of the top 10 income share largely follow the cyclical properties of profits in the model, finding too volatile top incomes is intimately linked to the literature that discusses tensions between model-implied markups/profits and the data (see, e.g., Andreasen and Dang, 2019; Nekarda and Ramey, 2020). The slow moving nature of top income shares suggest through the lens of our model that profit incomes need to be less volatile or less concentrated among the rich. This holds true despite the fact that we allow the model to choose in the estimation how much surprise

---

[29]The estimated standard deviations of measurement errors can be found in Appendix A.3.

[30]There is a growing literature on the rise of markups; see, e.g., Karabarbounis and Neiman (2019), Barkai (2020), Hall (2018), or Kehrig and Vincent (2020).

Figure 2: Historical decompositions: Inequality

Notes: Historical decompositions for the top 10 wealth and income shares for the estimated HANK and HANK-X models. Top panels show the implied series (HANK and HANK-X) obtained through a Kalman smoother in comparison to the data. The two bottom panels show the actual decomposition into shock contributions for the HANK-X estimates. Y-axis: Percent deviation from mean.

changes in profits get distributed to a broader set of households through the market for profit shares. In terms of the HANK-X specific shocks, we see that income uncertainty which drives the dispersion of human capital also has contributed to the increase in income inequality over the last 30 years. This can be viewed as the model's way to capture skill-biased technological change.

When we look at the historical evolution of wealth inequality, we observe that the increase in inequality since its trough in the 1980s is primarily driven by two factors. First, investment-

specific technology shocks that have driven up the price of capital. Since the wealthy hold a larger share of their wealth in terms of the illiquid asset in our model, this drives up wealth inequality and resembles the evidence from household balance sheets (Kuhn et al., 2020). The second most important factor is again the persistent increase in markups that drives up income inequality and thereby affects wealth inequality, too.

5.3 Robustness of estimation results

We check the robustness of our estimation results with respect to a number of potentially important modeling choices—where other alternatives would have been sensible as well. Concretely, we first estimate the model for the post-Volcker era—implicitly assuming a structural break in 1983. Second, we ask how the HANK estimation changes if we assume a risk aversion of 2 (instead of 4). Third, we model the distribution of union profits such that it leads to no wage-compression and the wage markup only affects labor supply but not directly the income distribution. Fourth, we allow (in HANK-X) for a systematic response of tax progressivity to income inequality and, finally, we assume King et al. (1988) preferences instead of Greenwood et al. (1988) ones, such that there is a wealth effect in labor supply. Modeling details, estimated parameters, and decompositions are reported in Appendix B.

We find that the view of the business cycle and inequality dynamics through the lens of the model is relatively robust to these variations, with the exception of the KPR-variant. In the post-Volcker period, markup shocks become somewhat more important, in a sense reflecting their larger impact starting in the 1980s that could be seen already in the historical decompositions for the estimation on the full sample. A lower risk aversion leaves the decomposition of output almost unchanged, but investment shocks become less important for consumption and more important for wealth inequality. The effect of wage compression is negligible throughout, with the exception of income inequality, similarly the estimated systematic response of tax progressivity is weak and thus affects results very little.

While the view of the business cycle is broadly similar between the KPR and GHH variants, marginal data densities clearly prefer the GHH specification both in RANK and HANK.[31] However, the differences become starker when looking at inequality dynamics. Here, the KPR-HANK model does not reproduce the U-shape in wealth inequality and misses the trend in income inequality, which makes the rejection of the KPR assumption by the data even stronger than the marginal data densities already indicate.

---

[31] The (log) marginal data densities for RANK and HANK with KPR preferences are 6555 and 6258, respectively, both lower than those of the models with GHH preferences (see Table 2). That RANK is preferred to HANK by the data when assuming KPR preferences also reflects the rejection of wealth effects on labor supply in the data, which are stronger in HANK.

6 Conclusion

How much does inequality matter for the business cycle and vice versa? To shed light on this two-way relationship, this paper develops a new reduction technique and provides a toolbox to estimate New-Keynesian business cycle models with household heterogeneity and portfolio choice in its state-space representation via Bayesian methods. Concretely, we leverage the fact that we have prior information for model reduction and use this to derive a factor representation of the heterogeneous-agent part of the model. The speed and precision of the proposed solution method is comparable to sequence-space approaches suggested by Auclert et al. (2021b) and Boppart et al. (2018). One advantage of the state-space approach lies in the ample tool set that has been developed for state-space models of the business cycle: for example, variance decomposition at the business cycle frequency or historical decompositions.

Using the same set of aggregate shocks and observables as in Smets and Wouters (2007) in the estimation, we find that heterogeneity in household portfolios gives more precedence to technology shocks in explaining consumption at the expense of markup shocks. They increase wages of poor households with high marginal propensities to consume, which leads to more comovement with investment relative to a setup with representative agent. When including cross-sectional data and shocks in the estimation, shocks to income risk start to play a role as well—especially for consumption of poor households.

The model successfully replicates the dynamics of US wealth inequality such that there is no strong tension between what business cycle analysis suggests as drivers of the cycle and what the model prefers as drivers of wealth inequality. Today's high inequality is a result of technology shocks driving up asset prices and markup shocks driving up profits, in line with empirical evidence from De Loecker and Eeckhout (2020) and Kuhn et al. (2020). However, some tension remains between the aggregate time series and income inequality dynamics when it comes to markup shocks. The model predicts that profit incomes and thereby top incomes are too volatile because fitting the aggregates requires large swings in markups in relative terms. This leaves room for future research using inequality data to inform researchers better about the calibration and estimation of the markup process. What is more, we show that different business cycle shocks impact the percentiles of the consumption distribution differently such that using detailed data on the consumption distribution dynamics might also help to tighten the identification of business cycle shocks in future research.

References

Acharya, S., Chen, W., Negro, M. D., Dogra, K., Matlin, E., and Sarfati, R. (2021). Estimating HANK: Macro time series and micro moments. Presentation, ASSA Meeting 2021.

Ahn, S., Kaplan, G., Moll, B., Winberry, T., and Wolf, C. (2018). When inequality matters for macro and macro matters for inequality. *NBER Macroeconomics Annual*, 32(1):1–75.

An, S. and Schorfheide, F. (2007). Bayesian analysis of DSGE models. *Econometric Reviews*, 26(2-4):113–172.

Andreasen, M. M. and Dang, M. (2019). Estimating the price markup in the New Keynesian model. CREATES Research Papers 2019-03.

Auclert, A. (2019). Monetary policy and the redistribution channel. *American Economic Review*, 109(6):2333–67.

Auclert, A., Bardóczy, B., and Rognlie, M. (2021a). MPCs, MPEs, and multipliers: A trilemma for New Keynesian models. *Review of Economics and Statistics*, pages 1–41.

Auclert, A., Bardóczy, B., Rognlie, M., and Straub, L. (2021b). Using the sequence-space Jacobian to solve and estimate heterogeneous-agent models. *Econometrica*, 89(5):2375–2408.

Auclert, A., Rognlie, M., and Straub, L. (2018). The intertemporal Keynesian Cross. NBER Working Paper 25020.

Auclert, A., Rognlie, M., and Straub, L. (2020). Micro jumps, macro humps: Monetary policy and business cycles in an estimated HANK model. NBER Working Paper 26647.

Barkai, S. (2020). Declining labor and capital shares. *Journal of Finance*, 75(5):2421–2463.

Bayer, C., Born, B., and Luetticke, R. (2022). The liquidity channel of fiscal policy. CEPR Discussion Paper 14883.

Bayer, C. and Luetticke, R. (2020). Solving heterogeneous agent models in discrete time with many idiosyncratic states by perturbation methods. *Quantitative Economics*, 11:1253–1288.

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.

Benabou, R. (2002). Tax and education policy in a heterogeneous-agent economy: What levels of redistribution maximize growth and efficiency? *Econometrica*, 70(2):481–517.

Berger, D., Bocola, L., and Dovis, A. (2022). Imperfect risk sharing and the business cycle. Technical report, Stanford University.

Bilbiie, F., Primiceri, G. E., and Tambalotti, A. (2022). Inequality and business cycles. Mimeo, Cambridge University.

Boar, C. and Midrigan, V. (2019). Markups and inequality. NBER Working Paper 25952.

Boppart, T., Krusell, P., and Mitman, K. (2018). Exploiting MIT shocks in heterogeneous-agent economies: The impulse response as a numerical derivative. *Journal of Economic Dynamics and Control*, 89:68–92.

Born, B. and Pfeifer, J. (2014). Policy risk and the business cycle. *Journal of Monetary Economics*, 68:68–85.

Bricker, J., Henriques, A., Krimmel, J., and Sabelhaus, J. (2016). Measuring income and wealth at the top using administrative and survey data. *Brookings Papers on Economic Activity*, 2016(1):261–331.

Broer, T., Hansen, N.-J. H., Krusell, P., and Öberg, E. (2019). The New Keynesian transmission mechanism: A heterogenous-agent perspective. *Review of Economic Studies*, 87(1):77–101.

Calvo, G. A. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, 12(3):383–398.

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

Castaneda, A., Diaz-Gimenez, J., and Rios-Rull, J.-V. (1998). Exploring the income distribution business cycle dynamics. *Journal of Monetary Economics*, 42(1):93 – 130.

Challe, E. and Ragot, X. (2015). Precautionary saving over the business cycle. *Economic Journal*, 126(590):135–164.

Chang, M., Chen, X., and Schorfheide, F. (2021). Heterogeneity and aggregate fluctuations. NBER Working Paper 28853.

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):471–475.

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.

De Loecker, J. and Eeckhout, J. (2020). The rise of market power and the macroeconomic implications. *Quarterly Journal of Economics*, 135(2):561–644.

Den Haan, W. J., Rendahl, P., and Riegler, M. (2017). Unemployment (fears) and deflationary spirals. *Journal of the European Economic Association*, 16(5):1281–1349.

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. Oxford University Press.

Fernández-Villaverde, J. (2010). The econometrics of DSGE models. *SERIEs*, 1(1):3–49.

Ferriere, A. and Navarro, G. (2018). The heterogeneous effects of government spending: It's all about taxes. *FRB International Finance Discussion Paper*, (1237).

Gabaix, X., Lasry, J.-M., Lions, P.-L., and Moll, B. (2016). The dynamics of inequality. *Econometrica*, 84(6):2071–2111.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics*, volume 4, pages 641–649. Clarendon Press, Oxford.

Gornemann, N., Kuester, K., and Nakajima, M. (2012). Monetary policy with heterogeneous agents. *Philadelphia Fed Working Paper 12-21*.

Greenwood, J., Hercowitz, Z., and Huffman, G. W. (1988). Investment, capacity utilization, and the real business cycle. *American Economic Review*, 78(3):402–417.

Guerrieri, V. and Lorenzoni, G. (2017). Credit crises, precautionary savings, and the liquidity trap. *Quarterly Journal of Economics*, 132(3):1427–1467.

Guvenen, F., Kaplan, G., and Song, J. (2014). How risky are recessions for top earners? *American Economic Review*, 104(5):148–153.

Hagedorn, M., Manovskii, I., and Mitman, K. (2018). Monetary policy in incomplete market models: Theory and evidence. Mimeo, University of Pennsylvania.

Hagedorn, M., Manovskii, I., and Mitman, K. (2019). The fiscal multiplier. NBER Working Paper 25571.

Hall, R. E. (2018). New evidence on the markup of prices over marginal costs and the role of mega-firms in the US economy. NBER Working Paper 24574.

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.

Heathcote, J., Storesletten, K., and Violante, G. L. (2017). Optimal tax progressivity: An analytical framework. *Quarterly Journal of Economics*, 132(4):1693–1754.

Herbst, E. and Schorfheide, F. (2014). Sequential Monte Carlo sampling for DSGE models. *Journal of Applied Econometrics*, 29(7):1073–1098.

Herbst, E. and Schorfheide, F. (2015). *Bayesian Estimation of DSGE Models*. Princeton University Press.

Hintermaier, T. and Koeniger, W. (2010). The method of endogenous gridpoints with occasionally binding constraints among endogenous variables. *Journal of Economic Dynamics and Control*, 34(10):2074–2088.

Hubmer, J., Krusell, P., and Smith Jr, A. A. (2020). Sources of U.S. wealth inequality: Past, present, and future. *NBER Macroeconomics Annual*, 35.

Jaimovich, N. and Rebelo, S. (2009). Can news about the future drive the business cycle? *American Economic Review*, 99(4):1097–1118.

Justiniano, A., Primiceri, G. E., and Tambalotti, A. (2010). Investment shocks and business cycles. *Journal of Monetary Economics*, 57(2):132 – 145.

Justiniano, A., Primiceri, G. E., and Tambalotti, A. (2011). Investment shocks and the relative price of investment. *Review of Economic Dynamics*, 14(1):101–121.

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

Kaplan, G. and Violante, G. L. (2014). A model of the consumption response to fiscal stimulus payments. *Econometrica*, 82:1199–1239.

Karabarbounis, L. and Neiman, B. (2019). Accounting for factorless income. *NBER Macroeconomics Annual*, 33(1):167–228.

Kaymak, B. and Poschke, M. (2016). The evolution of wealth inequality over half a century: The role of taxes, transfers and technology. *Journal of Monetary Economics*, 77:1–25.

Kehrig, M. and Vincent, N. (2020). The micro-level anatomy of the labor share decline. *Quarterly Journal of Economics*, 136(2):1031–1087.

King, R. G., Plosser, C. I., and Rebelo, S. T. (1988). Production, growth and business cycles: I. The basic neoclassical model. *Journal of Monetary Economics*, 21(2):195 – 232.

Klein, P. (2000). Using the generalized Schur form to solve a multivariate linear rational expectations model. *Journal of Economic Dynamics and Control*, 24(10):1405–1423.

Koopman, S. J. and Durbin, J. (2000). Fast filtering and smoothing for multivariate state space models. *Journal of Time Series Analysis*, 21(3):281–296.

Kopczuk, W. (2015). What do we know about the evolution of top wealth shares in the United States? *Journal of Economic Perspectives*, 29(1):47–66.

Kuhn, M., Schularick, M., and Steins, U. I. (2020). Income and wealth inequality in America, 1949-2013. *Journal of Political Economy*, 128(9):3469–3519.

Liu, L. and Plagborg-Møller, M. (2022). Full-information estimation of heterogeneous agent models using macro and micro data. *Quantitative Economics*, forthcoming.

Luetticke, R. (2021). Transmission of monetary policy with heterogeneity in household portfolios. *American Economic Journal: Macroeconomics*, 13(2):1–25.

McKay, A., Nakamura, E., and Steinsson, J. (2016). The power of forward guidance revisited. *American Economic Review*, 106(10):3133–3158.

McKay, A. and Reis, R. (2016). The role of automatic stabilizers in the US business cycle. *Econometrica*, 84(1):141–194.

McKay, A. and Reis, R. (2021). Optimal automatic stabilizers. *Review of Economic Studies*, 88(5):2375—2406.

Mertens, K. and Montiel Olea, J. L. (2018). Marginal tax rates and income: New time series evidence. *Quarterly Journal of Economics*, 133(4):1803–1884.

Nekarda, C. J. and Ramey, V. A. (2020). The cyclical behavior of the price-cost markup. *Journal of Money, Credit and Banking*, 52(S2):319–353.

Piketty, T. and Saez, E. (2003). Income inequality in the United States, 1913–1998. *Quarterly Journal of Economics*, 118(1):1–41.

Ravn, M. O. and Sterk, V. (2017). Job uncertainty and deep recessions. *Journal of Monetary Economics*, 90:125–141.

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

Saez, E. (2004). Reported incomes and marginal tax rates, 1960-2000: Evidence and policy implications. *Tax policy and the Economy*, 18:117–173.

Saez, E. and Zucman, G. (2016). Wealth inequality in the United States since 1913: Evidence from capitalized income tax data. *Quarterly Journal of Economics*, 131(2):519–578.

Schmitt-Grohé, S. and Uribe, M. (2004). Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of Economic Dynamics and Control*, 28(4):755–775.

Schmitt-Grohé, S. and Uribe, M. (2012). What's news in business cycles. *Econometrica*, 80(6):2733–2764.

Smets, F. and Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review*, 97(3):586–606.

Sterk, V. and Tenreyro, S. (2018). The transmission of monetary policy through redistributions and durable purchases. *Journal of Monetary Economics*, 99:124–137.

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

Taylor, J. B. (1993). Discretion versus policy rules in practice. In *Carnegie-Rochester Conference Series on Public Policy*, volume 39, pages 195–214. Elsevier.

Uhlig, H. (2001). A toolkit for analysing nonlinear dynamic stochastic models easily. In Marimon, R. and Scott, A., editors, *Computational Methods for the Study of Dynamic Economies*, chapter 3, pages 30–61. Oxford University Press.

Wong, A. (2019). Refinancing and the transmission of monetary policy to consumption. Mimeo, Princeton University.

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

Wu, J. C. and Xia, F. D. (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit and Banking*, 48(2-3):253–291.

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.

A Data and parameterization

In this appendix, we first list the data sources and transformations in Appendix A.1 that we employ in order to calibrate the parameters affecting the stationary distribution and to estimate via Bayesian methods those parameters that do not. Appendix A.2 then discusses the re-parameterization of the RANK model steady state. In Appendix A.3, we present the posterior estimates of the structural shock processes. Appendix A.4 contains the variance decompositions of observables not shown in the main text. In Appendix A.5, we provide the credible intervals for all variance decompositions of observables. Appendix A.6 contains historical decompositions of observables and further variables of interest based on the HANK and HANK-X models, and Appendix A.7 provides IRFs to all structural shocks for RANK, HANK, and HANK-X. Finally, Appendix A.8 provides convergence diagnostics for the MCMC chains.

A.1 Data: Sources and transformations

A.1.1 Data for calibration

The following list contains the data sources for the average data ratios we target in the calibration of the stationary equilibrium:

Mean illiquid assets. Fixed assets (NIPA table 1.1) over quarterly GDP (excluding net exports; see below), averaged over 1954 – 2019.

Mean liquidity. Liquid assets over illiquid assets, where liquid assets comprise the sum of checking, savings and call/money market accounts, as well as holdings in mutual funds, equity and other managed assets, and bonds other than saving bonds from the Survey of Consumer Finances. We use the SCF+, by Kuhn et al. (2020), which yields 20 years of cross-sectional data between 1950 and 2016.

Mean government debt. Gross federal debt held by the public as percent of GDP (FYPUGDA188S). Available from 1954 – 2019.

Fraction of borrowers. Taken from the Survey of Consumer Finances (1983 – 2013); see Bayer et al. (2019) for more details.

Average top 10 share of wealth. Source is the World Inequality Database (1954 – 2019).

A.1.2 Data for estimation

Formally, the vector of observable variables is given by:

$$
OBS_t =
\begin{bmatrix}
\Delta \log (Y_t) \\
\Delta \log (C_t) \\
\Delta \log (I_t) \\
\Delta \log (w_t^F) \\
\log (N_t) \\
\log (R_t^b) \\
\log (\pi_t) \\
\log (T10WShare_t) \\
\log (T10IShare_t) \\
\log (s_t) \\
\log (\tau_t^P)
\end{bmatrix}
-
\begin{bmatrix}
\overline{\Delta \log (Y_t)} \\
\overline{\Delta \log (C_t)} \\
\overline{\Delta \log (I_t)} \\
\overline{\Delta \log (w_t^F)} \\
\overline{\log (N_t)} \\
\overline{\log (R_t^b)} \\
\overline{\log (\pi_t)} \\
\overline{\log (T10WShare_t)} \\
\overline{\log (T10IShare_t)} \\
\overline{\log (s_t)} \\
\overline{\log (\tau_t^P)}
\end{bmatrix}
$$

where Δ denotes the temporal difference operator and bars above variables denote time-series averages.

Unless otherwise noted, all series are available at quarterly frequency from 1954Q3 to 2019Q4 from the St.Louis FED - FRED database (mnemonics in parentheses).

Output, Y_t. Sum of gross private domestic investment (GPDI), personal consumption expenditures for nondurable goods (PCND), durable goods (PCDG), and services (PCESV), and government consumption expenditures and gross investment (GCE) divided by the GDP deflator (GDPDEF) and the civilian noninstitutional population (CNP16OV).

Consumption, C_t. Sum of personal consumption expenditures for nondurable goods (PCND), durable goods (PCDG), and services (PCESV) divided by the GDP deflator (GDPDEF) and the civilian noninstitutional population (CNP16OV).

Investment, I_t. Gross private domestic investment (GPDI) divided by the GDP deflator (GDPDEF) and the civilian noninstitutional population (CNP16OV).

Real wage, w_t^F. Hourly compensation in the nonfarm business sector (COMPNFB) divided by the GDP deflator (GDPDEF).

Hours worked, N_t. Nonfarm business hours worked (COMPNFB) divided by the civilian noninstitutional population (CNP16OV).

Inflation, π_t. Computed as the log-difference of the GDP deflator (GDPDEF).

Nominal interest rate, R_t^b. Quarterly average of the effective federal funds rate (FED-FUNDS). From 2009Q1 to 2015Q4, we use the Wu and Xia (2016) shadow federal funds rate.

Wealth inequality, $T10WShare_t$. p90p100 of US net personal wealth from the World Inequality Database. Available annually 1954 to 2019.

Income inequality, $T10IShare_t$. p90p100 of US pre-tax national income from the World Inequality Database. Available annually 1954 to 2019.

Idiosyncratic income risk, s_t. We take the estimated time series for the variance of idiosyncratic income from Bayer et al. (2019) who use the Survey of Income and Program Participation. Available from 1983Q1 to 2013Q1.

Tax progressivity, τ_t^P. We follow Ferriere and Navarro (2018) and construct our measure of tax progressivity using the average and average marginal tax rate: P = (AMTR - ATR)/(1 - ATR). For a loglinear tax system, this measure equals the parameter capturing the curvature of the tax function. Available annually 1954 to 2017.

Details on the construction of the tax-progressivity measure

We extend the Mertens and Montiel Olea (2018)-calculations of average (ATR) and average marginal tax rates (AMTR) to the years 2013-2017. First, in constructing the ATR series, we obtain total tax liabilities for 1929-2017, from the National Income and Product Accounts (NIPA), *Table 3.2*. Federal social insurance contributions, which are added to total tax liability, come from NIPA, *Table 3.6*, line 3 and 21. For total income, we take Piketty and Saez (2003)'s income series, which uses a broader income concept based on adjusted gross income, excluding taxable social security and unemployment insurance benefits.

The AMTR is the sum of the average marginal individual income tax rate (AMIITR) and the average marginal payroll tax rate (AMPRT). We follow Ferriere and Navarro (2018) and use Saez (2004)'s income concept.[32] This income concept includes all income items reported on an individual's tax return before deductions and excluding capital gains. Income items include salaries and wages, small business/farm income, partnership and fiduciary income, dividends, interest, rents, royalties and other small items reported as other income. Realized capital gains are excluded in this measure of income.

---

[32] For a detailed explanation on the construction of the AMTRs; see Appendix A of Mertens and Montiel Olea (2018). We follow *method 1* for computing the AMIITRs.

To construct the AMTR, we first use several tables from the Statistics of Income (SOI) to construct the discrete distributions of adjusted gross income by income brackets needed for the AMIITR. *Table 1.1 All Returns* of the SOI archives contains information on number of returns, adjusted gross income (AGI), and taxable income for different ranges of AGI per return. These ranges define the discretization. Given the distribution is fit for every year and by filing status, *Table 1.2 All Returns: by Marital Status* provides the equivalent table distinguishing by filing status, e.g., married filing jointly or separately, head of household, single, and surviving spouse. *Table 1.3 All Returns: Sources of Income* provides information on how many of these returns reported income from salaries and wages. *Table 1.4 All Returns: Sources of Income, Adjustments, and Tax Items* contains data on taxable income and number of corresponding returns by bracket. *Table 3.3 All Returns: Tax Liability, Tax Credits, and Tax Payments* provides information on how many filed for self-employment and their tax liability. Finally, *Table 3.4* contains the number of returns and adjusted gross income by marginal tax bracket and filing status using.

To construct the Average Marginal Payroll Tax Rate (AMPTR), we collect data from the 2019 Annual Statistical supplement, *Table 2.A3* (columns 1, 2, 3 and 9), to obtain the taxation of labor and self-employed earnings under the Old Age, Survivors and Disability Insurance (OASDI) and Hospital Insurance (HI) programs. The columns respectively cover the number of covered workers and self employed with maximum earnings as well as total taxable earnings. Their difference allows us to calculate the total taxable earnings of covered workers with earnings *below* the maximum. Information on earnings can be found in *Table 4.B* from the same source.

A.2 RANK calibration

Table A.1 shows the steady-state parameterization of the representative-agent analogue of the HANK model. We adjust the discount factor to match a capital-to-output ratio of 11.44 (quarterly) and the level of the tax rate to match the ratio of government-spending-to-output (0.2). All other parameters are externally chosen and equal to the parameterizaton of the HANK model.

Table A.1: External/calibrated parameters in RANK (quarterly frequency)

Parameter	Value	Description	Target
Households			
β	0.996	Discount factor	K/Y=11.44
ξ	4.000	Relative risk aversion	Kaplan et al. (2018)
γ	2.000	Inverse of Frisch elasticity	Chetty et al. (2011)
Firms			
α	0.680	Share of labor	62% labor income
δ_0	0.018	Depreciation rate	7.0% p.a.
$\bar{\eta}$	11.000	Elasticity of substitution	Price markup 10%
$\bar{\zeta}$	11.000	Elasticity of substitution	Wage markup 10%
Government			
$\bar{\tau}^L$	0.250	Tax rate level	$G/Y = 0.2$
$\bar{\tau}^P$	0.120	Tax progressivity	SoI 1954 - 2019
$\bar{R}^b$	1.000	Nominal rate	Growth $\approx$ interest rate
$\bar{\pi}$	1.000	Inflation	Indexation, w.l.o.g.

A.3 Estimated structural shock processes

Table A.2 presents prior and posterior distributions of the estimated shock processes. The RANK and HANK version only include seven standard aggregate shocks, while the HANK-X version also includes shocks to income risk and tax progressivity.

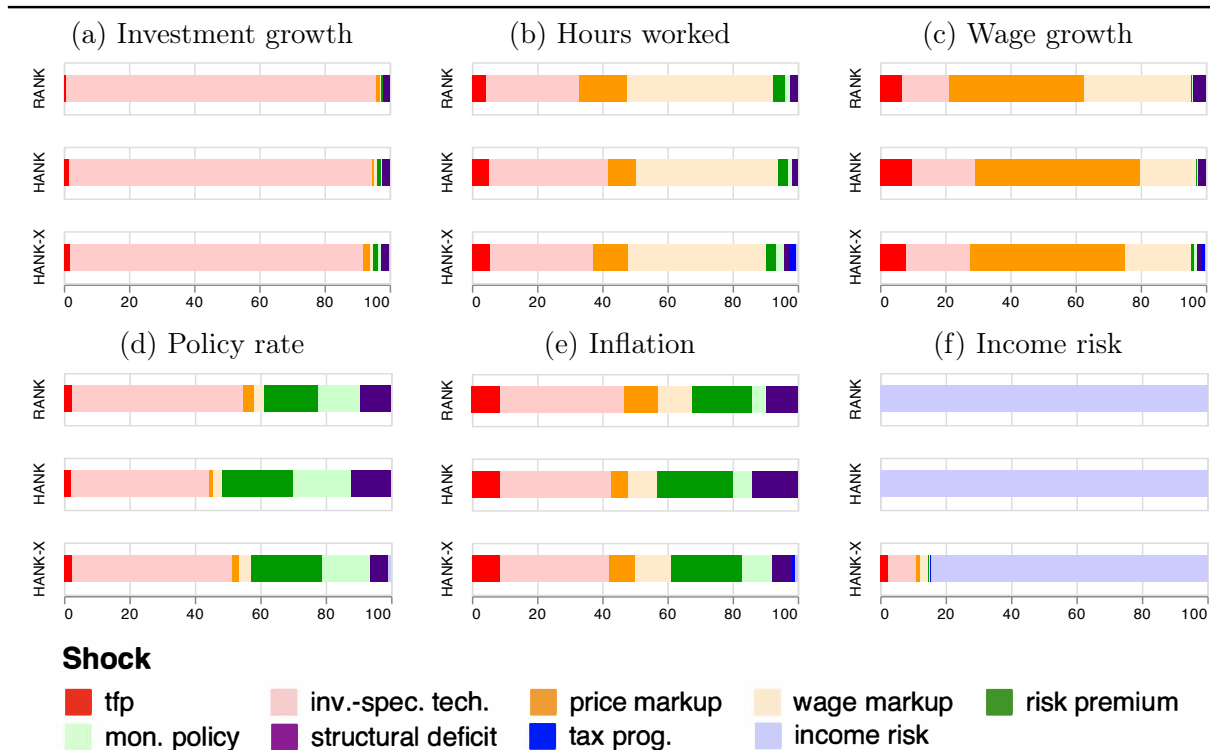**Table A.2:** Prior and posterior distributions of estimated shocks and measurement errors

Parameter	Distribution	Prior		Posterior		
		Mean	Std. Dev.	RANK	HANK	HANK-X
Structural Shocks						
ρ_A	Beta	0.50	0.20	0.943	0.943	0.972
				(0.917, 0.969)	(0.913, 0.970)	(0.947, 0.992)
σ_A	Inv.-Gamma	0.10	2.00	0.223	0.223	0.158
				(0.178, 0.272)	(0.170, 0.275)	(0.125, 0.193)
ρ_Z	Beta	0.50	0.20	0.996	0.996	0.998
				(0.994, 0.997)	(0.994, 0.998)	(0.996, 0.999)
σ_Z	Inv.-Gamma	0.10	2.00	0.579	0.628	0.601
				(0.529, 0.632)	(0.574, 0.685)	(0.552, 0.655)
ρ_Ψ	Beta	0.50	0.20	0.722	0.663	0.764
				(0.669, 0.773)	(0.603, 0.721)	(0.707, 0.817)
σ_Ψ	Inv.-Gamma	0.10	2.00	16.722	12.635	7.211
				(13.055, 20.65)	(10.465, 15.045)	(6.575, 7.924)
ρ_μ	Beta	0.50	0.20	0.963	0.906	0.904
				(0.933, 0.988)	(0.881, 0.928)	(0.877, 0.928)
σ_μ	Inv.-Gamma	0.10	2.00	1.281	1.225	1.351
				(1.118, 1.481)	(1.083, 1.392)	(1.184, 1.545)
$\rho_{\mu w}$	Beta	0.50	0.20	0.892	0.912	0.906
				(0.851, 0.927)	(0.887, 0.932)	(0.878, 0.929)
$\sigma_{\mu w}$	Inv.-Gamma	0.10	2.00	3.590	3.285	3.506
				(3.038, 4.301)	(2.871, 3.787)	(3.008, 4.098)
σ_D	Inv.-Gamma	0.10	2.00	0.534	0.522	0.376
				(0.451, 0.626)	(0.434, 0.615)	(0.323, 0.433)
ρ_P	Beta	0.50	0.20	—	—	0.919
				(—, —)	(—, —)	(0.884, 0.950)
σ_P	Inv.-Gamma	0.10	2.00	—	—	6.865
				(—, —)	(—, —)	(5.839, 8.102)
σ_s	Gamma	65.00	30.00	—	—	69.186
				(—, —)	(—, —)	(61.41, 77.882)
Measurement Errors						
σ_{I10}^{me}	Inv.-Gamma	0.05	0.01	—	—	2.418
				(—, —)	(—, —)	(2.057, 2.825)
σ_{W10}^{me}	Inv.-Gamma	0.05	0.01	—	—	7.616
				(—, —)	(—, —)	(6.454, 8.951)

Notes: The table displays the estimated shock processes and measurement errors, their priors and posterior means across three model variants: RANK, HANK, and HANK-X. The 90% credible intervals are shown in parentheses. Posteriors are obtained by an MCMC method. The standard deviations have been multiplied by 100 for better readability.

A.4 Variance decompositions of further observables

Figure A.1 shows the variance decomposition of all observables not shown in the main text for the estimated models.

Figure A.1: Variance decompositions of further observables

Table A.3 shows the credible intervals of all shown variance decomposition of for the RANK, the HANK, and the HANK-X model. The credible intervals are obtained by sampling 1000 times from the posterior.

**Table A.3:** Variance decompositions with credible intervals

	tfp	inv.-spec. tech.	price markup	wage markup	risk premium	mon. policy	structural deficit	tax progr.	income risk
					RANK				
output growth	14.3	49.6	12.2	17.4	0.9	0.8	4.8	–	–
	(11.3, 17.2)	(45.1, 54.8)	(9.7, 15.3)	(13.6, 21.0)	(0.4, 1.9)	(0.5, 1.3)	(3.5, 6.3)		
consumption growth	28.0	7.6	15.3	37.2	7.6	3.6	0.6	–	–
	(23.4, 32.8)	(5.9, 10.1)	(12.5, 18.6)	(31.4, 41.7)	(5.9, 10.2)	(2.7, 5.2)	(0.4, 1.0)		
investment growth	0.8	95.1	1.2	0.4	0.5	0.1	1.9	–	–
	(0.5, 1.2)	(92.9, 96.5)	(0.7, 2.2)	(0.2, 0.7)	(0.3, 1.0)	(0.1, 0.2)	(1.4, 2.5)		
employment	4.3	28.8	14.7	44.9	3.4	1.5	2.4	–	–
	(3.4, 5.4)	(24.1, 35.0)	(11.7, 18.2)	(37.9, 50.2)	(2.4, 5.0)	(1.1, 2.3)	(1.9, 3.1)		
wage growth	6.6	14.7	41.1	33.1	0.4	0.1	3.9	–	–
	(4.9, 8.8)	(12.1, 19.0)	(34.9, 46.7)	(26.4, 38.6)	(0.2, 1.0)	(0.1, 0.5)	(2.7, 5.3)		
nominal rate	2.6	52.0	3.6	3.1	16.3	13.1	9.3	–	–
	(1.6, 3.7)	(44.3, 59.2)	(2.2, 5.9)	(2.0, 4.3)	(12.8, 20.6)	(10.4, 16.6)	(5.9, 12.7)		
inflation	8.6	38.0	10.6	10.4	18.3	4.3	9.8	–	–
	(6.5, 10.9)	(30.4, 45.6)	(7.1, 15.7)	(7.7, 12.8)	(15.1, 22.3)	(3.2, 5.7)	(5.9, 14.0)		
					HANK				
output growth	17.1	57.9	7.3	14.2	0.8	0.6	2.1	–	–
	(14.4, 20.3)	(52.9, 62.5)	(5.8, 9.2)	(11.5, 16.8)	(0.4, 1.6)	(0.5, 0.9)	(1.4, 2.9)		
consumption growth	31.5	24.4	13.1	21.9	5.8	2.3	0.9	–	–
	(27.1, 36.6)	(19.8, 28.8)	(10.8, 15.6)	(17.8, 25.7)	(4.4, 7.5)	(1.7, 3.2)	(0.7, 1.6)		
investment growth	1.7	92.9	0.9	0.7	1.3	0.3	2.3	–	–
	(1.3, 2.3)	(91.0, 94.3)	(0.6, 1.3)	(0.5, 0.9)	(0.7, 2.0)	(0.2, 0.5)	(1.7, 3.0)		
employment	5.5	36.5	8.4	43.5	3.1	1.2	1.7	–	–
	(4.5, 6.9)	(30.7, 42.8)	(6.7, 10.4)	(37.1, 49.4)	(2.0, 4.6)	(0.9, 1.7)	(1.2, 2.4)		
wage growth	9.9	19.2	50.9	17.1	0.5	0.1	2.3	–	–
	(7.7, 12.4)	(16.0, 24.0)	(45.0, 55.3)	(13.6, 21.3)	(0.2, 1.0)	(0.1, 0.4)	(1.5, 3.3)		
nominal rate	2.4	42.0	1.4	2.5	22.0	17.7	12.0	–	–
	(1.4, 3.5)	(34.5, 50.0)	(0.9, 2.1)	(1.6, 3.6)	(17.5, 26.9)	(13.8, 22.0)	(7.0, 17.0)		
inflation	8.8	34.0	5.3	8.9	23.2	5.7	14.1	–	–
	(6.6, 11.0)	(27.9, 40.7)	(3.8, 7.2)	(6.7, 11.3)	(18.8, 28.1)	(4.3, 7.5)	(8.3, 19.9)		
					HANK-X				
output growth	16.0	53.0	10.1	14.8	1.6	1.6	1.8	0.6	0.4
	(13.2, 19.2)	(48.3, 58.0)	(8.2, 12.5)	(11.7, 17.7)	(1.0, 2.7)	(1.2, 2.2)	(1.2, 2.6)	(0.4, 0.9)	(0.3, 0.6)
consumption growth	28.3	17.8	14.0	21.4	5.7	4.1	1.0	0.6	7.2
	(23.8, 33.1)	(15.1, 21.0)	(11.6, 16.4)	(17.2, 24.8)	(4.5, 7.6)	(3.3, 5.2)	(0.8, 1.7)	(0.4, 0.9)	(5.6, 9.1)
investment growth	2.1	89.9	1.9	1.2	1.6	0.7	2.5	0.1	0.1
	(1.6, 2.7)	(88.0, 91.5)	(1.4, 2.5)	(0.9, 1.6)	(1.0, 2.2)	(0.5, 1.0)	(1.9, 3.1)	(0.0, 0.1)	(0.1, 0.1)
employment	5.7	31.4	10.7	42.5	3.1	2.5	1.6	1.8	0.6
	(4.7, 7.0)	(27.3, 36.4)	(8.8, 13.1)	(36.7, 46.8)	(2.2, 4.4)	(2.0, 3.2)	(1.1, 2.2)	(1.3, 2.6)	(0.4, 0.8)
wage growth	8.1	19.7	47.6	20.3	0.9	0.8	1.5	1.0	0.2
	(6.4, 10.3)	(16.8, 23.5)	(42.6, 51.7)	(16.3, 24.5)	(0.5, 1.7)	(0.4, 1.3)	(0.9, 2.2)	(0.7, 1.4)	(0.1, 0.3)
nominal rate	2.5	48.8	2.4	3.5	21.7	14.6	5.5	0.3	0.7
	(1.7, 3.5)	(41.3, 55.7)	(1.6, 3.4)	(2.4, 4.7)	(17.8, 26.3)	(11.6, 18.2)	(2.8, 8.7)	(0.2, 0.4)	(0.5, 1.0)
inflation	8.8	33.2	8.1	11.1	21.9	8.9	6.4	0.7	0.9
	(6.8, 11.0)	(27.3, 39.4)	(5.9, 10.6)	(8.6, 13.6)	(18.2, 26.4)	(7.2, 11.0)	(3.4, 9.8)	(0.5, 1.0)	(0.6, 1.2)
uncertainty	2.5	8.3	1.5	2.4	0.2	0.2	0.3	0.1	84.5
	(1.9, 3.4)	(6.2, 10.9)	(1.0, 2.1)	(1.8, 3.2)	(0.1, 0.4)	(0.1, 0.3)	(0.2, 0.4)	(0.1, 0.1)	(80.3, 87.8)
tax progressivity	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0)	(100.0, 100.0)	(0.0, 0.0)
T10 wealth share	1.4	49.6	34.7	1.8	4.6	2.2	1.3	4.1	0.3
	(1.0, 2.0)	(42.2, 56.3)	(29.0, 40.8)	(0.8, 3.0)	(3.2, 6.9)	(1.7, 2.8)	(0.8, 1.8)	(2.7, 5.9)	(0.2, 0.5)
T10 income share	6.4	36.8	33.6	17.4	1.5	0.6	1.2	0.5	1.9
	(5.2, 7.8)	(30.0, 44.0)	(28.1, 38.8)	(13.9, 21.7)	(1.0, 2.5)	(0.4, 0.9)	(1.0, 1.7)	(0.4, 0.8)	(1.4, 2.5)

Notes: The table displays variance decompositions at business cycle frequencies and their (5,95)-credible intervals for all observables and shocks in the RANK, HANK, and HANK-X models. The credible intervals are obtained by sampling 1000 times from the posterior.

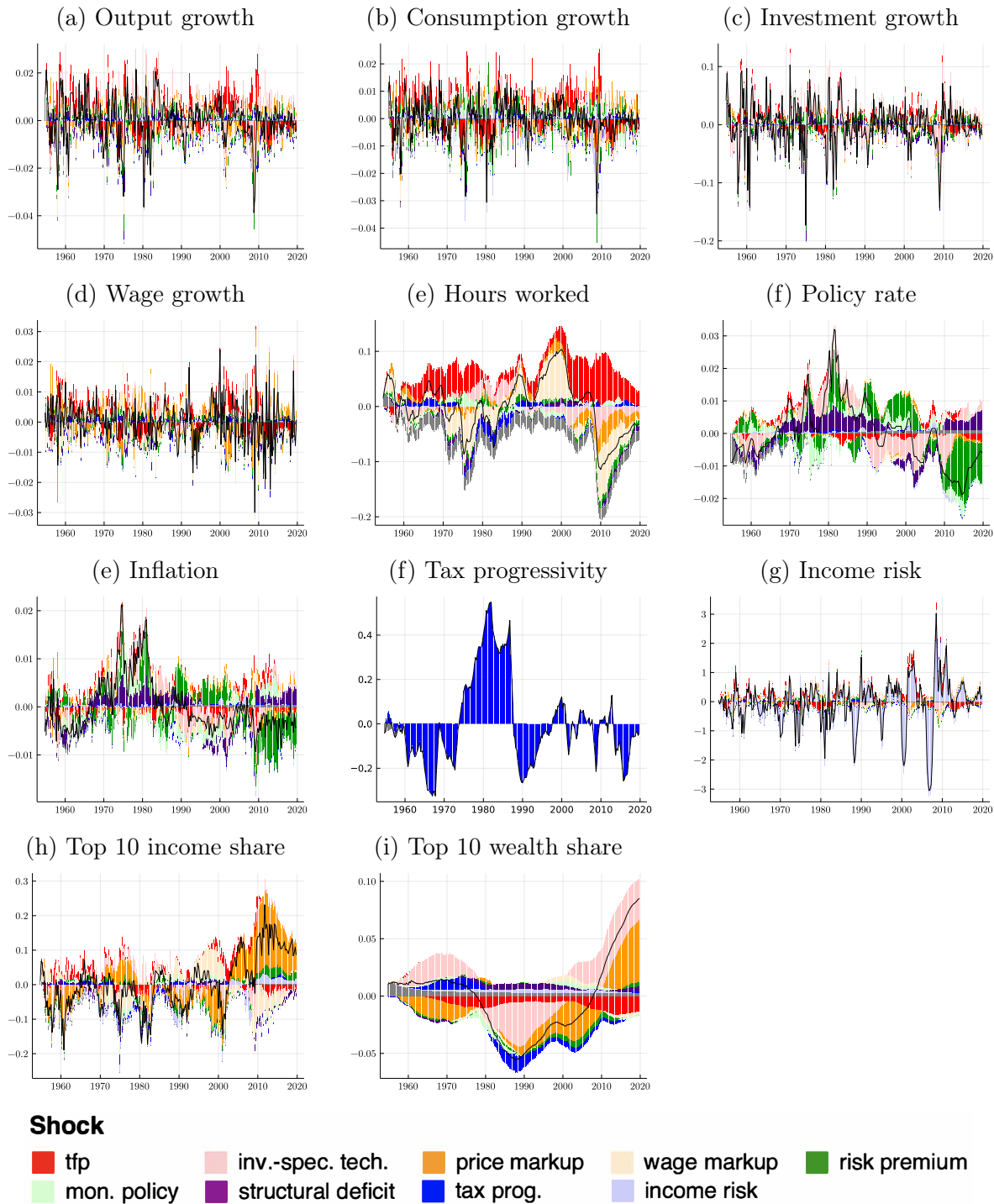## A.6 Historical decompositions of further observables

Figure A.2 shows the historical decomposition of all observables for the estimation of the HANK model and Figures A.3 for the HANK-X model. Figure A.4 shows the historical decomposition of non-observed variables target markups, profits, and the Top 1 percent share of income in the HANK-X model.

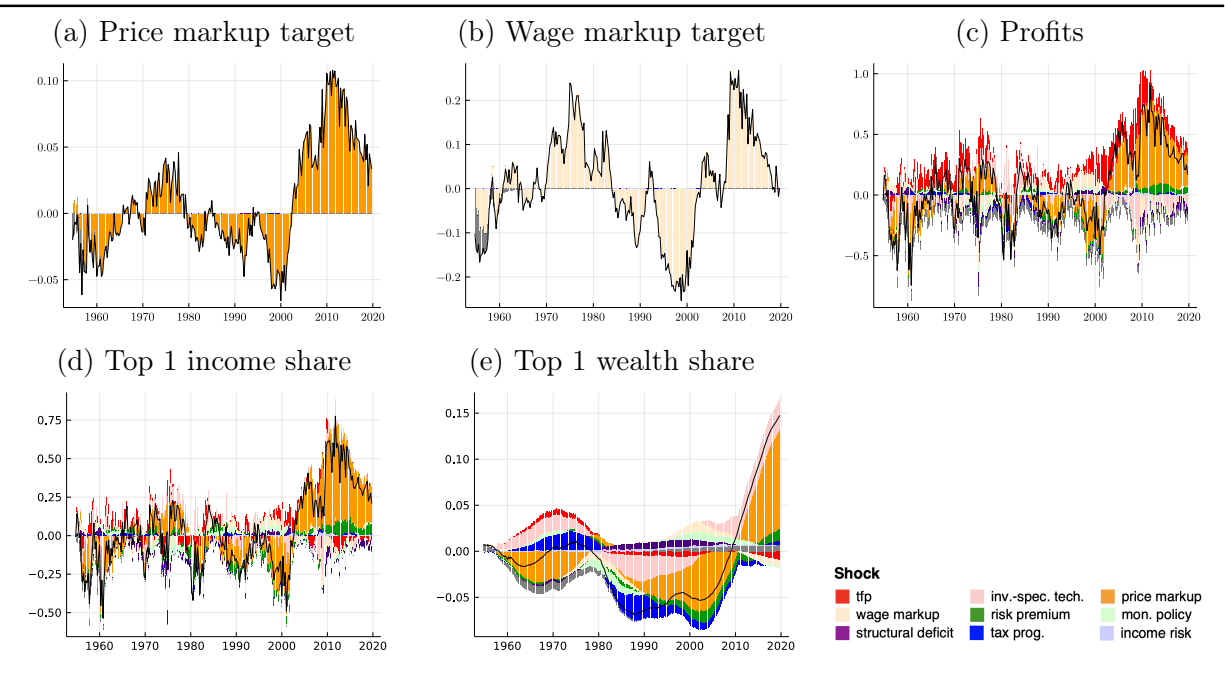**Figure A.2:** Historical decompositions of observables in HANK
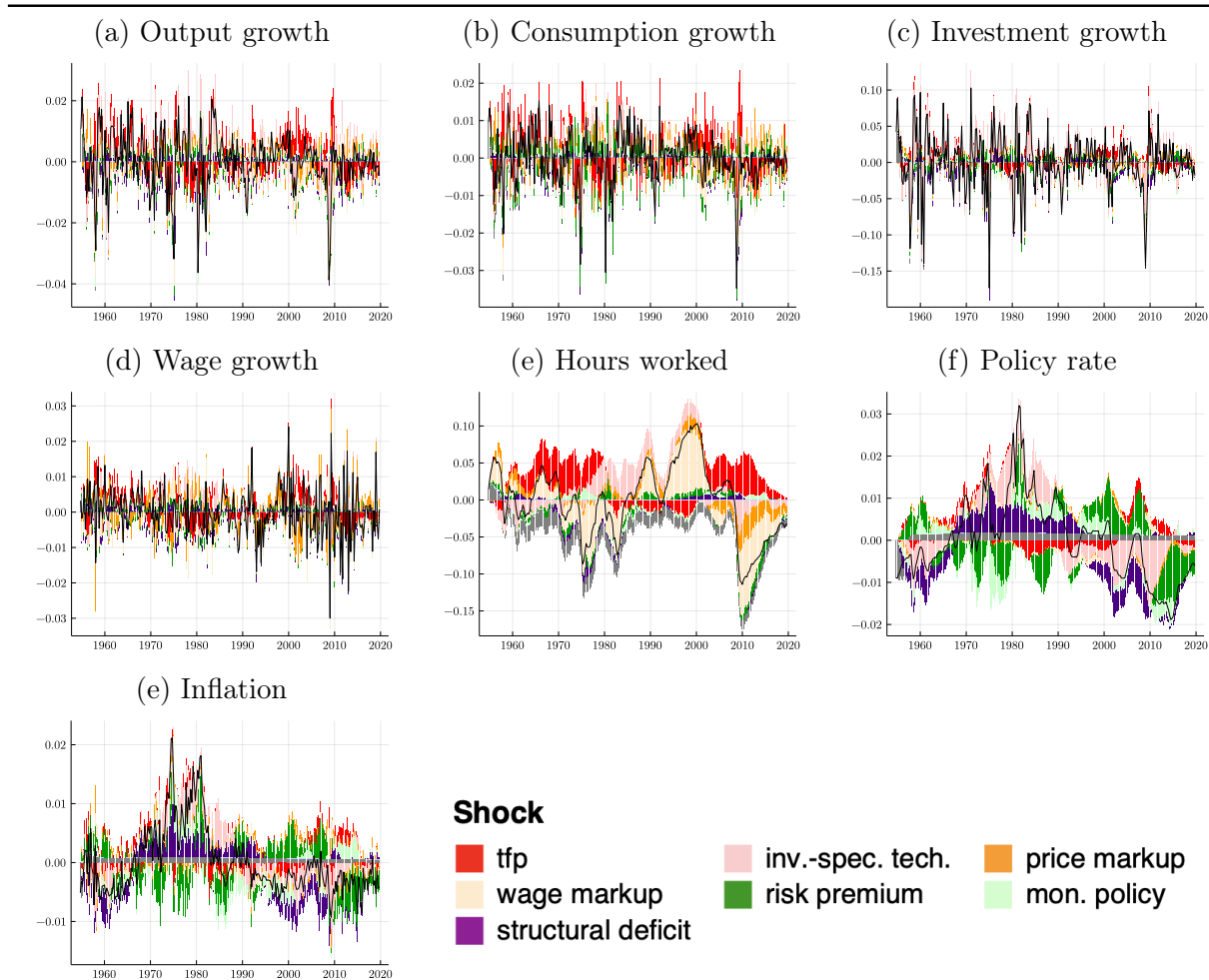
Notes: Historical decompositions of all observables in HANK. Y-axis: Percent deviation from mean.

Figure A.3: Historical decompositions of observables in HANK-X

Notes: Historical decompositions of all observables in HANK-X. Y-axis: Percent deviation from mean.

Figure A.4: Historical decompositions of further variables in HANK-X

(a) Price markup target

(b) Wage markup target

(c) Profits

(d) Top 1 income share

(e) Top 1 wealth share

Shock
tfp — inv.-spec. tech. — price markup
wage markup — risk premium — mon. policy
structural deficit — tax prog. — income risk

Notes: Historical decompositions of further unobserved variables in HANK-X. Y-axis: Percent deviation from mean.

A.7 Impulse Responses

Figures A.5 – A.9 plot the impulse response functions for the estimated RANK, HANK, and HANK-X model. The first panel on the top left corner of each figure shows the shock and the remaining panels show the responses of all potential observables.

Figure A.5: IRFs to structural deficit and monetary policy shocks
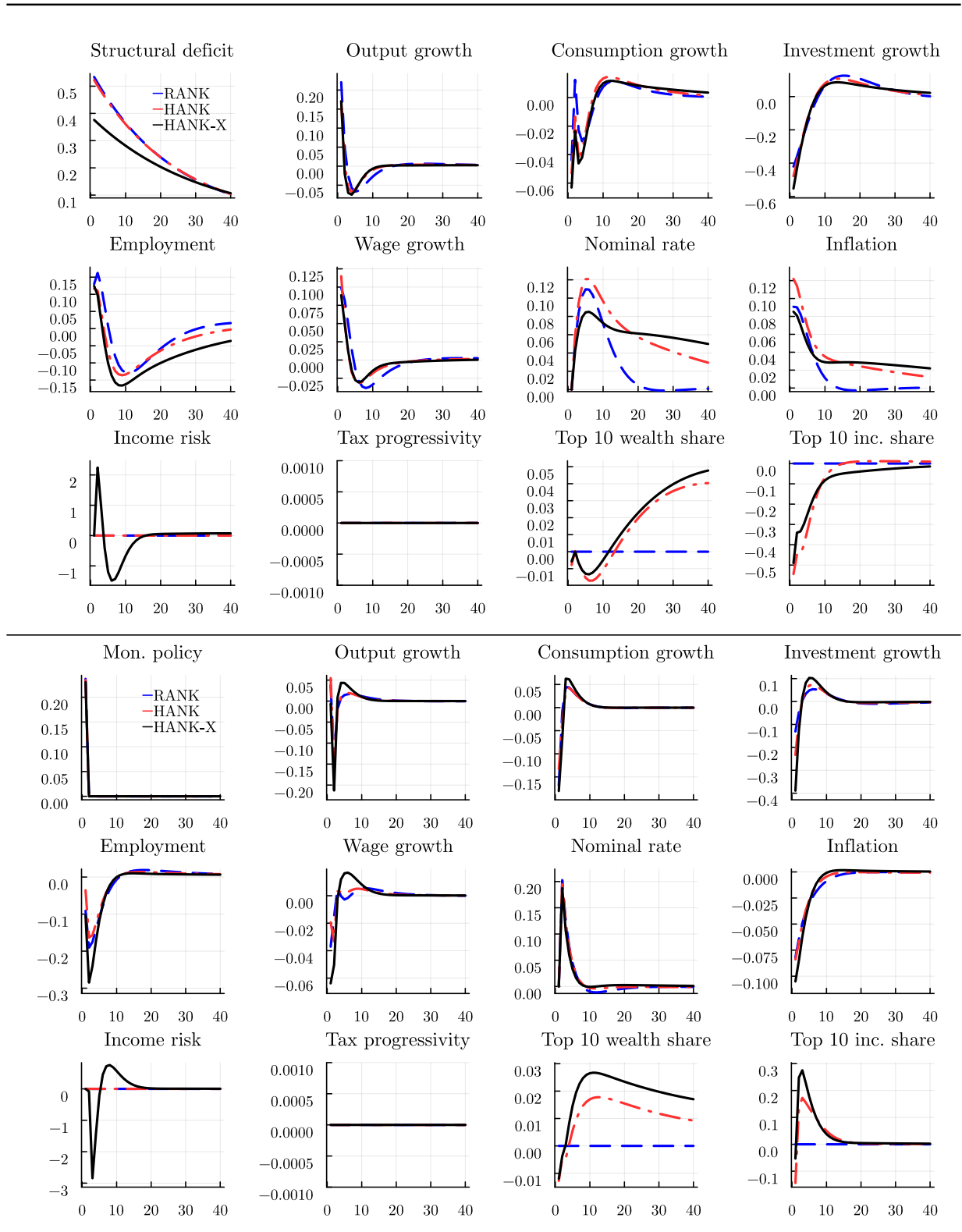
Notes: Top: IRF to a structural deficit shock. Bottom: IRF to a monetary policy shock. Blue-dashed line: RANK; red dashed-dotted line: HANK; black solid line: HANK-X. Y-axis: Percentage points for the nominal rate and inflation, otherwise percent.
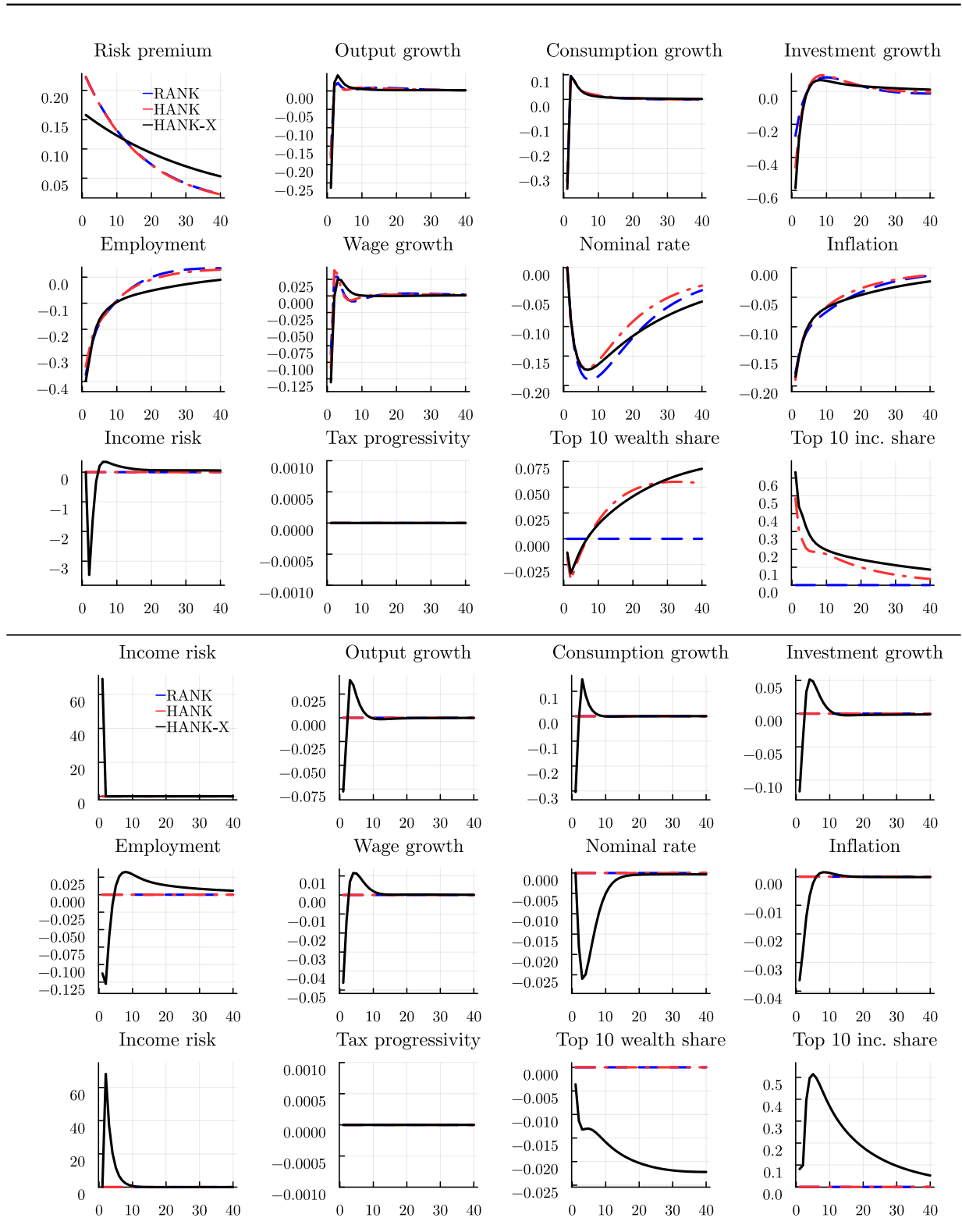
Figure A.6: IRFs to markup shocks
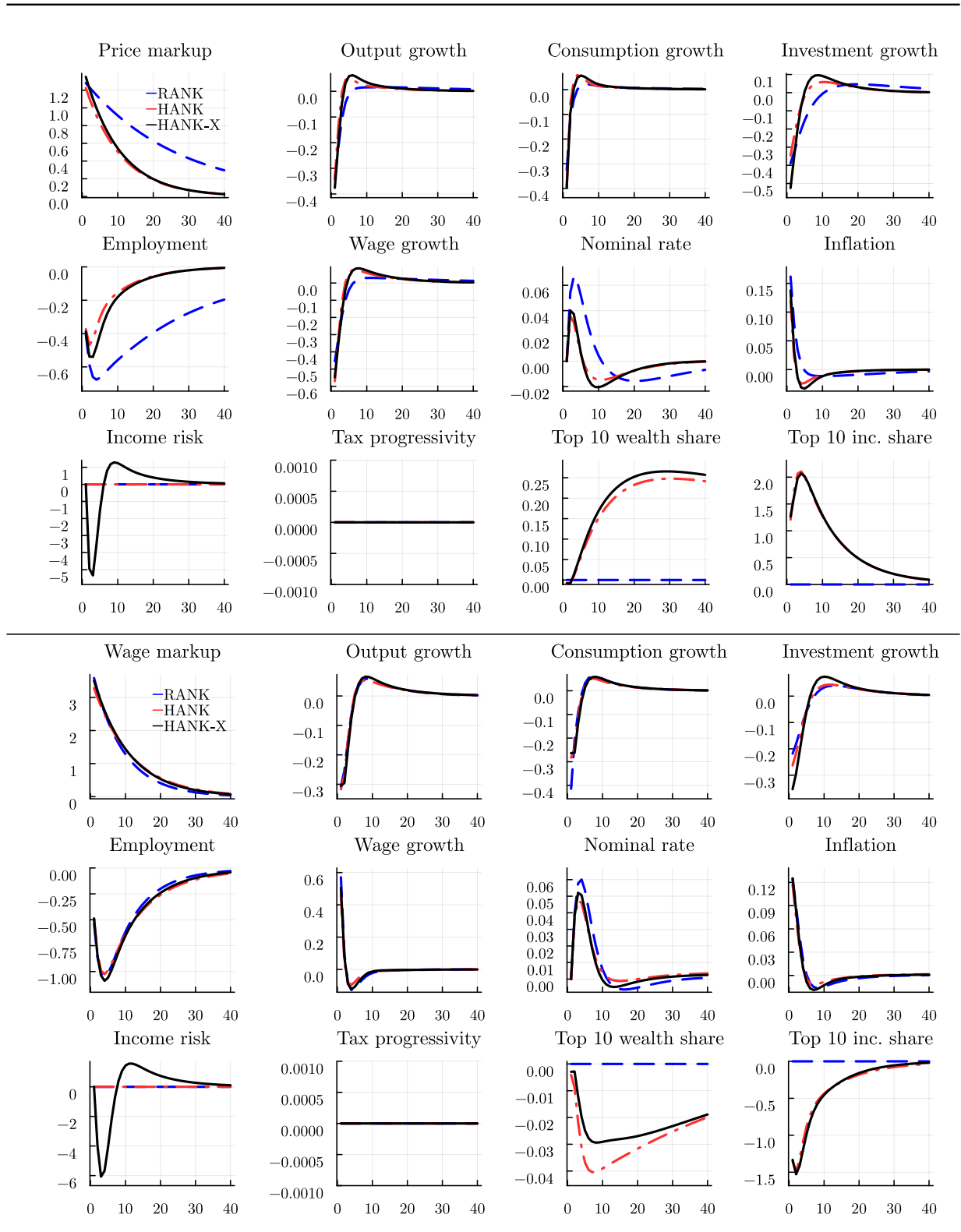
Notes: Top: IRF to a price-markup shock. Bottom: IRF to a wage-markup shock. Blue-dashed line: RANK; red dashed-dotted line: HANK; black solid line: HANK-X. Y-axis: Percentage points for the nominal rate and inflation, otherwise percent.

56

Figure A.7: IRFs to technology shocks
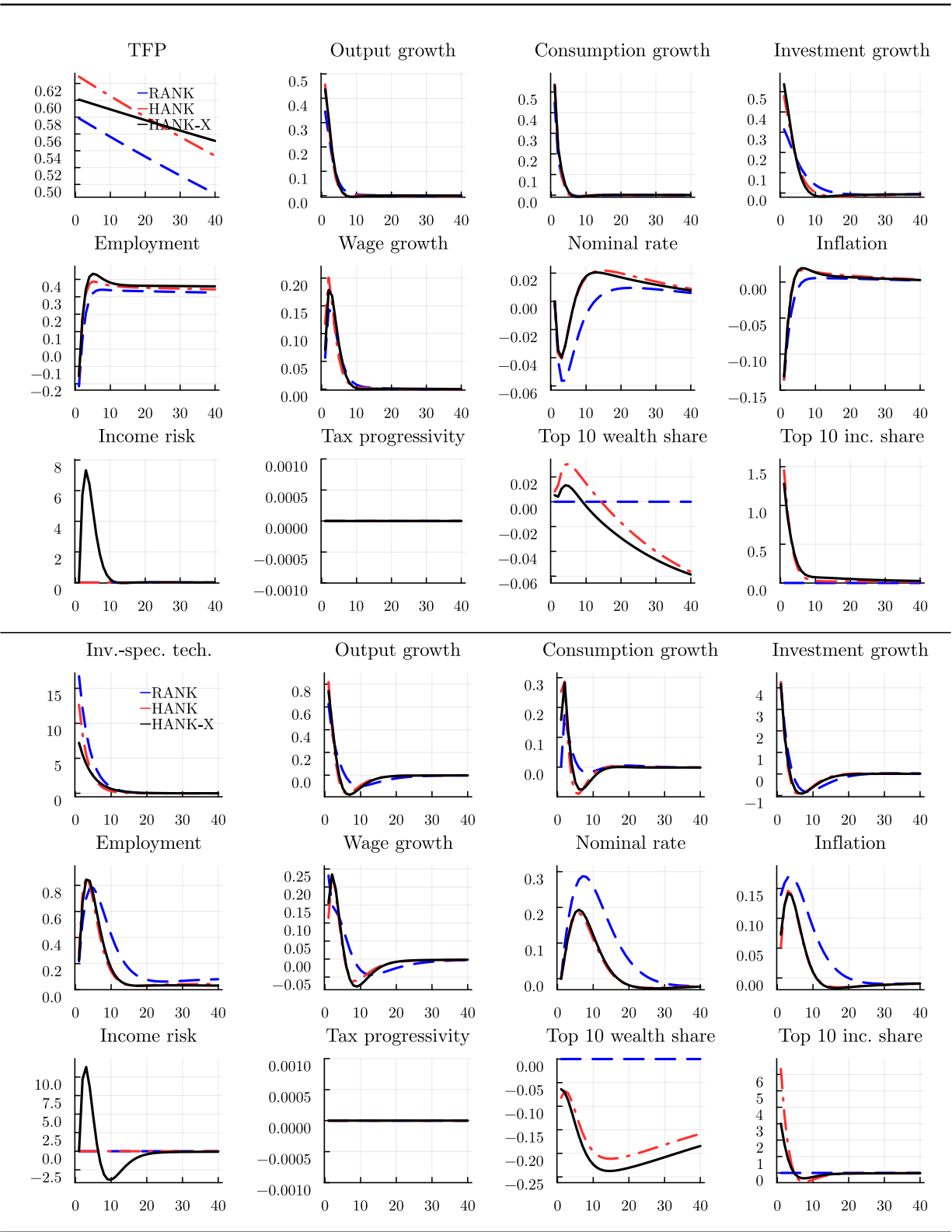
Notes: Top: IRF to a TFP shock. Bottom: IRF to an MEI shock. Blue-dashed line: RANK; red dashed-dotted line: HANK; black solid line: HANK-X. Y-axis: Percentage points for the nominal rate and inflation, otherwise percent.

Figure A.8: IRFs to risk premium and income risk shocks

Notes: Top: IRF to a risk premium shock. Bottom: IRF to an income risk shock. Blue-dashed line: RANK; red dashed-dotted line: HANK; black solid line: HANK-X. Y-axis: Percentage points for the nominal rate and inflation, otherwise percent.

Figure A.9: IRFs to a tax progressivity shock
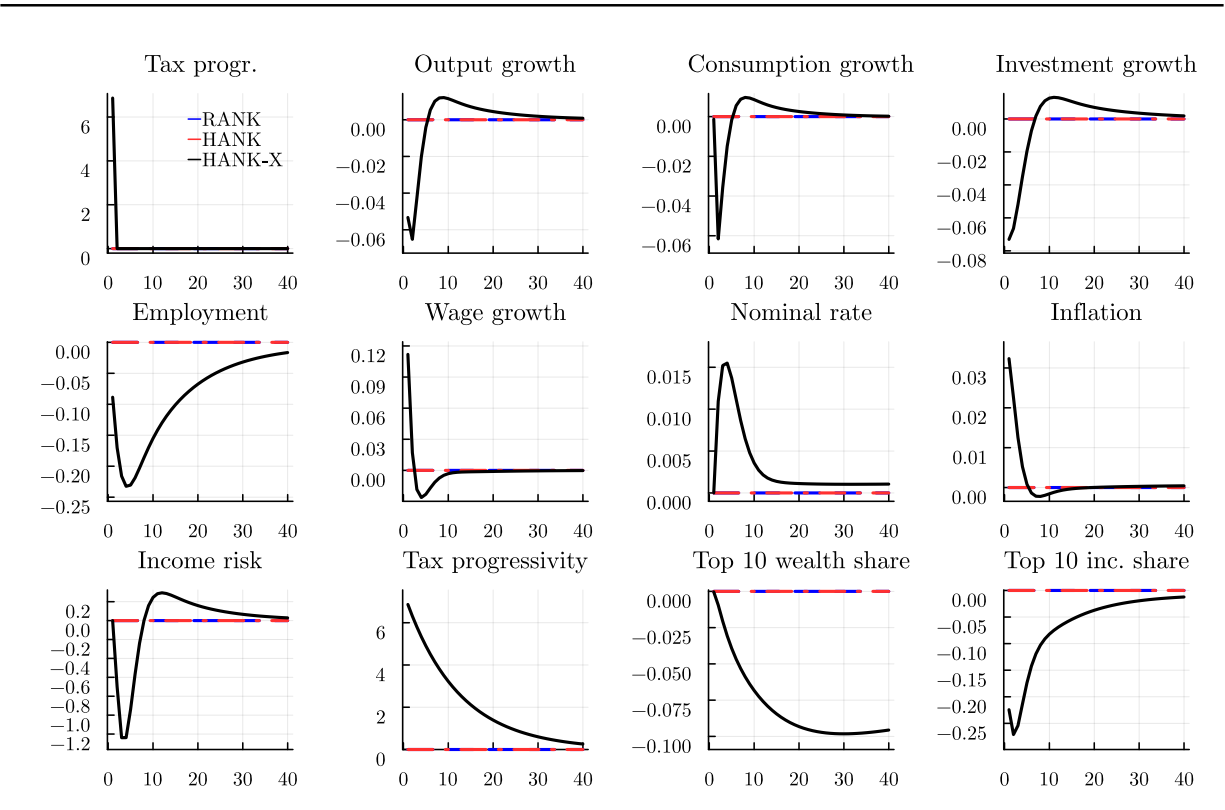
Notes: IRF to a tax progressivity shock. Blue-dashed line: RANK; red dashed-dotted line: HANK; black solid line: HANK-X. Y-axis: Percentage points for the nominal rate and inflation, otherwise percent.

A.8   MCMC diagnostics

We estimate each model using a single RWMH chain after an extensive mode search. After burn-in, 400,000 draws from the posterior distribution are used to compute the posterior statistics. The acceptance rates across chains are between 20% and 30%. Here, we provide Geweke (1992) convergence statistics for individual parameters of the RANK, HANK, and HANK-X models as well as traceplots for HANK and HANK-X. Geweke (1992) tests the equality of means of the first 10% of draws and the last 50% of draws (after burn-in). If the samples are drawn from the stationary distribution of the chain, the two means are equal and Geweke's statistic has an asymptotically standard normal distribution. Table A.4 reports the Geweke z-score statistic and the p-value for each parameter. Taking the evidence from Geweke (1992) and the traceplot graphs together, we conclude that our chains have converged.

Table A.4: Geweke (1992) convergence diagnostics

Parameter	RANK z-stat	RANK p-value	HANK z-stat	HANK p-value	HANK-X z-stat	HANK-X p-value
δ_s	-0.331	0.741	0.544	0.586	-0.991	0.322
ϕ	-0.044	0.965	0.791	0.429	1.011	0.312
κ	-0.623	0.533	0.636	0.525	-0.593	0.553
κ_w	-0.804	0.422	1.302	0.193	0.797	0.425
ι^Π	—	—	-1.639	0.101	-0.121	0.904
ρ_A	-1.465	0.143	-0.374	0.709	-0.509	0.611
σ_A	1.141	0.254	0.420	0.675	0.340	0.734
ρ_Z	-0.691	0.489	-0.324	0.746	-0.494	0.621
σ_Z	-0.392	0.695	0.507	0.612	-0.023	0.982
ρ_Ψ	-1.512	0.130	0.268	0.789	-0.96	0.337
σ_Ψ	0.943	0.346	0.968	0.333	1.107	0.268
ρ_μ	0.570	0.569	1.053	0.292	0.070	0.945
σ_μ	-0.196	0.845	-1.242	0.214	-1.52	0.128
$\rho_{\mu w}$	0.200	0.842	0.158	0.875	0.941	0.346
$\sigma_{\mu w}$	0.325	0.745	-0.874	0.382	-0.714	0.475
ρ_s	—	—	—	—	-0.975	0.329
σ_s	—	—	—	—	0.121	0.904
Σ_y	—	—	—	—	-1.056	0.291
ρ_R	1.037	0.300	0.157	0.875	1.111	0.267
σ_R	-0.567	0.571	0.916	0.360	-0.943	0.346
θ_π	1.293	0.196	1.070	0.285	-0.694	0.488
θ_Y	0.422	0.673	-0.429	0.668	1.337	0.181
γ_B	1.114	0.265	-1.349	0.177	0.026	0.979
γ_π	-1.19	0.234	-0.839	0.402	-0.718	0.473
γ_Y	-1.373	0.170	-0.923	0.356	0.508	0.611
ρ_D	1.735	0.083	0.352	0.725	0.333	0.740
σ_D	0.221	0.825	0.549	0.583	-0.582	0.560
ρ_τ	-0.369	0.712	-1.253	0.210	-0.432	0.666
γ_B^τ	0.251	0.802	-1.923	0.054	-0.963	0.335
γ_Y^τ	2.330	0.020	-1.179	0.238	-0.337	0.736
ρ_P	—	—	—	—	-1.267	0.205
σ_P	—	—	—	—	0.666	0.505
σ_{W10}^{me}	—	—	—	—	1.064	0.288
σ_{I10}^{me}	—	—	—	—	0.471	0.638

Note: Geweke (1992) equality of means test of the first 10% vs. the last 50% of draws. Failure to reject the null of equal means indicates convergence.
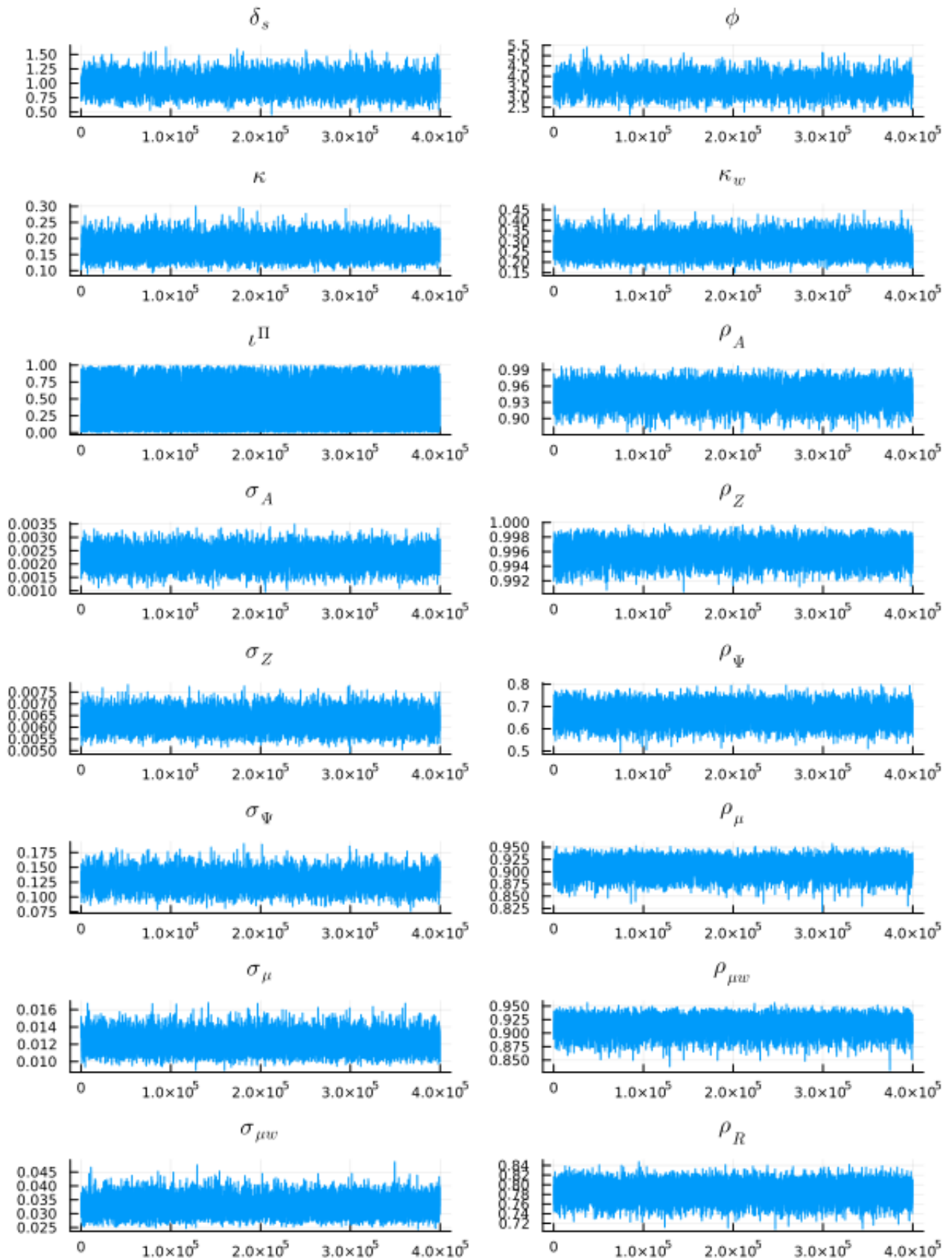
Figure A.10: MCMC draws of HANK model
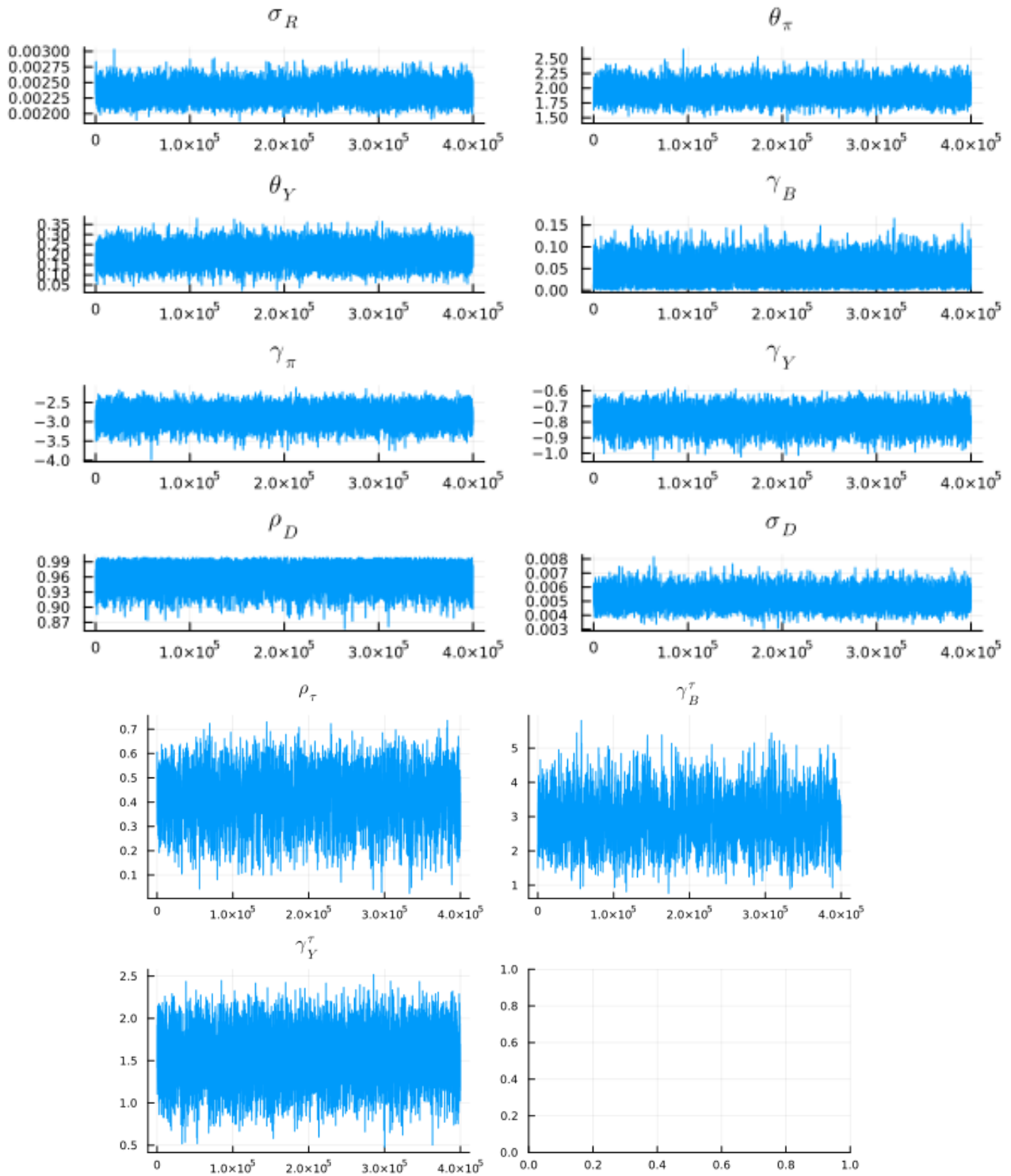
Figure A.11: MCMC draws of HANK model
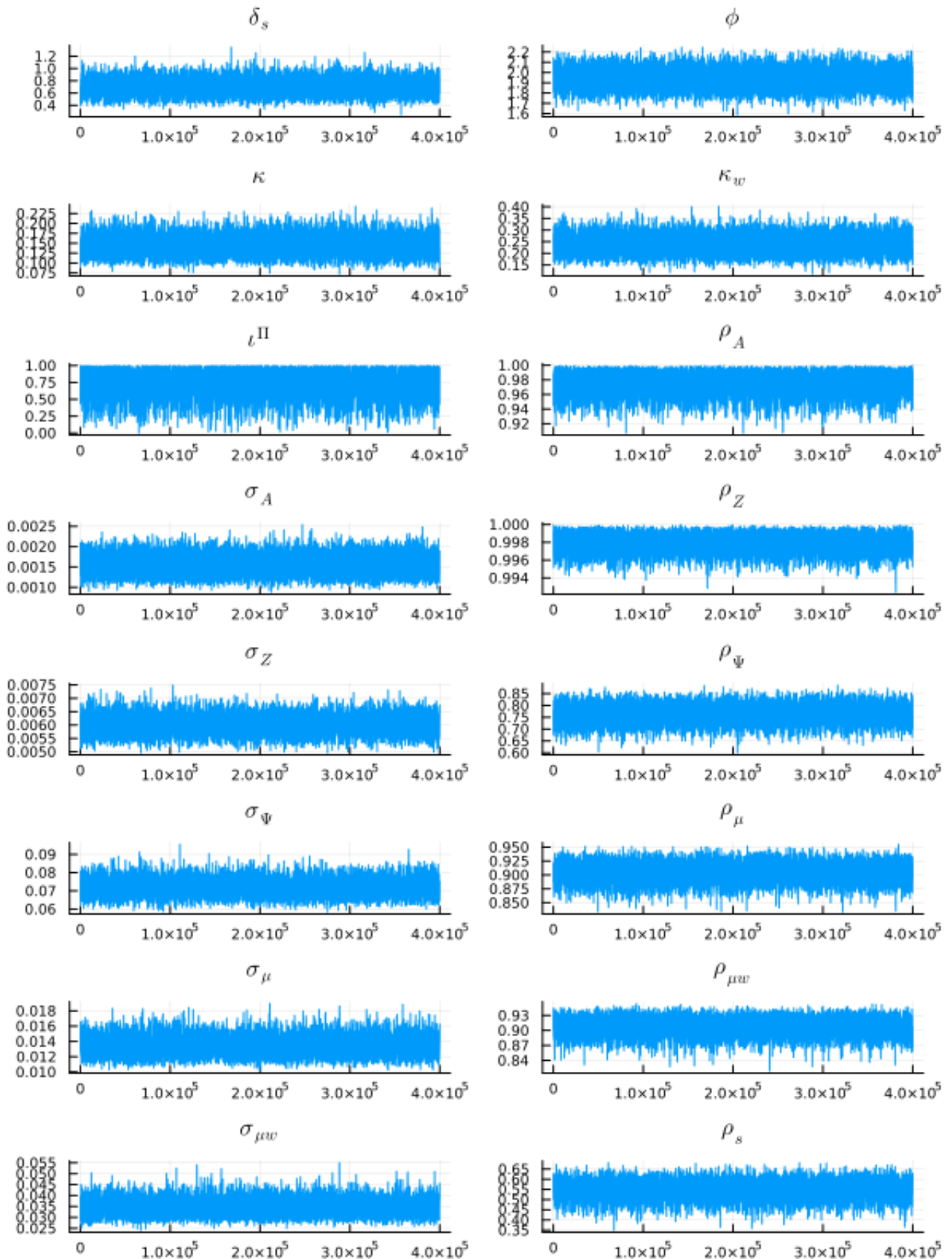
Figure A.12: MCMC draws of HANK-X model

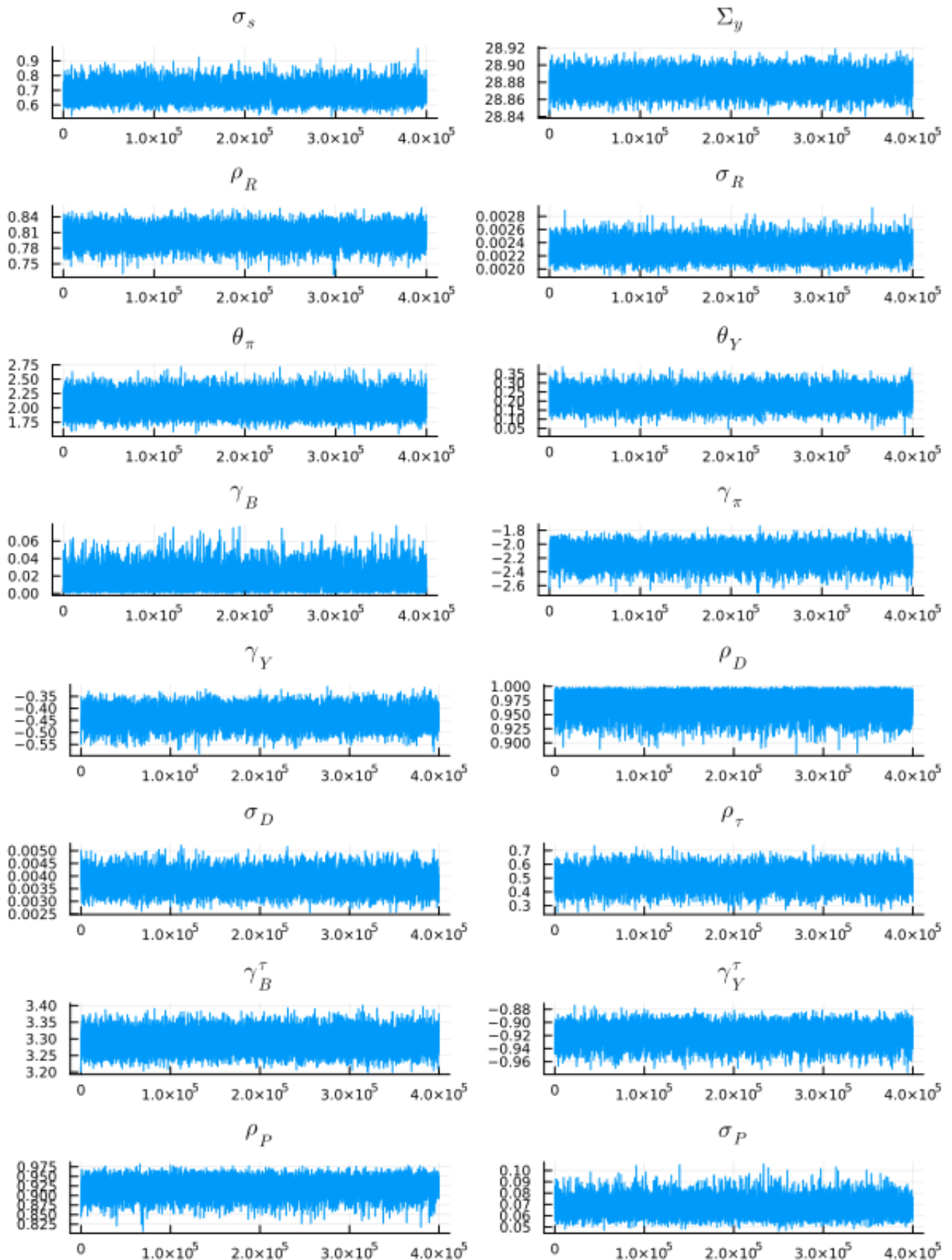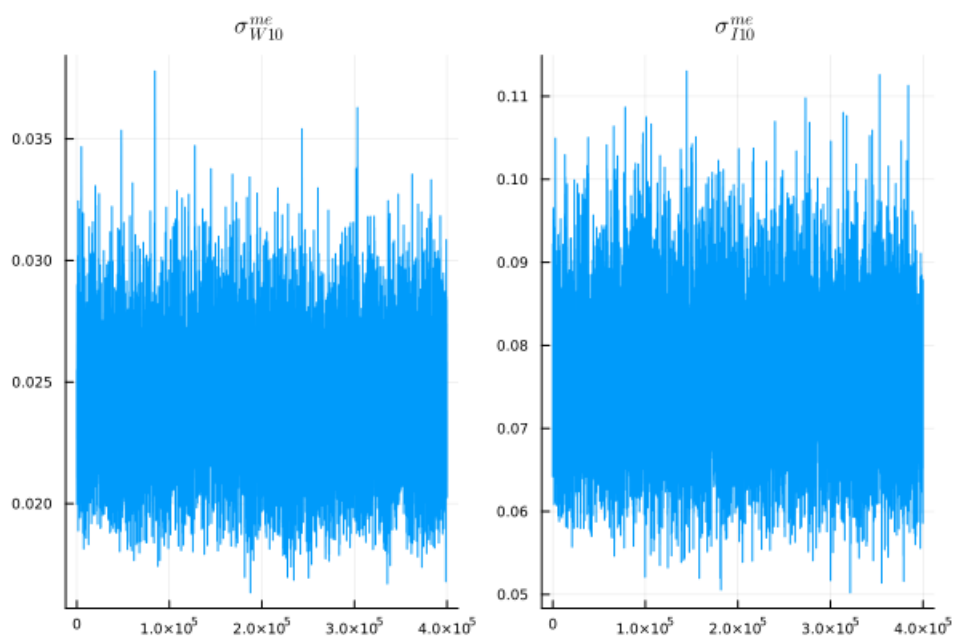**Figure A.13:** MCMC draws of HANK-X model

Figure A.14: MCMC draws of HANK-X model

B Robustness to alternative specifications

We estimate five variants of our model to understand the effect of potentially important data and modeling choices: 1) sample 1983-2019, 2) risk aversion (2 instead of 4), 3) paying out union profits proportional to idiosyncratic productivity (no wage compression), 4) systematic response of tax progressivity to income inequality, 5) King et al. (1988) preferences instead of Greenwood et al. (1988).

Appendix B.1 provides more details on each variant. Appendix B.2 contains the estimated parameters, Appendix B.3 the variance decompositions for all variants, and Appendix B.4 the historical decomposition of income and wealth inequality for the variants with risk aversion 2 and KPR preferences.

B.1 Description of variants

Below we quickly describe the recalibration of the steady state for variants 2) risk aversion, 3) union profits, and 5) KPR preferences. The other two variants, 1) sample split and 4) fiscal response to inequality, do not require a recalibration of the steady state. The sample split estimation is run using the same model and calibration as in the baseline. Allowing for a feedback coefficient of tax progressivity to the top 10 income share only affects the aggregate model part.

Risk aversion 2

Changing the coefficient of relative risk aversion to 2 (instead of 4) requires a recalibration of the steady state to match the same targets as listed in Table 1. In particular, we adjust the discount factor, the asset market participation frequency, the fraction of entrepreneurs, and the borrowing penalty. The re-calibration yields $\beta = 0.992$, $\lambda = 4.5\%$, $\zeta = 1/3750$, and $\bar{R} = 2.18\%$.

Proportional union profits

Paying out union profits proportional to idiosyncratic productivity (instead of lump sum) affects the steady-state distribution of income and requires a recalibration. Again, we adjust the discount factor, the asset market participation frequency, the fraction of entrepreneurs, and the borrowing penalty. The re-calibration yields $\beta = 0.982$, $\lambda = 7.0\%$, $\zeta = 1/7500$, and $\bar{R} = 1.35\%$.

Fiscal response to inequality

We change the policy rule for the tax progressivity parameter, τ_t^P, in HANK-X to the following:

$$\frac{\tau_t^P}{\bar{\tau}^P} = \left(\frac{\tau_{t-1}^P}{\bar{\tau}^P}\right)^{\rho_{\tau^P}} \left(\frac{T10IShare_t}{\overline{T10IShare}}\right)^{(1-\rho_{\tau^P})\gamma_W^{\tau^P}} \epsilon_t^P, \tag{B.1}$$

where the new parameter $\gamma_W^{\tau^P}$ captures the response of tax progressivity to income inequality. Its prior follows a standard normal distribution . We find that tax progressivity does respond to the top 10 income share with an estimated elasticity of 0.41. In the US, the fiscal authority responds to higher income inequality by increasing the progressivity of taxes thereby mitigating the increase in pre-tax income inequality to post-tax income inequality. However, tax progressivity is still largerly driven by exogenous shocks ϵ_t^P as the feedback from inequality is quantitatively small.

KPR preferences

Changing the preferences to King et al. (1988) preferences (instead of Greenwood et al. (1988)) also requires the recalibration of the steady state. The felicity function u, additively separable in consumption and leisure, now reads:

$$u(c_{it}, n_{it}) = \frac{c_{it}^{1-\xi} - 1}{1-\xi} - \gamma^{shift}\frac{n_{it}^{1+\gamma} - 1}{1+\gamma}, \tag{B.2}$$

with risk aversion parameter $\xi > 0$ and inverse Frisch elasticity $\gamma > 0$. The first-order condition for labor supply is:

$$n_{it} = \left[\frac{1}{\gamma^{shift}}u'(c)(1-\bar{\tau}^P)(1-\tau_t^L)(wh_{it})^{(1-\bar{\tau}^P)}\right]^{\left(\frac{1}{\gamma+\bar{\tau}^P}\right)}. \tag{B.3}$$

We recalibrate the steady state to match the capital-to-output ratio, the bonds-to-capital ratio, the fraction of borrowers, and the top 10 wealth share as reported in Table 1. This yields a discount factor of $\beta = 0.988$, a portfolio adjustment probability of $\lambda = 8.25\%$, a borrowing penalty of $\bar{R} = 3.56\%$, and a probability of becoming an entrepreneur of $1/2000$.

B.2   Parameter estimates

Table B.5 displays the estimation results for the model variants. The estimated parameters are broadly similar across variants with some exceptions. The KPR estimates feature lower real frictions and a different parameterization of the tax rule.

Table B.5: Posterior distributions: model variants

Parameter	Posterior				
	HANK (Post-83)	HANK (RA2)	HANK (Union)	HANK-X (Tax)	HANK (KPR)
Frictions					
δ_s	1.066	2.431	0.686	0.709	0.188
	(0.989, 1.143)	(1.835, 3.121)	(0.501, 0.891)	(0.475, 0.969)	(0.090, 0.325)
ϕ	2.767	3.846	3.758	1.918	0.226
	(2.072, 3.467)	(2.973, 4.853)	(3.118, 4.463)	(1.292, 2.603)	(0.151, 0.313)
κ	0.085	0.331	0.307	0.156	0.106
	(0.060, 0.116)	(0.262, 0.412)	(0.237, 0.389)	(0.120, 0.195)	(0.091, 0.123)
κ_w	0.279	0.584	0.670	0.252	0.324
	(0.208, 0.357)	(0.449, 0.741)	(0.515, 0.851)	(0.193, 0.316)	(0.247, 0.407)
ι^Π	0.704	0.399	0.506	0.604	0.230
	(0.608, 0.799)	(0.063, 0.824)	(0.139, 0.864)	(0.213, 0.919)	(0.201, 0.257)
Debt and monetary policy rules					
ρ_R	0.861	0.776	0.774	0.802	0.717
	(0.840, 0.880)	(0.748, 0.803)	(0.746, 0.800)	(0.776, 0.827)	(0.683, 0.749)
σ_R	0.138	0.248	0.237	0.232	0.269
	(0.123, 0.154)	(0.227, 0.271)	(0.217, 0.257)	(0.213, 0.253)	(0.245, 0.297)
θ_π	2.967	2.142	1.851	2.081	2.010
	(2.670, 3.272)	(1.955, 2.341)	(1.659, 2.054)	(1.845, 2.335)	(1.859, 2.173)
θ_Y	0.195	0.192	0.169	0.221	0.344
	(0.121, 0.269)	(0.123, 0.261)	(0.103, 0.236)	(0.151, 0.290)	(0.284, 0.403)
γ_B	0.020	0.107	0.058	0.026	0.009
	(0.005, 0.042)	(0.065, 0.153)	(0.015, 0.111)	(0.005, 0.060)	(0.001, 0.026)
γ_π	-2.307	-2.96	-3.457	-2.226	-1.808
	(-2.642, -2.005)	(-3.288, -2.659)	(-3.956, -3.008)	(-2.482, -1.985)	(-1.956, -1.664)
γ_Y	-0.635	-0.84	-1.068	-0.494	-0.287
	(-0.743, -0.536)	(-0.938, -0.746)	(-1.251, -0.892)	(-0.615, -0.387)	(-0.334, -0.244)
ρ_D	0.970	0.965	0.938	0.970	0.982
	(0.941, 0.991)	(0.935, 0.988)	(0.895, 0.974)	(0.933, 0.994)	(0.963, 0.995)
σ_D	0.270	0.516	0.681	0.382	0.313
	(0.202, 0.346)	(0.438, 0.599)	(0.575, 0.803)	(0.318, 0.452)	(0.279, 0.351)
Tax rules					
ρ_τ	0.394	0.311	0.366	0.464	0.425
	(0.263, 0.531)	(0.154, 0.457)	(0.217, 0.501)	(0.294, 0.602)	(0.416, 0.434)
γ_B^τ	3.114	2.379	2.511	3.384	-0.198
	(3.064, 3.166)	(1.464, 3.379)	(1.531, 3.554)	(2.386, 4.470)	(-0.218, -0.181)
γ_Y^τ	0.790	2.252	3.387	-0.148	-0.455
	(0.760, 0.819)	(2.142, 2.368)	(2.264, 4.441)	(-1.584, 1.232)	(-0.471, -0.441)
ρ_P	—	—	—	0.921	—
	(—, —)	(—, —)	(—, —)	(0.885, 0.950)	(—, —)
σ_P	—	—	—	6.927	—
	(—, —)	(—, —)	(—, —)	(5.877, 8.286)	(—, —)
γ_W^P	—	—	—	0.407	—
	(—, —)	(—, —)	(—, —)	(0.147, 0.659)	(—, —)

Table B.5: Posterior distributions: model variants – continued

Parameter	Posterior				
	HANK (Post-83)	HANK (RA2)	HANK (Union)	HANK-X (Tax)	HANK (KPR)
Structural shocks					
ρ_A	0.937	0.915	0.930	0.967	0.938
	(0.903, 0.965)	(0.881, 0.947)	(0.898, 0.961)	(0.936, 0.992)	(0.923, 0.953)
σ_A	0.185	0.244	0.228	0.165	0.163
	(0.156, 0.218)	(0.203, 0.288)	(0.168, 0.289)	(0.124, 0.207)	(0.144, 0.184)
ρ_Z	0.995	0.991	0.998	0.998	0.906
	(0.992, 0.997)	(0.988, 0.995)	(0.997, 0.999)	(0.996, 0.999)	(0.891, 0.920)
σ_Z	0.555	0.588	0.622	0.599	1.598
	(0.494, 0.623)	(0.545, 0.635)	(0.573, 0.675)	(0.551, 0.650)	(1.456, 1.742)
ρ_Ψ	0.794	0.728	0.599	0.769	0.969
	(0.724, 0.854)	(0.667, 0.787)	(0.518, 0.676)	(0.705, 0.830)	(0.962, 0.974)
σ_Ψ	6.609	13.588	13.122	7.170	3.250
	(5.302, 8.073)	(10.987, 16.515)	(10.943, 15.505)	(5.347, 9.219)	(2.814, 3.740)
ρ_μ	0.824	0.932	0.957	0.908	0.986
	(0.767, 0.872)	(0.908, 0.956)	(0.936, 0.975)	(0.882, 0.932)	(0.962, 0.998)
σ_μ	1.756	0.959	0.925	1.302	0.442
	(1.426, 2.183)	(0.867, 1.062)	(0.838, 1.024)	(1.144, 1.497)	(0.343, 0.587)
$\rho_{\mu w}$	0.922	0.944	0.963	0.909	0.819
	(0.894, 0.944)	(0.928, 0.959)	(0.947, 0.977)	(0.881, 0.933)	(0.777, 0.856)
$\sigma_{\mu w}$	3.586	2.477	2.363	3.453	2.810
	(2.996, 4.318)	(2.229, 2.758)	(2.142, 2.606)	(2.965, 4.051)	(2.367, 3.349)
Income risk process					
ρ_s	—	—	—	0.543	—
	(—, —)	(—, —)	(—, —)	(0.471, 0.606)	(—, —)
σ_s	—	—	—	68.923	—
	(—, —)	(—, —)	(—, —)	(61.172, 77.84)	(—, —)
Σ_y	—	—	—	22.169	—
	(—, —)	(—, —)	(—, —)	(22.101, 22.237)	(—, —)
Measurement errors					
σ_{W10}^{me}	—	—	—	2.438	—
	(—, —)	(—, —)	(—, —)	(2.069, 2.856)	(—, —)
σ_{I10}^{me}	—	—	—	7.701	—
	(—, —)	(—, —)	(—, —)	(6.411, 9.154)	(—, —)

Notes: The standard deviations of the shocks and measurement errors have been transformed into percentages by multiplying by 100. HANK (Post-83): HANK model estimated on post-Volcker data only; HANK (RA2): HANK model with risk aversion 2 instead of 4; HANK (Union): HANK model in which union profits are payed out proportionally to idiosyncratic productivity; HANK-X (Tax): HANK-X model with income-inequality feedback to tax progressivity; HANK (KPR): HANK model with KPR instead of GHH preferences. For more details see text.
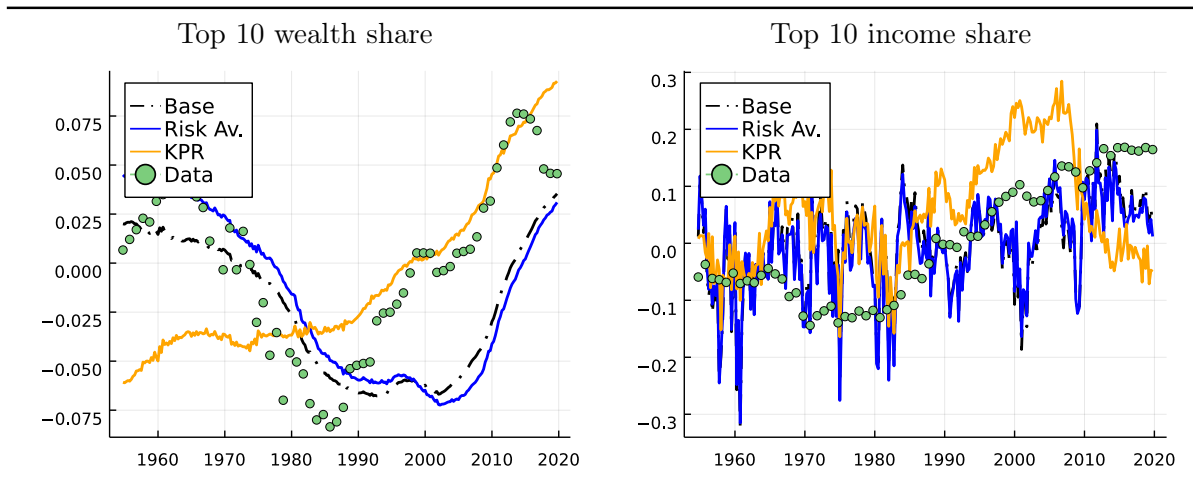
B.3   Variance decompositions

Figure B.15 shows that the variance decompositions are similar across all variants. Shocks to investment specific technology are by far the most important driver of output growth (explaining 40-60%), followed with some distance by shocks to TFP and wage markups. The same three shocks are prominent in consumption growth but of more equal importance and with TFP being the most important one. The variance decompositions of top 10 wealth and income shares are also quite similar. The outliers are KPR preferences and risk aversion 2. The former variant finds a larger role for TFP shocks in explaining inequality, while the latter finds a larger role for investment specific technology shocks.

Figure B.15: Variance decompositions: Output and consumption growth

Figure B.16: Historical decompositions: Inequality

Notes: Kalman smoother in comparison to the data for the top 10 wealth and income shares for the baseline and the estimated variants risk aversion 2 and KPR. Y-axis: Percent deviation from mean.

B.4 Historical decomposition of inequality

Figure B.16 shows the historical decomposition of inequality for these two variants, KPR and risk aversion 2, that differ most from the baseline in the previous section. Estimating the model with risk aversion 2 does not affect the implied time path of the top 10 income and wealth shares much. KPR preferences, however, do change the estimated results. Wealth inequality is now rising throughout the whole period, missing the U-shape. While income inequality is too high from 1970-2010 and too low afterwards such that the top 10 income share does not display a significant trend over the whole sample.

C Further details on the solution technique

C.1 Deviations of functionals from steady state

Our solution technique, following Reiter (2009), is based on writing the sequential equilibrium as a non-linear difference equation in function space. For this purpose, we write the marginal value functions, $\frac{\partial \mathbb{W}_t}{\partial b}$ and $\frac{\partial \mathbb{W}_t}{\partial k}$, as a sum of the stationary equilibrium function, $\bar{\mathbb{W}}_{b/k}$, and time-t deviations thereof, $\hat{\mathbb{W}}_{b/k,t}$. Since we work with Young's 2010 formulation of off-grid policies as fair gambles between grid points, we represent all functions as linear interpolants based on a set of node values for the full tensor grid of b, k, h. However, we represent the nodal values by their DCT coefficients, that is by the coefficients, $\theta^{p,q,r}$, of Chebychev polynomials, $T_{p/q/r}(\cdot)$, where we assume that the grid nodes were transformed to the corresponding Chebychev nodes:

$$\hat{\mathbb{W}}_{b/k,t}(b_i, k_j, h_l) = \sum_{p,q,r} \theta^{p,q,r}_{\mathbb{W}_{b/k},t} T_p(i) T_q(j) T_r(l). \tag{C.1}$$

The advantage of this formulation is that we can read off from the stationary equilibrium solution, which sparse polynomial would have been a good approximation to the non-sparse solution by comparing the absolute values of $\theta^{p,q,r}$. One way to do this is to look at the function values in the stationary distribution and fit the polynomials. If we had restricted the stationary equilibrium solution to the sparse polynomial class that forces the small coefficients to zero, then the solution would not have changed much. While we do not enforce this restriction in calculating $\bar{\mathbb{W}}_{b/k}$, we use it to select a baseline set of polynomials, i.e., the coefficients $\theta^{p,q,r}$ in (C.1), to be perturbed when we linearize the system.

In addition, we can use the envelop theorem, to calculate recursively the response of the value functions (or derivatives thereof) to a change in an expected future price P_{t+h}. Assuming that we wrote the problem such that prices do only show up contemporaneously in the Bellman equation, we have for $h > 0$:

$$\frac{\partial \mathbb{W}_t}{\partial P_{t+h}} = \left(\frac{\partial u}{\partial x_{t+1}} + \beta \Gamma \frac{\partial \mathbb{W}_{t+1}}{\partial x_{t+1}} \right) \frac{\partial x_{t+1}}{\partial P_{t+h}} + \beta \Gamma \frac{\partial \mathbb{W}_{t+1}}{\partial P_{t+h}}, \tag{C.2}$$

where Γ is the transition matrix induced by stationary equilibrium policies and income shocks (i.e., it includes the expectations operator). Here, x_{t+1} are the endogenous idiosyncratic states. Importantly, the sum of the first two terms is zero when the choice of x_{t+1} is not constrained because the borrowing constraint does not bind. When it binds, however, $\frac{\partial x_{t+1}}{\partial P_{t+h}} = 0$. This implies that the product of the two terms is always zero and we can write $\frac{\partial \mathbb{W}_t}{\partial P_{t+h}}$

recursively as

$$\frac{\partial \mathbb{W}_t}{\partial P_{t+h}} = \beta^h \Gamma^h \underbrace{\frac{\partial \mathbb{W}_t}{\partial P_t}}_{=:\mathbf{w}_P}. \tag{C.3}$$

Selecting polynomial terms on the basis of how well they allow to describe the response of the value function to price changes, $\mathbf{w}_P$, helps to obtain a good approximation of the planning problem with a sparse polynomial basis.

For the distribution function, we extend the approach of Bayer and Luetticke (2020). Again following Young (2010), we write the distribution function in terms of its histogram over the discrete nodes b, k, h. We then re-interpret this histogram as the histogram of its copula (i.e., the joint-distribution of marginal probabilities) by translating the axes from the b, k, h space to the space of the marginal distributions F_t^b, F_t^k, F_t^h. This allows us to split the joint distribution of b, k, h into three separate objects: First, marginal distributions at time t, second the copula in the stationary equilibrium $\bar{C}(F_{i,t}^b, F_{j,t}^k, F_{l,t}^h)$, at the grid points of b, k, h with indices i, j, l evaluated at these marginals and, third, deviations of the copula, $\hat{C}_t$.

The advantage of this splitting the distribution into three objects is that we can work with different degrees of precision for the different objects. Again, we write all functionals as linear interpolants over a set of nodal values. The nodal values of $\bar{C}$ are simply given by the stationary distribution. This means, we define the node grid $\{F_i^b, F_j^k, F_l^h\}$ in line with the stationary marginal distributions over the b, k, and h grid, respectively.

The deviation of the copula is again given by a linear interpolant of the pdf $d\hat{C}$ over nodal values represented by a discrete-cosine transform that uses a subset of the nodal grid of $\bar{C}$:

$$d\hat{C}_t(F_i^b, F_j^k, F_l^h) = \sum_{p,q,r} \theta_{C,t}^{p,q,r} T_p(i) T_q(j) T_r(l). \tag{C.4}$$

A sparser grid for $\hat{C}$ implies that we need to perturb less coefficients. Working with the DCT-transformation on top, allows us to easily formulate the constraints that are posed by making sure that the combined copula $\bar{C} + \hat{C}$ remains a copula (fulfills the restrictions on partial integrals). This constraint translates into parameter restrictions on $\theta_C^{p,q,r}$, where $\theta_C^{p,q,r} = 0$ for $p = q = 1$, $q = r = 1$, or $p = r = 1$. This restriction ensures that $\int dC_t = 0$ and reflects that $\sum_s T_s(m) = 0$ for $s > 1$ where m is the Chebychev node index. We do not restrict the perturbed coefficients any further than this before running the second-step model reduction.

C.2 Intuition for the possibility of a strong model reduction

The procedure above gives us the first-stage model reduction. It is based only on objects calculated from the stationary equilibrium. While this renders solving for a sequential equilibrium feasible, because the model becomes sufficiently small in terms of the number of variables involved, this number is still large and would thus yield long estimation times. Our second-stage model reduction leverages the Bayesian setup, using prior knowledge about the dynamics to derive a factor representation of the idiosyncratic model part. We find that it reduces the model dramatically in the number of variables, making estimation feasible.

To gain some intuition for why such strong further model reduction is possible, it is useful to draw insights from the sequence-space solution techniques (Auclert et al., 2021b). The key idea, which sequence-space techniques leverage, is that the household's decision problem depends only on the expected sequence of a small set of "prices" P_t.[33]

The sequence-space method assumes that it is possible to approximate the impact of a shock by a finite T period sequence of prices. Given this assumption, we know that we can write the equilibrium sequence of prices as an impulse response

$$\mathbb{E}_t dP_{t+h} = \Phi_h \epsilon_t. \tag{C.5}$$

Stability requires that $\lim_{h \to \infty} \Phi_h = 0$ and if the sequence-space solution is exact at horizon T, $\Phi_h \approx 0 \quad \forall h \geq T$.

If we now consider infinitesimally small shocks, we can write the deviations of the value functions (in a total differential notation) as

$$d\mathbb{W}_t = \mathbb{E}_t \sum_{h=0}^{T} \frac{\partial \mathbb{W}_t}{\partial P_{t+h}} dP_{t+h} = \mathbb{E}_t \sum_{h=0}^{T} (\beta\Gamma)^h \mathbf{w}_P dP_{t+h} = \sum_{s=0}^{T} \underbrace{\sum_{h=0}^{T-s} (\beta\Gamma)^h \mathbf{w}_P \Phi_{s+h}}_{=:C_s} \epsilon_{t-s}. \tag{C.6}$$

The second equality uses the envelope result from (C.3). This implies for the variance covariance of deviations in the value functions:

$$\mathbb{E} d\mathbb{W}_t d\mathbb{W}_t' = \begin{bmatrix} C_0 & \cdots & C_T \end{bmatrix} \begin{bmatrix} \Sigma_\epsilon & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \Sigma_\epsilon \end{bmatrix} \begin{bmatrix} C_0' \\ \vdots \\ C_T' \end{bmatrix} = \sum_{s=0}^{T} C_s \Sigma_\epsilon C_s' \tag{C.7}$$

The rank of the sum of matrices is bounded from above by $T \times J$, where J is the number

---

[33]These are: level and progressivity of taxes, income risk, wage rate, real interest rate on liquid assets, price of capital, return on capital, entrepreneurial profits, and union profits; see equation (36) for example.

of shocks. This means that, under the assumption that a T-period approximation is good enough (for a sequence-space solution), there are at most $T \times J$ factors in the value functions. This upper bound is, however, too loose: The matrices C_s shrink in s towards zero because of discounting in the planning problem, $\lim_{s \to \infty} \beta^s = 0$, and the stability of the price process, $\lim_{h \to \infty} \Phi_h = 0$. What is more, the convergence of Γ^s also implies that C_s become, in some sense, more similar when s increases. This effectively means that C_s converges more quickly to zero than Φ_s or β^s alone and the sum (C.7) can be approximated well using a smaller T.

Similarly, the change in distribution is accumulated from up to T periods back. This accumulation involves effects of past prices and expectations about future prices. Whenever prices or their expectations change in period $t - h$, the transition matrix in that period Γ_{t-h} changes. These changes, as they are infinitesimal in a first-order approximation, get accumulated through the stationary equilibrium transition matrix Γ in the following way:

$$
\begin{aligned}
d\Theta_t &= \sum_{h=0}^{T} \Gamma^{h\prime} d\Gamma'_{t-h} \bar{\Theta} = \sum_{h=0}^{T} \Gamma^{h\prime} \left[\frac{\partial \Gamma' \bar{\Theta}}{\partial P} dP_{t-h} + \frac{\partial \Gamma' \bar{\Theta}}{\partial \mathbb{W}_{+1}} \mathbb{E}_{t-h} d\mathbb{W}_{t-h+1} \right] \\
&= \sum_{h=0}^{T} \Gamma^{h\prime} \left[\frac{\partial \Gamma' \bar{\Theta}}{\partial P} \sum_{j=0}^{T} \Phi_j \epsilon_{t-h-j} + \frac{\partial \Gamma' \bar{\Theta}}{\partial \mathbb{W}_{+1}} \sum_{j=0}^{T} C_{j+1} \epsilon_{t-h-j} \right] \\
&= \sum_{h=0}^{T} \sum_{j=0}^{T} \Gamma^{h\prime} \left[\frac{\partial \Gamma' \bar{\Theta}}{\partial P} \Phi_j + \frac{\partial \Gamma' \bar{\Theta}}{\partial \mathbb{W}_{+1}} C_{j+1} \right] \epsilon_{t-h-j} \\
&= \sum_{s=0}^{2T} \sum_{h=0}^{s} \sum_{j=0}^{s-h} \Gamma^{h\prime} \left[\frac{\partial \Gamma' \bar{\Theta}}{\partial P} \Phi_j + \frac{\partial \Gamma' \bar{\Theta}}{\partial \mathbb{W}_{+1}} C_{j+1} \right] \epsilon_{t-s} = \sum_{s=0}^{2T} D_s \epsilon_{t-s},
\end{aligned}
\tag{C.8}
$$

where $\partial \mathbb{W}_{+1}$ denotes the derivative with respect to the continuation value.

As before the stability of the price process adds to the convergence, but now the variance of the distribution accumulates twice as many variance covariance matrices. What is more, D_s for low s accumulates more terms and thus does converge to zero as fast as C_s does. Here, the key role for convergence lies in Γ^h converging against a matrix with identical rows, $\bar{\Theta}$, and the fact that when summing over grid points $\sum \frac{\partial \Gamma' \bar{\Theta}}{\partial P} = \sum \frac{\partial \Gamma' \bar{\Theta}}{\partial \mathbb{W}_{+1}} = 0$ because the total mass of the distribution cannot change. Overall, we can expect the variance covariance matrix of the distribution to have a rank that is below the number of grid points, at most $2T \times J$, but larger than the one of the value functions.

C.3   Intuition for local invariance of model reduction

What is important, in both (C.7) and (C.8) the parameters we estimate only enter through their effect on price dynamics Φ_h. They affect neither the stationary equilibrium transition

matrix Γ, nor the response of the value functions to price changes $\mathbf{w}_P$, and they also have no effect on how the optimal household policy responds to price or continuation value changes, $\frac{\partial \Gamma' \bar{\Theta}}{\partial P_t}$ and $\frac{\partial \Gamma' \bar{\Theta}}{\partial \mathbb{W}_{+1}}$.

While the price dynamics changes in parameters, the changes are bounded. The priors, the model structure, and the data impose a restriction on how much the price process (C.5) changes between two likelihood evaluations. This implies that an ideal reduction basis under some parameters will remain good in their vicinity.

Another way to express this can be obtained by rewriting the linearized difference equation (42). As above we define objects in a way such that only contemporaneous prices enter the value functions, such that $B_{fX}, B_{Xf} = 0$ holds.[34] What is more, B_{ff} is typically invertible.[35] This implies that we can write the model as:

$$\underbrace{\begin{bmatrix} B_{ff}^{-1} A_{ff} & B_{ff}^{-1} A_{fX} \\ A_{Xf} & A_{XX} \end{bmatrix}}_{:=\tilde{A}} \begin{bmatrix} f_t \\ X_t \end{bmatrix} = -\mathbb{E}_t \begin{bmatrix} f_{t+1} \\ B_{XX} X_{t+1} \end{bmatrix}, \tag{C.9}$$

to obtain the equivalent formulation:

$$\begin{bmatrix} \tilde{A}_{ff} & \tilde{A}_{fX} \\ \tilde{A}_{Xf} & \tilde{A}_{XX} \end{bmatrix} \begin{bmatrix} f_t \\ X_t \end{bmatrix} = -\mathbb{E}_t \begin{bmatrix} f_{t+1} \\ B_{XX} X_{t+1} \end{bmatrix}. \tag{C.10}$$

Applying a singular value decomposition on $\tilde{A}_{ff}$, the idiosyncratic to idiosyncratic feedback matrix that does not depend on any of the estimated parameters, we can rewrite the system as:

$$\begin{bmatrix} U \Sigma V' & \tilde{A}_{fX} \\ \tilde{A}_{Xf} & \tilde{A}_{XX} \end{bmatrix} \begin{bmatrix} f_t \\ X_t \end{bmatrix} = -\mathbb{E}_t \begin{bmatrix} f_{t+1} \\ B_{XX} X_{t+1} \end{bmatrix}, \tag{C.11}$$

and invoke the Eckart-Young-Mirsky theorem to reduce the system in a first step to

$$\begin{bmatrix} V'U \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} & V' \tilde{A}_{fX} \\ \tilde{A}_{Xf} V & \tilde{A}_{XX} \end{bmatrix} \begin{bmatrix} V' f_t \\ X_t \end{bmatrix} \approx -\mathbb{E}_t \begin{bmatrix} V' f_{t+1} \\ B_{XX} X_{t+1} \end{bmatrix}, \tag{C.12}$$

---

[34] $B_{Xf} = 0$ holds because the distribution only *directly* affects summary variables at t but no variables at $t + 1$. The policy functions only affect the distribution. To obtain $B_{fX} = 0$ one needs to write the model such that only prices with a time-index t show up in the household's problem and all time $t + 1$ prices are implicitly captured by the continuation value/time $t + 1$ policy functions.

[35] The matrix B_{ff} has a block tridiagonal structure, where the derivative of the discrete time Fokker-Planck equation $\Theta'_{t+1} - \Gamma' \Theta'_t = 0$ with respect to Θ_{t+1} is unity. The derivatives with respect to continuation values $\mathbb{W}_{+1}$ of the Bellman equation is $\beta \Gamma$, while it is $\frac{\partial}{\partial \mathbb{W}_{+1}} \left(\Gamma' \bar{\Theta}' \right) = \left[\sum_h \bar{\Theta}_h \frac{\partial \Gamma_{h,i}}{\partial \mathbb{W}_{j,+1}} \right]_{i,j=1...n}$ for the Fokker-Planck equation.

where only the largest singular values Σ_1 are retained and it becomes sufficient to solve the reduced system

$$\begin{bmatrix} V_1'U\Sigma_1 & V_1'\tilde{A}_{fX} \\ \tilde{A}_{Xf}V_1 & \tilde{A}_{XX} \end{bmatrix} \begin{bmatrix} Y_t \\ X_t \end{bmatrix} \approx -\mathbb{E}_t \begin{bmatrix} Y_{t+1} \\ B_{XX}X_{t+1} \end{bmatrix}, \tag{C.13}$$

where V_1 refers to the rows in V that correspond to the largest singular values.

This means that a large component of the actual model reduction is strictly independent of the estimated parameters. The actual model reduction goes beyond the singular value decomposition above, and for that reason, needs to be updated very infrequently during the estimation procedure.

This first stage reduction is necessary despite this result because an unreduced system would have a too large matrix $\tilde{A}_{ff}$ to perform the SVD. In our application this matrix would have $660,000^2$ entries.

C.4   Direct IRF comparison across solution techniques

Figures C.17 to C.20 compare the impulse responses of the observables used in the estimation of the HANK model obtained from our solution method to those obtained from a sequence-space method assuming a 300 period transition. The terminal values are assumed to be given by the state-space solution instead of the stationary equilibrium. The figures are organized by observable variables and show the responses to the various shocks in one figure. Figures C.21 to C.26 repeat this exercise for the HANK-X estimates.

The figures show that the differences in the IRFs are almost zero. What the IRFs also show is that the TFP shock leads to a persistent change in the capital stock (which can be seen in the persistent increase of employment). We also compared the sequence space solution with a 300 period transition to itself using the state-space solution as terminal outcome and the stationary equilibrium. Given the persistent change in the capital stock after a TFP shock, a 300 periods transition is not a good approximation and we find that the approximation error between the two solutions is for persistent variables more than one order of magnitude larger than between sequence and state-space solution. Results are available upon request.

Figure C.17: Comparison of IRFs across solution methods (HANK model)
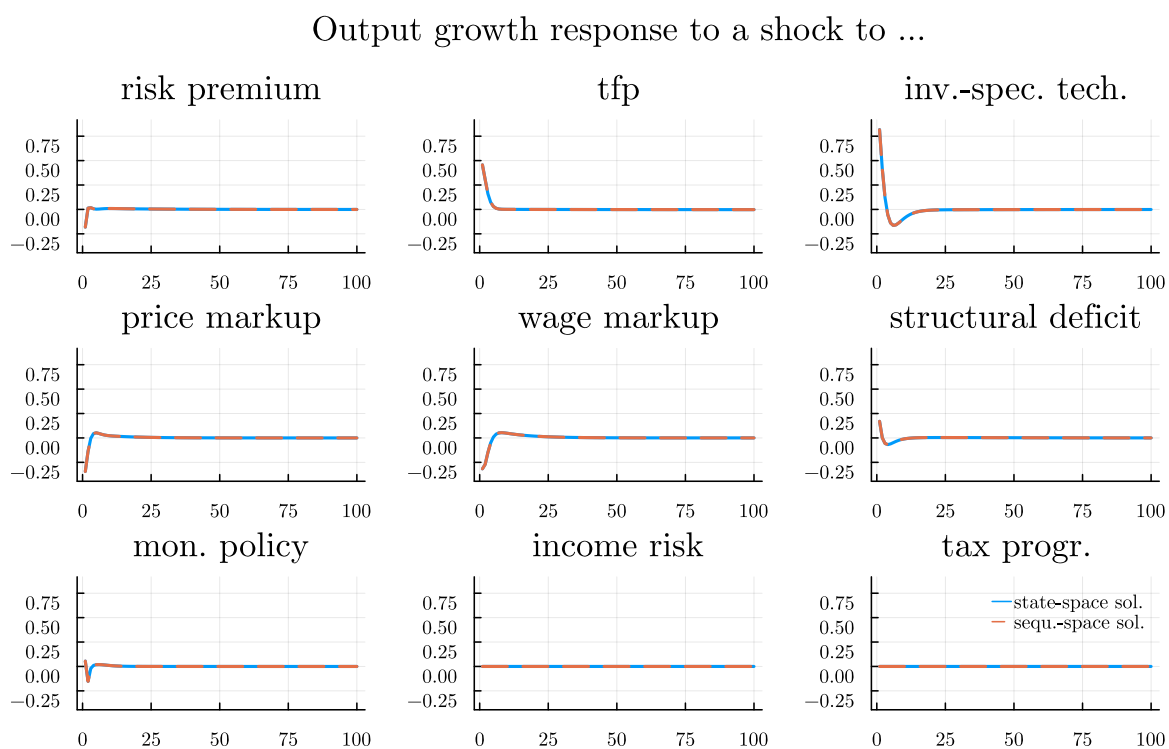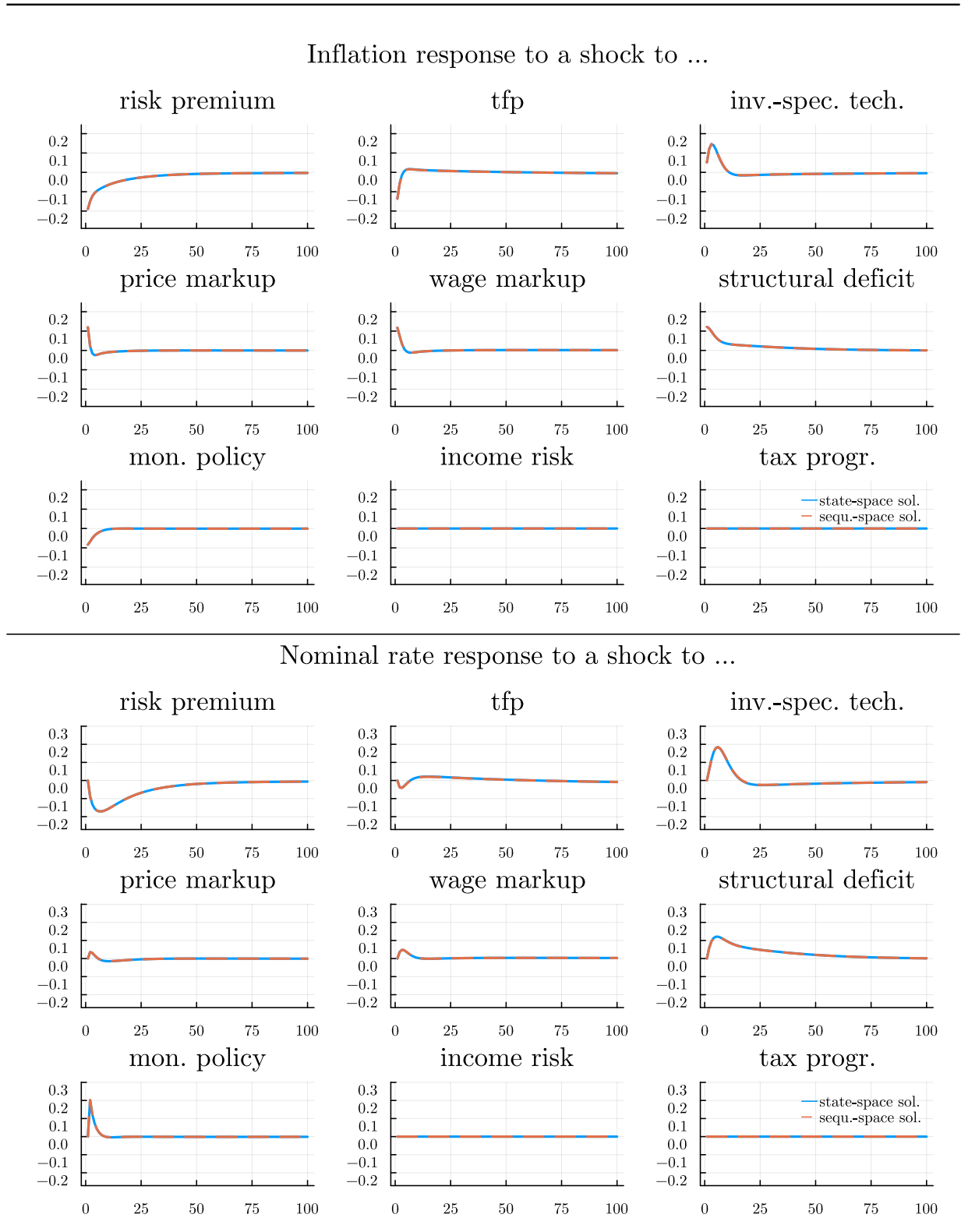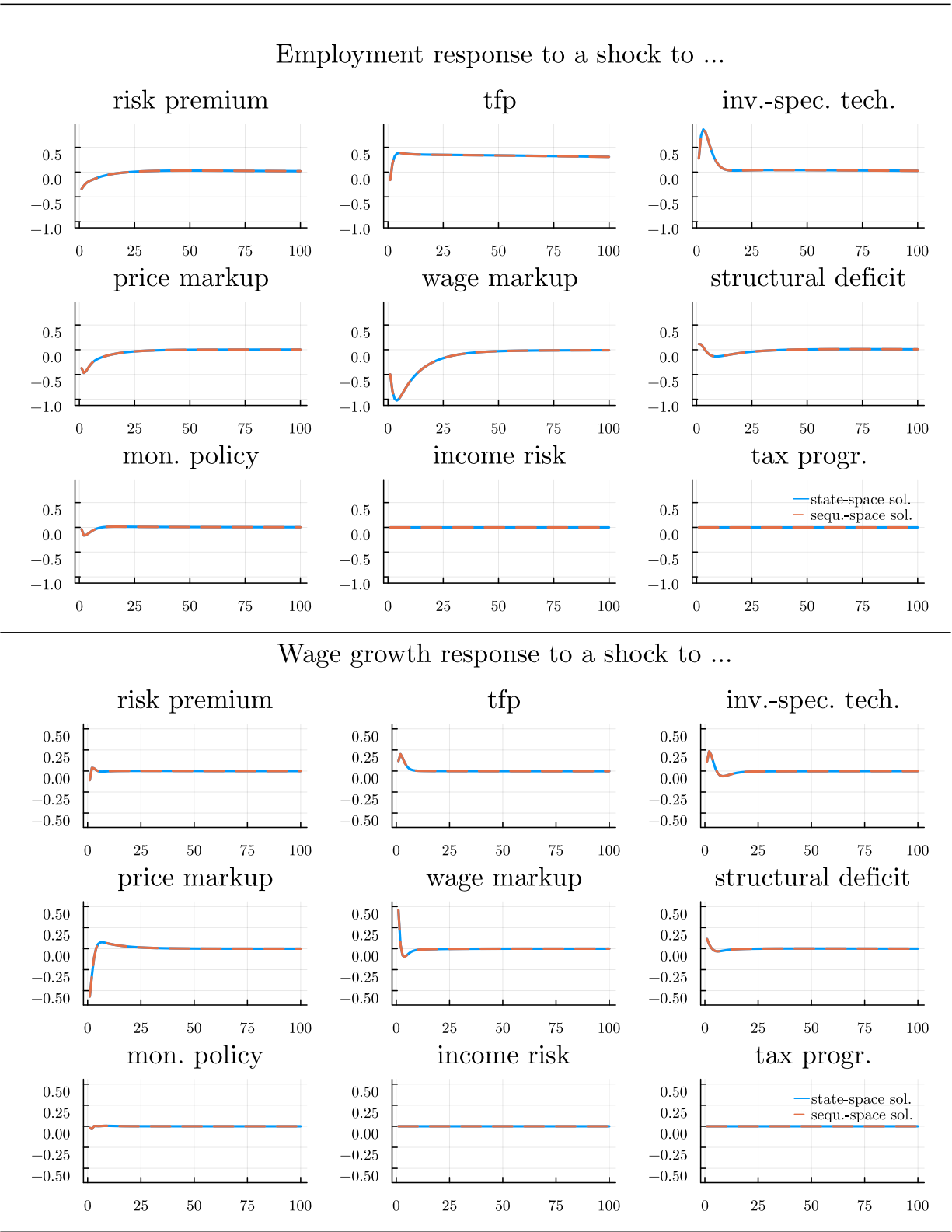

Output growth response to a shock to ...

Notes: The figure shows the impulse response to the various shocks in the HANK model, comparing a sequence-space solution (red dashed line) to our state-space solution (blue solid).

Figure C.18: Comparison of IRFs across solution methods (HANK model)
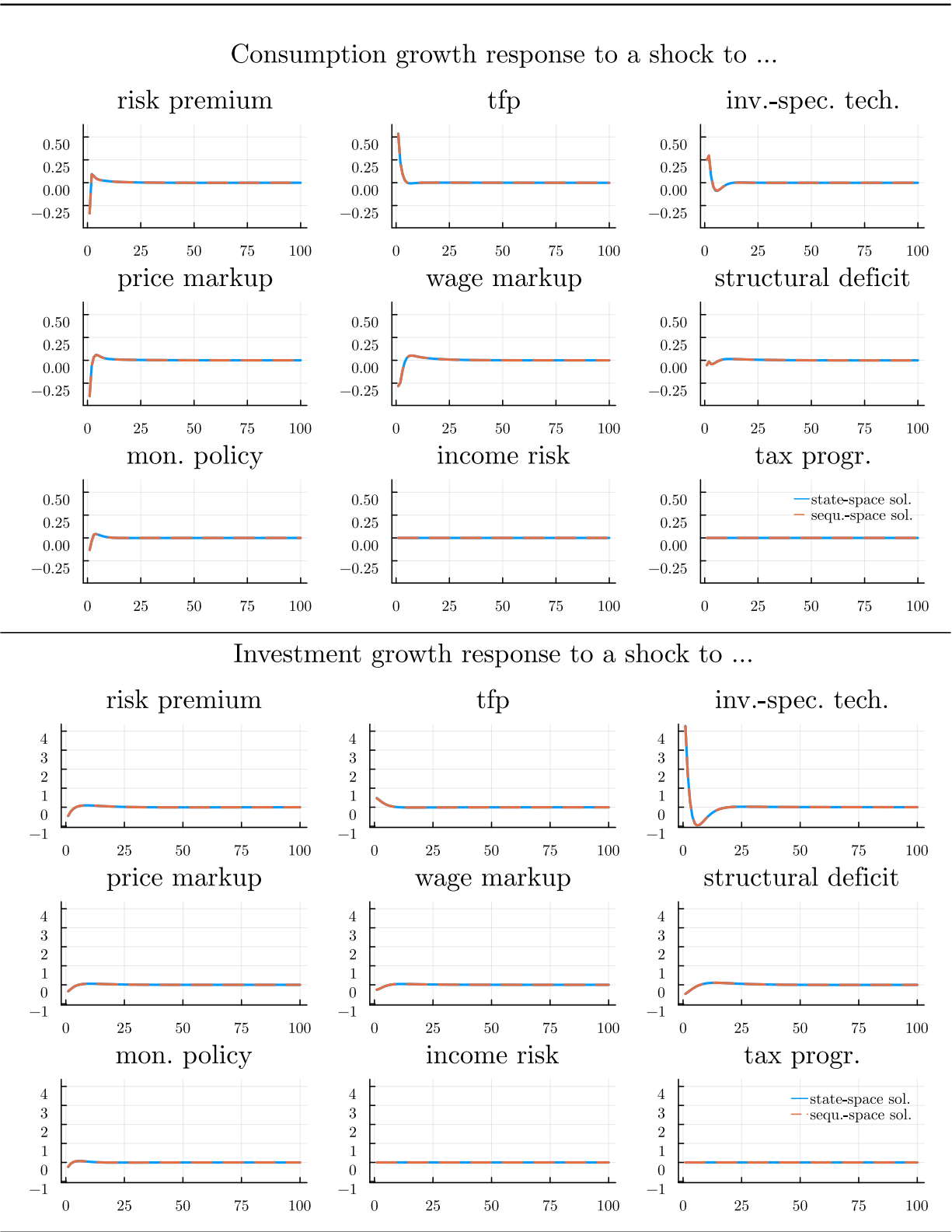
Consumption growth response to a shock to ...

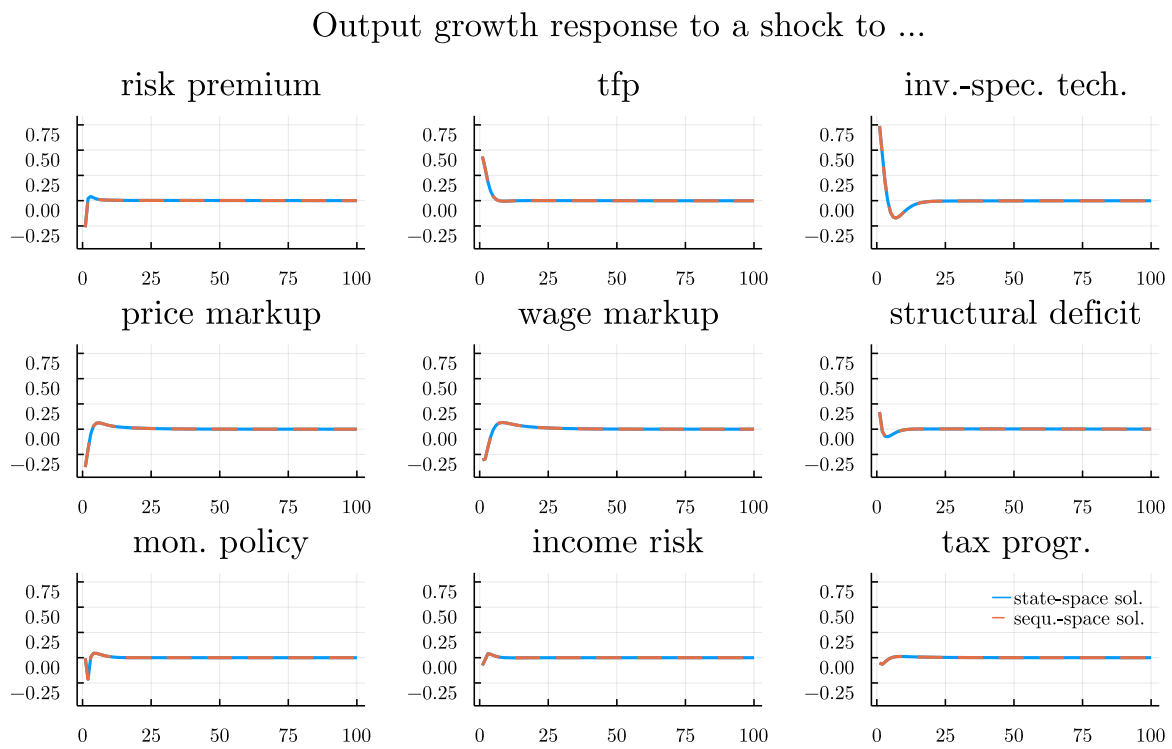Investment growth response to a shock to ...

Notes: The figure shows the impulse response to the various shocks in the HANK model, comparing a sequence-space solution (red dashed line) to our state-space solution (blue solid).

Figure C.19: Comparison of IRFs across solution methods (HANK model)

Inflation response to a shock to ...
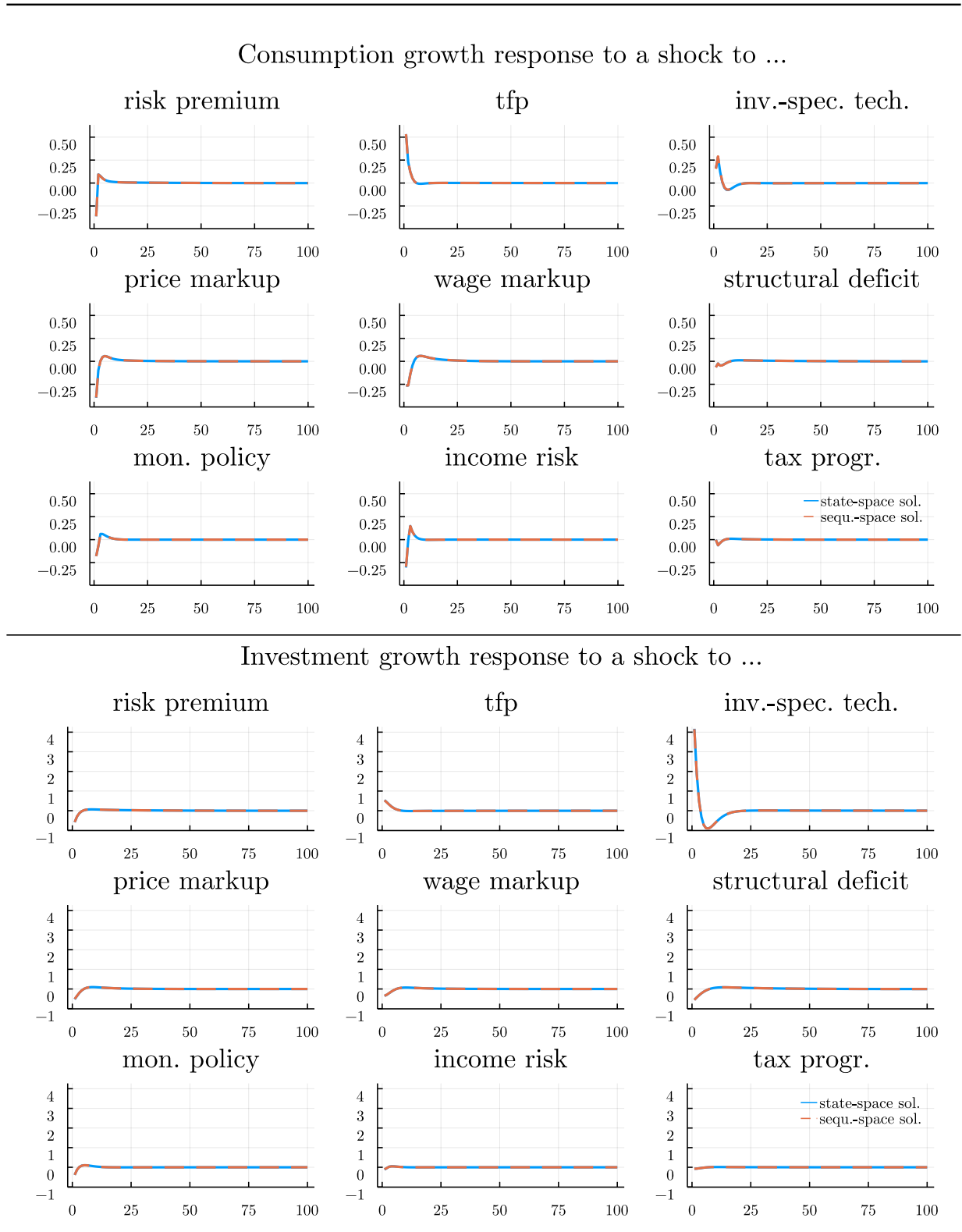

Nominal rate response to a shock to ...

Notes: The figure shows the impulse response to the various shocks in the HANK model, comparing a sequence-space solution (red dashed line) to our state-space solution (blue solid).

Figure C.20: Comparison of IRFs across solution methods (HANK model)

Employment response to a shock to ...
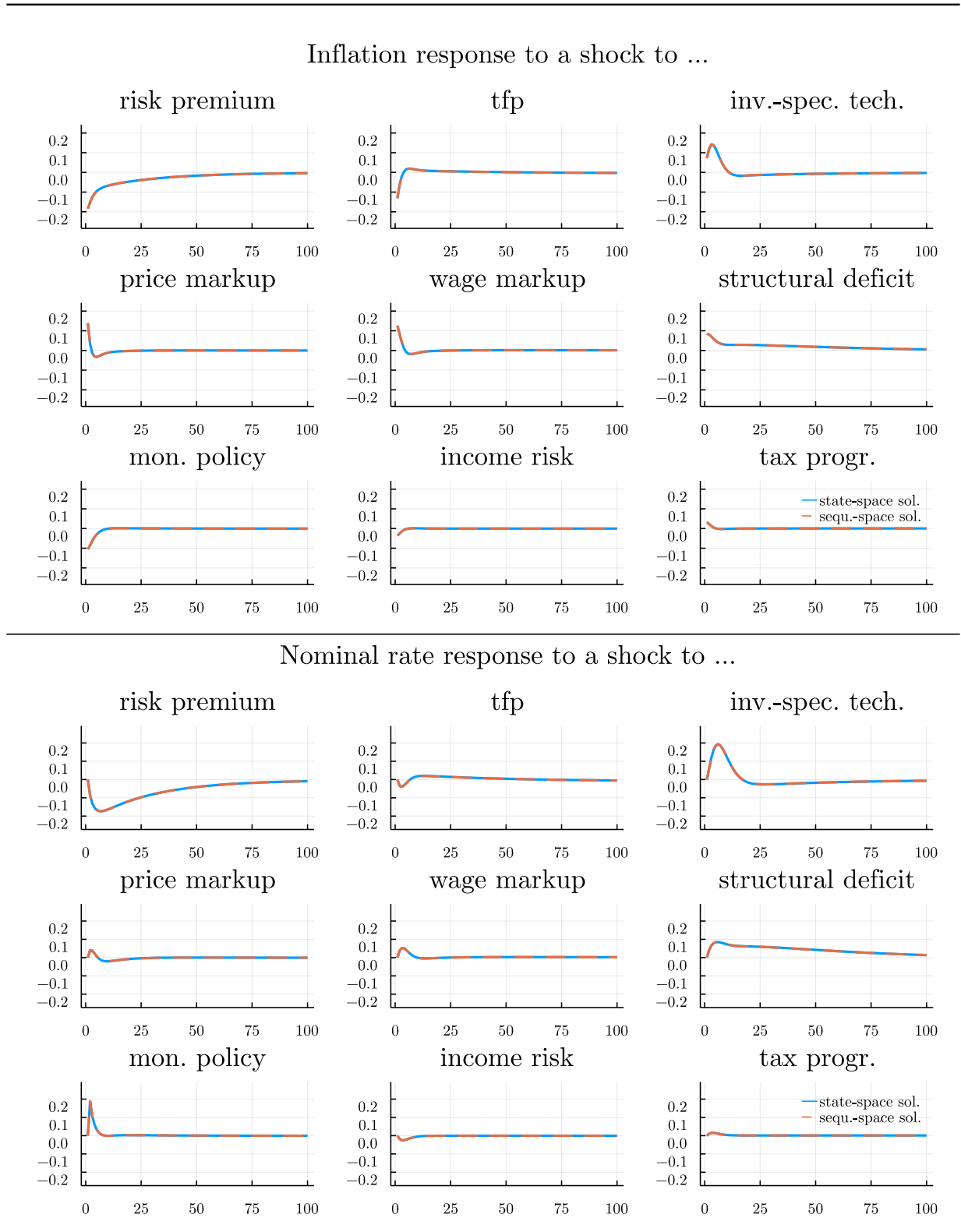

Wage growth response to a shock to ...

Notes: The figure shows the impulse response to the various shocks in the HANK model, comparing a sequence-space solution (red dashed line) to our state-space solution (blue solid).

Figure C.21: Comparison of IRFs across solution methods (HANK-X model)

Output growth response to a shock to ...

Notes: The figure shows the impulse response to the various shocks in the HANK-X model, comparing a sequence-space solution (red dashed line) to our state-space solution (blue solid).

Figure C.22: Comparison of IRFs across solution methods (HANK-X model)

Consumption growth response to a shock to ...

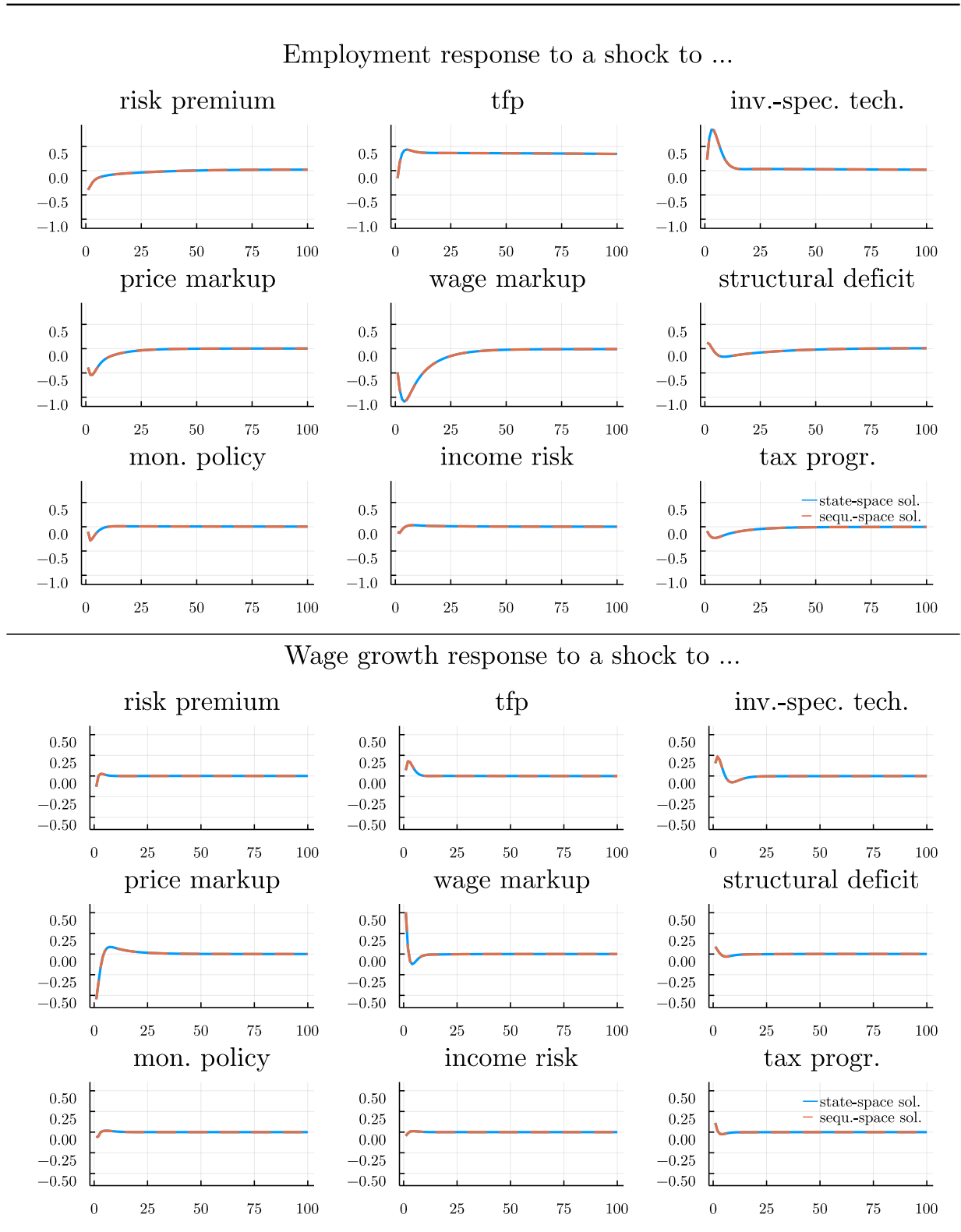Investment growth response to a shock to ...

Notes: The figure shows the impulse response to the various shocks in the HANK-X model, comparing a sequence-space solution (red dashed line) to our state-space solution (blue solid).

Figure C.23: Comparison of IRFs across solution methods (HANK-X model)

Inflation response to a shock to ...
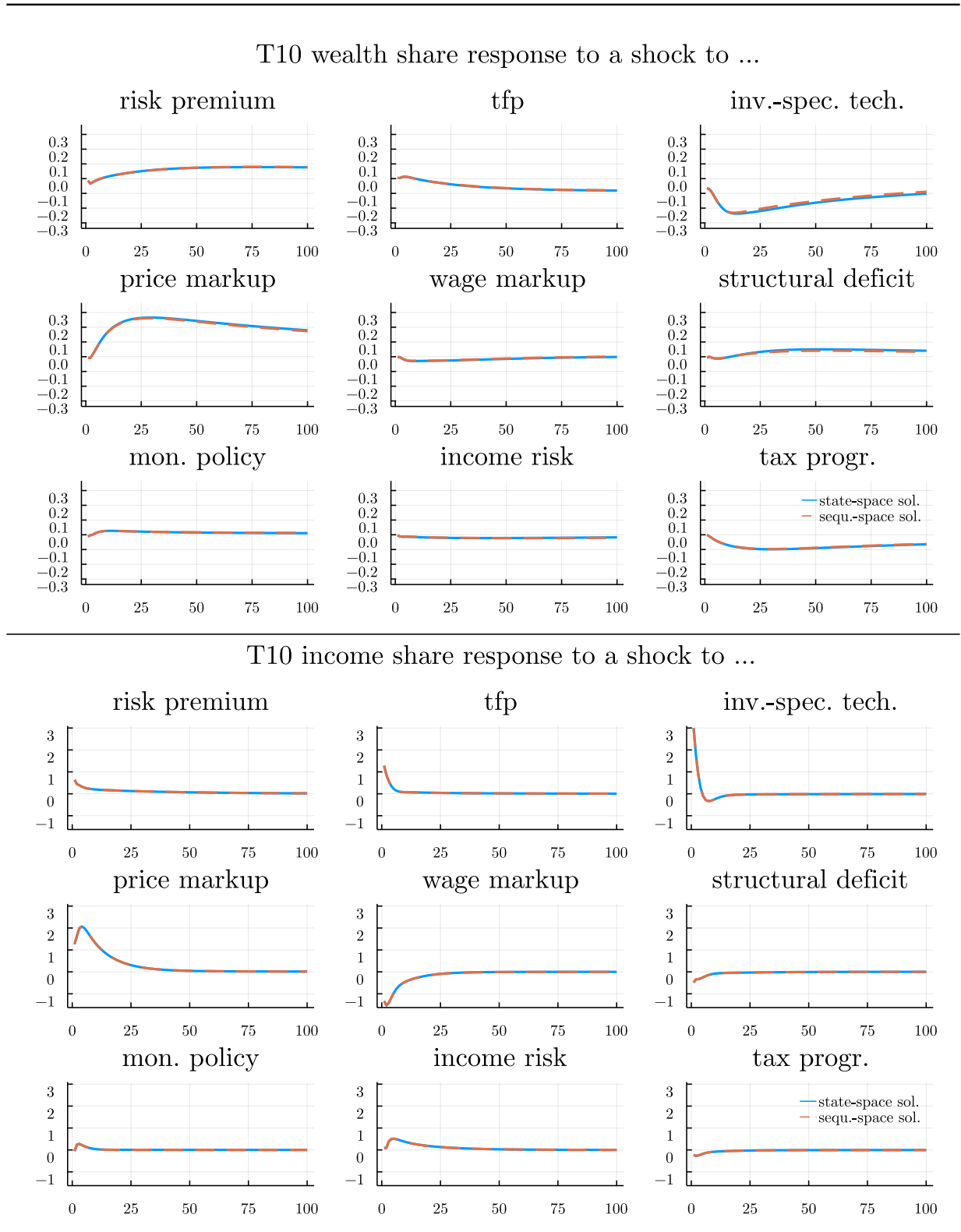

Nominal rate response to a shock to ...

Notes: The figure shows the impulse response to the various shocks in the HANK-X model, comparing a sequence-space solution (red dashed line) to our state-space solution (blue solid).

Figure C.24: Comparison of IRFs across solution methods (HANK-X model)

Employment response to a shock to ...
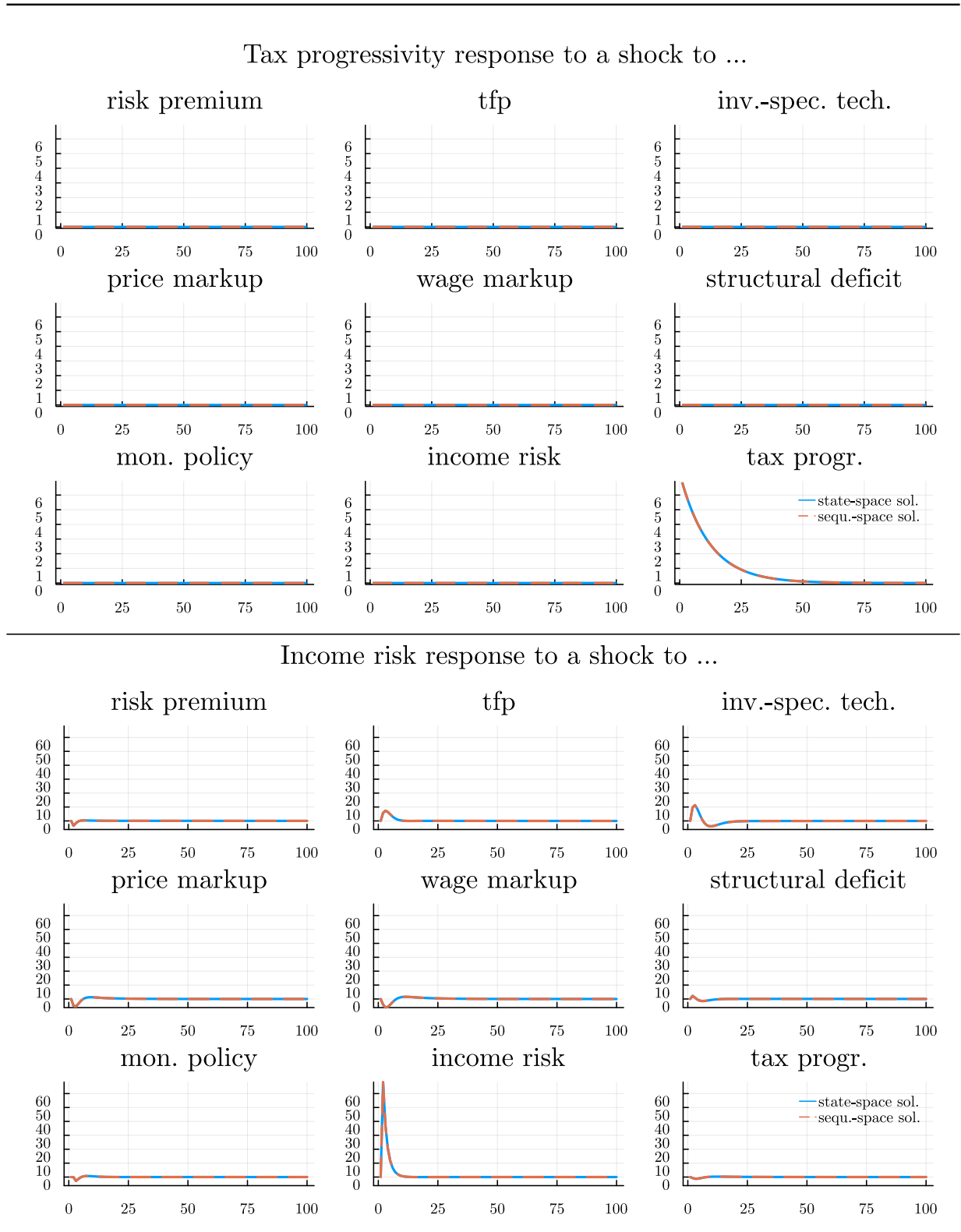

Wage growth response to a shock to ...

Notes: The figure shows the impulse response to the various shocks in the HANK-X model, comparing a sequence-space solution (red dashed line) to our state-space solution (blue solid).

Figure C.25: Comparison of IRFs across solution methods (HANK-X model)

T10 wealth share response to a shock to ...

T10 income share response to a shock to ...

Notes: The figure shows the impulse response to the various shocks in the HANK-X model, comparing a sequence-space solution (red dashed line) to our state-space solution (blue solid).

Figure C.26: Comparison of IRFs across solution methods (HANK-X model)

Tax progressivity response to a shock to ...

Income risk response to a shock to ...

Notes: The figure shows the impulse response to the various shocks in the HANK-X model, comparing a sequence-space solution (red dashed line) to our state-space solution (blue solid).