STAFF REPORTS

NO. 1071
AUGUST 2023

Estimating HANK for Central Banks

Sushant Acharya | William Chen | Marco Del Negro | Keshav Dogra | Aidan Gleich | Shlok Goyal | Ethan Matlin | Donggyu Lee | Reca Sarfati | Sikata Sengupta

FEDERAL RESERVE BANK of NEW YORK

Estimating HANK for Central Banks

Sushant Acharya, William Chen, Marco Del Negro, Keshav Dogra, Aidan Gleich, Shlok Goyal,
Ethan Matlin, Donggyu Lee, Reca Sarfati, and Sikata Sengupta

Federal Reserve Bank of New York Staff Reports, no. 1071
August 2023
https://doi.org/10.59576/sr.1071

Abstract

We provide a toolkit for efficient online estimation of heterogeneous agent (HA) New Keynesian (NK) models based on Sequential Monte Carlo methods. We use this toolkit to compare the out-of-sample forecasting accuracy of a prominent HANK model, Bayer et al. (2022), to that of the representative agent (RA) NK model of Smets and Wouters (2007, SW). We find that HANK's accuracy for real activity variables is notably inferior to that of SW. The results for consumption in particular are disappointing since the main difference between RANK and HANK is the replacement of the RA Euler equation with the aggregation of individual households' consumption policy functions, which reflects inequality.

JEL classification: C11, C32, D31, E32, E37, E52
Keywords: HANK, Bayesian inference, sequential Monte Carlo methods

---

Del Negro, Dogra, Gleich, Lee, Sengupta: Federal Reserve Bank of New York (emails: marco.delnegro@ny.frb.org, keshav.dogra@ny.frb.org, donggyu.lee@ny.frb.org). Acharya: Bank of Canada, CEPR (email: sushant.acharya.econ@gmail.com). Chen, Sarfati: MIT Economics (emails: wyc1@mit.edu, sarfati@mit.edu). Goyal, Matlin: Harvard University. The authors thank participants at the XXV Annual Conference of the Central Bank of Chile, "Heterogeneity in Macroeconomics: Implications for Monetary Policy," for which this paper was written. They also thank Markus Kirchner (discussant) for helpful comments, Mikkel Plagborg-Møller for useful feedback, Pranay Gundam, Ramya Nallamotu, and Brian Pacula for research assistance, and Christian Bayer, Benjamin Born, and Ralph Luetticke for sharing their code, data, and results.

1 Introduction

Central banks are very interested in investigating questions surrounding inequality and its relationship with monetary policy, arguably for very good reasons. First of all, inequality has become a central issue in many countries. It is therefore important to ask how central bank policies affect inequality. Second, even if central bankers were not concerned with the answer to the above question, they ought to be concerned with the fact that inequality changes the transmission mechanism of monetary policy, as forcefully argued in Kaplan et al. (2018) and Ahn et al. (2018). Several central banks have indeed shown interest in these topics (in fact, the title of the conference on which this volume is based is "Heterogeneity in Macroeconomics: Implications for Monetary Policy") and a few have begun to develop models that speak to the interaction of monetary policy and inequality, such as heterogeneous agent New Keynesian (HANK) models that follow the seminal work by Kaplan et al. (2018).

Models serve many purposes, and for some of these purposes a model's ability to fit the data—that is, to adequately describe the data from a quantitative point of view—is important, especially for central banks. After all, the popularity of representative agent DSGE models such as Smets and Wouters (2007, henceforth, SW) since the beginning of the century is largely due to these models' ability to forecast with an accuracy that is at least comparable to that of other models previously used in central banks, such as vector autoregressions. Even if forecasting is not the main purpose of a model—and arguably it is not the main purpose of DSGE models—it is a way to test its reliability in providing answers to quantitative questions: forecasting accuracy lends quantitative credibility.

These considerations prompt us to ask: What is the forecasting accuracy of HANK models? To the extent that these models have a more realistic transmission mechanism than representative agent models, one would hope that this translates into a better forecasting performance. This is particularly true for aggregate consumption, since the main difference between SW-type DSGEs and HANK models is the replacement of the representative agent Euler equation, which determines consumption in standard DSGEs, with the aggregation of individual households' consumption policy functions. These consumption policy functions depend, among other things, on the wealth distribution in the economy—that is, on inequality. This is the first paper to our knowledge that provides an assessment of the out-of-sample forecasting accuracy of HANK models.

From a computational point of view, the task of performing an out-of-sample forecasting accuracy exercise is not trivial, as it involves estimating a HANK model over and over, for each of the several vintages of data for which we want to compute forecasts. Concretely, our forecasting exercise begins in the first quarter of 2000 and ends in the last quarter of 2019, for a total of 80 periods. For each period we estimate the model using Bayesian methods—the same approach used by SW and much of the HANK literature.

Each estimation is very costly in computational terms for HANK if one calculates the likelihood using the Kalman filter, since these models have a very large state space that includes the distribution of wealth (both liquid and illiquid, in a two-asset HANK model) across households.[1]

All of the growing literature estimating HANK models using Bayesian methods (see Winberry, 2018; Auclert et al., 2021; Bayer et al., 2022; and Lee, 2021, among others) use the standard Markov Chain Monte Carlo approach followed in the representative DSGE model literature to obtain draws from the posterior distribution (e.g., An and Schorfheide, 2007) and featured in popular packages such as `Dynare`. This approach has two drawbacks. First, it cannot be naturally parallelized, being a Markov Chain-based algorithm. Second, one has to start from scratch every new estimation. For example, if one just estimated the model up to 2000Q1 and then added only one more quarter of data, with Markov Chain methods one has to start the Markov Chain anew even though the posterior distribution may not be all that different.

This paper deviates from this trend and uses a Monte Carlo method that can be readily parallelized: Sequential Monte Carlo. This parallelization makes it feasible to estimate models even when each likelihood computation takes a substantial amount of time. This method has another crucial advantage—namely, that models can be estimated "online." What online estimation means is that the swarm of particles describing the posterior distribution computed for the estimation up to 2000Q1 can be used to jump-start the estimation with one or more quarters of data, thereby making it considerably faster. This online feature is what makes repeated estimation, and therefore our forecasting accuracy exercise, possible.[2] While these methods are not new (see Cai et al., 2021), one contribution of this paper is to explain how and why they work to an audience with little or no background in Monte Carlo methods. Accordingly, this paper may serve as a blueprint for central bank researchers planning to estimate HANK models and use them in routine policy analysis and forecasting exercises. We also plan to share the code used in our forecasting exercise on `GitHub`.

As anticipated above, the other contribution of this paper is to provide a forecasting accuracy assessment of a HANK model. While several HANK models have been developed, we use that of Bayer et al. (2022, henceforth, BBL) in this paper. We do this because in their frontier contribution the authors put particular care in the empirical fit of their model, making sure that they include all the shocks and frictions that make SW-type models empirically successful. In other words, the BBL model is the closest thing to a HANK version of SW. We then ask: Does this model forecast macro time series better than the original SW? Unfortunately, the answer from our preliminary investigation is no. For some series such as inflation, the forecasting performance is similar. For other series, notably for consumption growth, the accuracy for the

[1] The advantage of the so-called sequence-space Jacobian approach to solving and estimating HANK models championed by Auclert et al. (2021) is that it circumvents the issue of the large state space associated with carrying around a set of distributions.

[2] The methods described in this paper can be used in the context of limited information approaches such as those used by Hagedorn et al. (2018), who estimate a HANK model using impulse-response matching as in Christiano et al. (2005).

HANK model is much worse than for the representative agent model, which is particularly disappointing for the reasons discussed above.

What are the reasons for, and the implications of, the relatively worse forecasting performance of this HANK model compared to SW? We suspect that one key reason is that many parameters in HANK—namely those affecting the model's steady state—are still calibrated. This is not necessarily a philosophical choice on the part of the HANK modelers; rather, it is done because recomputing the steady state is extremely costly. If this suspicion is correct, these findings pose a computational challenge to HANK researchers interested in estimation. Certainly the findings should be interpreted as a motivation to do more research on HANK models, as opposed to sticking with representative agent models. Inequality is one of the critical issues of our time—no matter the forecasting performance of HANK models. The fact that the latter can be improved is a stimulus for further efforts, especially from central bank researchers who want to use these models for quantitative purposes.

In the remainder of the paper, section 2 presents BBL's model and solution approach so to make the paper self-contained, section 3 describes the Sequential Monte Carlo algorithm and the online estimation approach used to perform the forecasting exercise, section 4 discusses the results, and section 5 concludes.

2 Model

This paper employs the HANK model developed by BBL, which augments standard New Keynesian DSGE models, such as those presented in SW or Christiano et al. (2005), with heterogeneous agents and incomplete markets. The model incorporates standard shocks and frictions utilized in DSGE models. Moreover, it is also capable of reproducing notable characteristics of household heterogeneity deemed important in the literature, such as heterogeneous wealth and income composition and the presence of wealthy hand-to-mouth households. BBL show that when the model is estimated on aggregate data, it can reproduce the business cycle dynamics of aggregate data as well as of observed U.S. inequality. As the model is taken entirely from BBL, we will provide only a brief description of the model environment below in order to make the paper self-contained.[3]

[3]We use the version of the model available at `https://github.com/BASEforHANK/HANK_BusinessCycleAndInequality` as of June 2022, when we began this project. The latest version of the model, as described in Bayer et al. (2022), has two minor differences from the model adopted in this paper. First, the latest version of the model has a different formulation for the liquid asset return. Specifically, BBL assume that entrepreneurs sell claims to a fraction of profits as liquid shares, and the liquid asset return is the weighted average of the interest on government bonds and the return on these shares, which consists of profit payouts and the realized capital gain. In addition, BBL assume that time-varying income risks respond to output growth, which makes income risks either procyclical or countercyclical. These modifications allow BBL to better explain inequality series and their income risk estimates with their model.

2.1 Households

There exists a unit mass of infinitely-lived households, indexed by i, that maxmize their lifetime utility

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_{it}, n_{it} | h_{it}), \tag{1}$$

where β denotes the discount factor, c_{it} denotes consumption, and n_{it} denotes hours worked. The instantaneous utility function $u(\cdot)$ follows Greenwood et al. (1988)

$$u(c_{it}, n_{it} | h_{it}) = \frac{x_{it}^{1-\xi} - 1}{1-\xi} \ , \ x_{it} = c_{it} - h_{it}^{1-\bar{\tau}^P} \frac{n_{it}^{1+\gamma}}{1+\gamma}, \tag{2}$$

where $\bar{\tau}^P$ is the steady-state level of tax progressivity, ζ is the coefficient of relative risk aversion, γ is the inverse of the Frisch elasticity, and h_{it} is idiosyncratic labor productivity. There are two types of households: workers ($h_{it} \neq 0$) and entrepreneurs ($h_{it} = 0$). Idiosyncratic productivity $h_{it} = \dfrac{\tilde{h}_{it}}{\int \tilde{h}_{it} di}$ evolves as follows:

$$\tilde{h}_{it} = \begin{cases} \exp\left(\rho_h \log \tilde{h}_{it-1} + \epsilon_{it}^h\right) & \text{with probability } 1-\zeta \text{ if } h_{it-1} \neq 0, \\ 1 & \text{with probability } \iota \text{ if } h_{it-1} = 0, \\ 0 & \text{else.} \end{cases} \tag{3}$$

The above equation implies that workers become entrepreneurs with the probability ζ or continue to be workers with the probability $1-\zeta$. While being workers, labor productivity h_{it} evolves according to an AR(1) process in logs with an autocorrelation coefficient ρ_h. The shocks ϵ_{it}^h are normally distributed with variance $\sigma_{h,t}^2$. This variance changes over time according to the process

$$\sigma_{h,t}^2 = \bar{\sigma}_h^2 \exp(\hat{s}_t), \tag{4}$$

$$\hat{s}_{t+1} = \rho_s \hat{s}_t + \epsilon_t^{\sigma}, \tag{5}$$

where the shocks ϵ_t^{σ} follow a normal distribution with zero mean and the standard deviation σ_s.[4]

Workers earn wage income $w_t h_{it} n_{it}$, where w_t is the real wage paid to households by labor unions. In addition, rents from unions Π_t^U are equally distributed among workers. Entrepreneurs become workers with the probability ι or maintain their entrepreneur status with the probability $1-\iota$. When entrepreneurs become workers, their productivity becomes one. Entrepreneurs do not supply labor and, instead, receive profits Π_t^F generated in the firm sector, except for rents of unions.

Markets are incomplete, and households self-insure against income risks by saving in two types of assets: illiquid capital and liquid bonds. Capital as an asset is illiquid in the sense that only a fraction λ of households

[4]The latest version of BBL assumes that the level of income risks is affected by the output growth, i.e., $\hat{s}_{t+1} = \rho_s \hat{s}_t + \Sigma_Y \dfrac{Y_{t+1}}{Y_t} + \epsilon_t^{\sigma}$. Depending on the sign of the coefficient Σ_Y, idiosyncratic income risks are either pro- or countercyclical in the model. This setup allows BBL to better capture the dynamics of income risks with their model.

are allowed to adjust their capital holdings in each period. In contrast, households can freely adjust their liquid bond holdings.

The household's budget constraint can be written as

$$c_{it} + b_{it+1} + q_t k_{it+1} = b_{it}\frac{R_{it}}{\pi_t} + (q_t + r_t)k_{it} + (1-\tau_t^L)(w_t h_{it} n_{it})^{1-\bar{\tau}_t^P}$$
$$+ \mathbb{1}_{h_{it}\neq 0}(1-\tau_t)\Pi_t^U + \mathbb{1}_{h_{it}=0}(1-\tau_t^L)(\Pi_t^E)^{1-\bar{\tau}^P} \ , \quad k_{it+1} \geq 0 \ , \quad b_{it+1} \geq \underline{\mathrm{b}} \ , \tag{6}$$

where b_{it} is real liquid bonds, k_{it} is capital stock, q_t is the price of capital, r_t is the dividend on capital holdings, $\pi_t = \frac{P_t}{P_{t-1}}$ is the gross inflation rate, and $\underline{\mathrm{b}} < 0$ is an exogenous borrowing limit. Workers' labor income and entrepreneurs' profit income are taxed progressively. The two tax rates, τ_t^L and τ_t^P, determine the degree of tax progressivity. The union profit is taxed uniformly at the average tax rate τ_t. Finally, the return on the liquid assets R_{it} depends on whether households are borrowers:

$$R_{it} = \begin{cases} A_t R_t^b & \text{if } \ b_{it} \geq 0 \\ A_t R_t^b + \bar{R} & \text{if } \ b_{it} < 0. \end{cases} \tag{7}$$

The coefficient A_t is the so-called "risk-premium shock" (see SW), which reflects intermediation efficiency, and $\bar{R}$ is the borrowing premium. R_t^b is the nominal interest rate on government bonds, which is determined by the monetary authority.[5]

Since households may or may not be able to adjust their illiquid asset holdings, the household's problem is characterized by three functions: the value function V_t^a when households are allowed to adjust their capital holdings, the function V_t^n when households are not allowed to adjust, and the expected value in the next period $\mathbb{W}_{t+1}$:

$$V_t^a(b,k,h) = \max_{b'_a, k'} u[(x(b, b'_a, k, k', h))] + \beta \mathbb{E}_t \mathbb{W}_{t+1}(b'_a, k', h'), \tag{8}$$

$$V_t^n(b,k,h) = \max_{b'_n} u[(x(b, b'_n, k, k, h))] + \beta \mathbb{E}_t \mathbb{W}_{t+1}(b'_n, k', h'), \tag{9}$$

$$\mathbb{W}_{t+1}(b', k', h') = \lambda V_{t+1}^a(b', k', h') + (1-\lambda)V_{t+1}^n(b', k', h'), \tag{10}$$

where $x(b, b', k, k', h) = c(b, b', k, k', h) - h^{1-\bar{\tau}^P}\frac{n(w)^{1+\gamma}}{1+\gamma}$ is the household's composite demand for goods and leisure.[6] Maximization is subject to the budget constraint described above.

[5] In their model, Bayer et al. (2022) assume that entrepreneurs sell claims to a fraction ω_Π of profits at the price of q_t^Π as liquid shares and these shares become a part of the household's liquid asset portfolio as well. Thus, the liquid asset return is the weighted average of the return on government bonds and the return on profit shares. Consequently, dynamics of profit shares also affect the liquid asset return in the model.

[6] Because of the specific form of GHH preference used in the model, all workers supply the same amount of labor, depending on the level of the real wage only.

2.2 Firms

The firm sector comprises four types of firms: 1) final goods producers, 2) intermediate goods producers, 3) capital producers, and 4) labor packers. Final goods producers transform intermediate goods into final consumption goods. Intermediate goods producers produce differentiated goods using capital and labor service as inputs. Capital producers transform final goods into new capital stock, subject to adjustment frictions, and rent out capital to intermediate goods producers. Labor packers combine differentiated labor supplied by unions and rent out homogeneous labor services to intermediate goods producers. Intermediate goods producers and unions operate under a monopolistically competitive environment and set prices subject to nominal rigidity á la Calvo (1983).

2.2.1 Final goods producers

Final goods producers combine differentiated intermediate goods and make final consumption goods according to a CES aggregation techonology

$$Y_t = \left(\int y_{jt}^{\frac{\eta_t - 1}{\eta_t}} dj \right)^{\frac{\eta_t}{\eta_t - 1}}, \tag{11}$$

where y_{jt} is intermediate good j and η_t is the time-varying elasticity of substitution. Profit maximization yields the following individual good demand and the aggregate price index:

$$y_{jt} = \left(\frac{p_{jt}}{P_t} \right)^{-\eta_t} Y_t \tag{12}$$

$$P_t = \int p_{jt}^{1-\eta_t} dj, \tag{13}$$

where p_{jt} is individual good j's price.

2.3 Intermediate goods producers

There is a continuum of intermediate good firms, indexed by j, that produce differentiated goods using capital and labor services, according to a constant return-to-scale production function

$$y_{j,t} = Z_t N_{jt}^{\alpha} (u_{jt} K_{jt})^{1-\alpha}, \tag{14}$$

where α is the labor share in production, Z_t is total factor productivity that follows a AR(1) process in logs, N_{jt} is labor input, and $u_{jt} K_{jt}$ is capital input with the utilization rate u_{jt}. The capital depreciation rate depends on the degree of utilization according to $\delta(u_{jt}) = \delta_0 + \delta_1(u_{jt} - 1) + \frac{\delta_2}{2}(u_{jt} - 1)^2$. First-order conditions associated with the cost minimization are as follows:

$$w_t^F = \alpha \mathrm{mc}_{jt} Z_t \left(\frac{u_{jt} K_{jt}}{N_{jt}} \right)^{1-\alpha} \tag{15}$$

$$r_t + q_t \delta(u_{jt}) = u_{jt}(1-\alpha)\mathrm{mc}_{jt} Z_t \left(\frac{N_{jt}}{u_{jt} K_{jt}} \right)^{\alpha} \tag{16}$$

$$q_t[\delta_1 + \delta_2(u_{jt} - 1)] = (1-\alpha)\mathrm{mc}_{jt} Z_t \left(\frac{N_{jt}}{u_{jt} K_{jt}} \right)^{\alpha}, \tag{17}$$

where mc_{jt} is the marginal cost of production of firm j. Since the production function exhibits constant return-to-scale, the above optimality conditions imply that marginal costs are identical across producers, i.e., $\mathrm{mc}_{jt} = \mathrm{mc}_t$.

Firms operate under monopolistically competitive environments and set prices for their goods subject to price adjustment frictions á la Calvo (1983): only a fraction $1 - \lambda_Y$ of firms can adjust their prices, while the rest index their prices to the steady-state inflation rate $\bar{\pi}$. Firms maximize the present value of real profits

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \lambda_Y^t (1 - \tau_t^L) Y_t^{1-\tau_t^P} \left\{ \left(\frac{p_{jt}\bar{\pi}^t}{P_t} - \mathrm{mc}_t \right) \left(\frac{p_{jt}\bar{\pi}^t}{P_t} \right)^{-\eta_t} \right\}^{1-\tau_t^P}. \tag{18}$$

The corresponding optimality condition, with a first-order approximation, implies the following Phillips curve:

$$\log\left(\frac{\pi_t}{\bar{\pi}}\right) = \beta \log\left(\frac{\pi_{t+1}}{\bar{\pi}}\right) + \kappa_Y \left(\mathrm{mc}_t - \frac{1}{\mu_t^Y} \right), \tag{19}$$

where $\kappa = \dfrac{(1-\lambda_Y)(1-\lambda_Y\beta)}{\lambda_Y}$ is the slope of the Phillips curve and $\mu_t^Y = \dfrac{\eta_t}{\eta_t - 1}$ is the target markup. The target markup follows an AR(1) process with shock $\epsilon_t^{\mu^Y}$.

2.3.1 Capital producers

Capital producers transform final goods into new capital stock, subject to adjustment frictions, while taking the price of capital q_t as given. They maximize

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t I_t \left\{ q_t \Psi_t \left[1 - \frac{\phi}{2} \left(\log \frac{I_t}{I_{t-1}} \right)^2 \right] - 1 \right\}, \tag{20}$$

where ϕ governs the degree of investment adjustment frictions and Ψ_t represents marginal efficiency of investment á la Justiniano et al. (2011), which follows an AR(1) process in logs with innovation ϵ_t^{Ψ}. Up to first order, the optimality condition for the maximization problem is

$$q_t \Psi_t \left[1 - \phi \log \frac{I_t}{I_{t-1}} \right] = 1 - \beta \mathbb{E}_t \left[q_{t+1} \Psi_{t+1} \phi \log\left(\frac{I_{t+1}}{I_t} \right) \right]. \tag{21}$$

Finally, the law of motion for aggregate capital is given by

$$K_t - (1-\delta(u_t))K_{t-1} = \Psi_t \left[1 - \frac{\phi}{2}\left(\log \frac{I_t}{I_{t-1}}\right)^2\right]. \tag{22}$$

2.3.2 Unions and labor packers

There exists a unit mass of unions, indexed by l, who purchase labor services from workers and sell a different variety of labor to labor packers. Labor packers combine a different variety of labor into homogeneous labor input according to the following CES aggregation technology:

$$N_t = \left(\int \hat{n}_{lt}^{\frac{\zeta_t - 1}{\zeta_t}} dj\right)^{\frac{\zeta_t}{\zeta_t - 1}}, \tag{23}$$

where $\hat{n}_{lt}$ is a variety l of labor service and ζ_t is the elasticity of substitution. Labor packers' cost minimization implies the following demand for each variety l of labor service:

$$\hat{n}_{lt} = \left(\frac{W_{lt}}{W_t^F}\right)^{-\zeta_t} N_t, \tag{24}$$

where W_{lt} is the nominal wage set by union l, and W_t^F is the nominal wage at which labor packers sell labor input to intermediate goods producers.

Unions have monoplistic power and maximize their stream of profits by setting prices w_{lt} for their labor variety, subject to nominal rigidity á la Calvo (1983). Specifically, only $1-\lambda_w$ fraction of unions can adjust wages, while the rest of unions index wages to the steady-state wage inflation rate. Thus, they maximize

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \lambda_w^t \frac{W_t^F}{P_t} N_t \left\{ \left(\frac{W_{lt}^{\bar{\pi}_W^t}}{W_t^F} - \frac{W_t}{W_t^F}\right)\left(\frac{W_{lt}^{\bar{\pi}_W^t}}{W_t^F}\right)^{-\zeta_t} \right\}. \tag{25}$$

From the optimality condition for the maximization problem, we obtain under a first-order approximation the wage Phillips curve

$$\log\left(\frac{\pi_t^W}{\bar{\pi}_W}\right) = \beta \mathbb{E}_t \log\left(\frac{\pi_{t+1}^W}{\bar{\pi}_W}\right) + \kappa_\omega \left(\mathrm{mc}_t^\omega - \frac{1}{\mu_t^\omega}\right), \tag{26}$$

where $\kappa_\omega = \dfrac{(1-\lambda_w)(1-\lambda_w\beta)}{\lambda_W}$ is the slope of the Phillips curve, and $\pi_t^W \equiv \dfrac{W_t^F}{W_{t-1}^F} = \dfrac{w_t^F}{w_{t-1}^F}\pi_t^Y$ is the gross wage inflation rate with w_t and w_t^F being the real wages for households and firms, respectively. $\mathrm{mc}_t^\omega = \dfrac{w_t}{w_t^F}$ is the actual and $\dfrac{1}{\mu_t^\omega} = \dfrac{\zeta_t - 1}{\zeta_t}$ is the target markdown of wages that unions pay to households relative to wages they charge to intermediate goods producers. The target markdown follows an AR(1) process in logs that is subject to the wage markup shock ϵ_t^w.

2.4 Government

The government sector consists of a fiscal and a monetary authority. The fiscal authority issues government bonds, levies taxes, and makes government purchases. The issuance of government bonds is governed by the following rule:

$$\frac{B_{t+1}}{B_t} = \left(\frac{B_t}{\bar{B}}\right)^{-\gamma_B} \left(\frac{\pi_t}{\bar{\pi}}\right)^{\gamma_\pi} \left(\frac{Y_t}{Y_{t-1}}\right)^{\gamma_Y} D_t \ , \quad D_t = D_{t-1}^{\rho_D} \epsilon_t^D, \tag{27}$$

where D_t represents the government structural deficit, which evolves exogenously as an AR(1) process subject to the shock ϵ_t^D. The parameters γ_B, γ_π, and γ_Y represent how sensitively the deficit responds to the existing debt, the evolution of the inflation rate, and the output growth, respectively. The government also sets the average tax rate according to the rule

$$\frac{\tau_t}{\bar{\tau}} = \left(\frac{\tau_{t-1}}{\bar{\tau}}\right)^{\rho_\tau} \left(\frac{B_t}{B_{t-1}}\right)^{(1-\rho_\tau)\gamma_B^\tau} \left(\frac{Y_t}{Y_{t-1}}\right)^{(1-\rho_\tau)\gamma_Y^\tau}. \tag{28}$$

The fiscal authority ensures that the average tax rate equals the target tax rate τ_t by adjusting the level parameter τ_L:

$$\tau_t = \frac{\mathbb{E}_t\big(w_t n_{it} h_{it} + \mathbb{1}_{h_{it}=0}\Pi_t^E\big) - \tau_L \mathbb{E}_t\big(w_t n_{it} h_{it} + \mathbb{1}_{h_{it}=0}\Pi_t^E\big)^{\bar{\tau}^P}}{\mathbb{E}_t\big(w_t n_{it} h_{it} + \mathbb{1}_{h_{it}=0}\Pi_t^E\big)}. \tag{29}$$

The total tax revenue is $T_t = \tau_t\big(w_t n_{it} h_{it} + \mathbb{1}_{h_{it}=1}\Pi_t^U + \mathbb{1}_{h_{it}=0}\Pi_t^E\big)$ and government purchases are determined by the balanced budget constraint, i.e., $G_t. = B_{t+1} + T_t - R_t^b/\pi_t B_t$.

The monetary authority determines the nominal interest rate on government bonds according to the following Taylor rule with interest rate smoothing:

$$\frac{R_{t+1}^b}{R_t^b} = \left(\frac{R_t^b}{\bar{R}^b}\right)^{\rho_R} \left(\frac{\pi_t}{\bar{\pi}}\right)^{(1-\rho_R)\theta_\pi} \left(\frac{Y_t}{Y_{t-1}}\right)^{(1-\rho_R)\theta_Y} \epsilon_t^R, \tag{30}$$

where $\bar{R}^b$ is the steady-state nominal interest rate. The coefficients ϕ_π and ϕ_Y represents the sensitivity of the policy rate to the evolution of price and output gap, respectively. The parameter ρ_R represents the degree of interest rate smoothing.

2.5 Market-clearing conditions

The model has four markets; goods, labor, liquid assets, and illiquid assets. The only liquid asset in the model is government bonds.[7] Thus, the liquid asset market clearing condition is given by

$$B_{t+1} = B_{t+1}^d \equiv \int \{\lambda b_a^*(b,k,h) + (1-\lambda) b_n^*(b,k,h)\} d\phi_t(b,k,h), \tag{31}$$

where b_a^* and b_n^* are the optimal liquid asset choice of adjusting and non-adjusting households with liquid asset holding b, illiquid asset holding k, and productivity level h, respectively, and ϕ_t is the distribution of

[7] In contrast, the most recent version of BBL has two kinds of liquid assets: government bonds and profit shares.

households over the idiosyncratic state space. The left- and right-hand sides of the above equation represent the aggregate liquid asset supply and demand, respectively. Similarly, the illiquid asset, i.e., capital, market-clearing condition is given by

$$K_{t+1} = K_{t+1}^d \equiv \int \{\lambda k_a^*(b,k,h) + (1-\lambda) k_n^*(b,k,h)\} d\phi_t(b,k,h), \tag{32}$$

with k_a^* and k_n^* being the optimal capital holding choice of adjusting and non-adjusting households with liquid asset holding b, illiquid asset holding k, and productivity level h.

The labor market clears when the following equation holds:

$$\int \hat{n}_{lt} dl = D_t^w N_t = \int h_{it} n_{it} d\phi_t(b,k,h), \tag{33}$$

where $D_t^w = \int (\frac{W_{lt}}{W_t^F})^{-\zeta_t} dl$ is the dispersion of wages set by unions and N_t is the aggregate labor input. The first two items of the above equation represent the demand of intermediate goods producers for a variety of labor, while the last item is the aggregate labor variety supplied by households. Once assets and labor markets clear, the goods market also clears because of Walras's law.

2.6 Numerical method

Following Reiter (2009), BBL solve the model using a linearized solution technique. The first step is to write the equilibrium as a system of nonlinear difference equations as follows:

$$\mathbb{E}_t F(X_t^*, X_{t+1}^*) = 0, \tag{34}$$

where X_t^* is a vector of state and control variables in period t. Then, BBL linearize the above system around the non-stochastic steady-state and apply a standard perturbation method, such as those proposed by Klein (2000). However, without any further treatments, applying a standard perturbation method is infeasible since the size of the above system is very large due to many idiosyncratic state variables such as asset holdings, productivity levels, and working statuses. Thus, BBL follow Bayer and Luetticke (2020) and reduce the size of the two largest components of the system: value functions and the household distributions.

For the value functions, BBL use a discrete cosine transformation (DCT), as proposed by Bayer and Luetticke (2020). All the value functions are written as linear interpolants based on a set of nodal values, and these nodal values are represented by DCT coefficients of Chebyshev polynomials as follows:

$$\hat{\mathbb{W}}_{b/k,t}(b_i, k_j, h_l) = \sum_{p,q,r} \theta^{p,q,r}_{\mathbb{W}_{b/k},t} T_p(i) T_q(j) T_r(l), \tag{35}$$

where $\hat{W}_{b/k}$ is the partial derivative of the continuation value $\mathbb{W}$ with respect to bond b, (capital k) holdings, $T_{p/q/r}(\cdot)$ are Chebyshev polynomials, and $\theta^{p,q,r}_{\mathbb{W}}$ are the corresponding DCT coefficients. In the above expression, BBL force very small coefficients to zero in order to achieve size reduction. That is, they only keep enough of these coefficients to approximate the original value functions with a certain threshold level of precision. In perturbing the system, they perturb these coefficients instead of the function values themselves.

BBL reduce the size of the distribution in a similar way. For the distribution, they keep only marginal distributions F_t^b, F_t^k, and F_t^h in the system and use a copula $C_t(\cdot)$, a functional relationship between marginals and the joint distribution, to recover the joint distribution from marginals. Then, the copula $C_t(\cdot)$ at time t is approximated using Chebyshev polynomials

$$\hat{C}_t(F_i^b, F_j^k, F_l^h) = \sum_{p,q,r} \theta_{C,t}^{p,q,r} T_p(i) T_q(j) T_r(l), \tag{36}$$

where $\hat{C}(\cdot)$ is the deviation of the copula at time t from its steady-state counterpart. Again, BBL reduce the size of the system by keeping only a small number of DCT coefficients $\theta_C^{p,q,r}$.

After the state-space reduction, the dimension of the system decreases substantially and one can find a linearized solution rather quickly. However, for the purpose of estimation, further acceleration of the solution method is required since one needs to efficiently evaluate the model's likelihood. To this end, BBL follow Bayer and Luetticke (2020) and estimate only a subset of parameters that do not affect the households' problem. BBL first partition X^* into the part related to household choices f and the aggregate part X:

$$\mathbb{E}_t F(X_t^*, X_{t+1}^*) = \mathbb{E}_t F(f_t, X_t, f_{t+1}, X_{t+1}). \tag{37}$$

Then, they obtain the following linearized system:

$$\begin{bmatrix} A_{ff} & A_{fX} \\ A_{Xf} & A_{XX} \end{bmatrix} \begin{bmatrix} f_t \\ X_t \end{bmatrix} = -\mathbb{E}_t \begin{bmatrix} B_{ff} & B_{fX} \\ B_{Xf} & B_{XX} \end{bmatrix} \begin{bmatrix} f_{t+1} \\ X_{t+1} \end{bmatrix}. \tag{38}$$

If only the parameters that do not affect the household problem are estimated, one only needs to update A_{XX} and B_{XX} during the estimation. Since the size of aggregate blocks A_{XX} and B_{XX} is relatively small, one can update the Jacobian rather quickly.

Finally, BBL perform a further model reduction, which relies on a factor representation of the idiosyncratic model part, i.e., the part related to household choices. Once they define objects in a way such that $B_{fX} = B_{Xf} = 0$, they reduce the size of the system by applying a singular value decomposition (SVD) on the idiosyncratic model part. Specifically, they rewrite the linearized system as

$$\begin{bmatrix} B_{ff}^{-1} A_{ff} & B_{ff}^{-1} A_{fX} \\ A_{Xf} & A_{XX} \end{bmatrix} \begin{bmatrix} f_t \\ X_t \end{bmatrix} = \begin{bmatrix} \tilde{A}_{ff} & \tilde{A}_{fX} \\ \tilde{A}_{Xf} & \tilde{A}_{XX} \end{bmatrix} \begin{bmatrix} f_t \\ X_t \end{bmatrix} = -\mathbb{E}_t \begin{bmatrix} f_{t+1} \\ B_{XX} X_{t+1} \end{bmatrix}. \tag{39}$$

Then, by applying a SVD on $\tilde{A}_{ff}$, i.e., $\tilde{A}_{ff} = U \Sigma V'$, and the Eckart-Young-Mirsky theorem, they obtain

$$\begin{bmatrix} V_1' U \Sigma_1 & V_1' \tilde{A}_{fX} \\ \tilde{A}_{Xf} V_1 & \tilde{A}_{XX} \end{bmatrix} \begin{bmatrix} Y_t \\ X_t \end{bmatrix} \approx -\mathbb{E}_t \begin{bmatrix} Y_{t+1} \\ B_{XX} X_{t+1} \end{bmatrix}, \tag{40}$$

where V_1 refers to the rows in V that correspond to the largest singular values and $Y_t = V' f_t$. Since $\tilde{A}_{ff}$ is independent of the estimated parameters, the SVD needs to be performed only infrequently. With this second-stage model reduction, the size of the model decreases drastically once again and the QZ-decomposition needed to solve the system can take place within a relatively short amount of time, which makes the estimation feasible.

3 Online estimation of HANK models

This section describes a Monte Carlo approach that makes possible what we call "online" estimation of HANK models. By online estimation, we mean estimation that can be conducted without starting from scratch as the dataset changes because, say, a new quarter of data becomes available. If estimating a model from scratch is nowadays a relatively trivial computational task for (linear) medium-scale DSGE models of the size of Smets and Wouters (2007)'s, it becomes much more time consuming and computer intensive when the size of the state space becomes very large, as is the case for HANK models.

Online estimation can be useful to central bank researchers who would like to use HANK models for forecasting. It can also be useful for academics who intend to run pseudo out-of-sample forecasting comparisons to assess the forecasting ability of HANK models, as we do in this paper, since these comparisons involve re-estimating the model(s) for each vintage of data.[8] Finally, online estimation can also be used to quickly re-estimate a model after small changes, such as modifications of the prior or any other relatively minor (or perhaps even major) alterations to the model.

The first part of this section describes the estimation problem and why its current handling by popular DSGE estimation packages such as `Dynare` may not be ideal for HANK models. The following subsection provides an intuitive description of an alternative estimation method—Sequential Monte Carlo (henceforth, SMC)—and explains why this approach is suitable for online estimation. While this section borrows much of the material from Cai et al. (2021), it strives to be accessible to an audience with little or no prior knowledge of Monte Carlo methods.[9]

[8] Edge and Gürkaynak (2010) and Del Negro and Schorfheide (2013) are examples of forecasting comparisons using medium-scale DSGEs.

[9] A terrific introduction to such methods is provided in textbooks such as Gelman et al. (1995), Geweke (2005), and Herbst and Schorfheide (2015), with the latter focusing specifically on DSGE model estimation. We refer the reader to these textbooks for a more formal treatment of the ideas described below.

3.1 Bayesian estimation of HANK models using state-space methods

The solution of the log-linearized version of the model described in section 2 produces the following transition equation:

$$s_t = T(\theta)s_{t-1} + R(\theta)\varepsilon_t,\ t = 1, .., T, \tag{41}$$

where s_t is the vector of states, θ is the parameter vector, and the shocks ε_t are independently and identically distributed according to $\varepsilon_t \sim N(0, Q(\theta))$. The measurement equation

$$y_t = Z(\theta)s_t + D(\theta) + u_t,\ t = 1, .., T \tag{42}$$

connects the latent states s_t to the vector of observables y_t, where the measurement error shocks are independently and identically distributed according to $u_t \sim N(0, H(\theta))$. The likelihood of this linear, Gaussian state-space model $p(y_{1:T}|\theta)$ can be readily computed via the Kalman filter, where we use the notation $y_{1:T}$ to denote the sequence of observations $\{y_1, ..., y_T\}$. Using Bayes' law, the posterior distribution of the parameters $p(\theta|y_{1:T})$ is obtained from

$$p(\theta|y_{1:T}) = \frac{p(y_{1:T}|\theta)p(\theta)}{\int p(y_{1:T}|\theta)p(\theta)d\theta}, \tag{43}$$

where $p(\theta)$ represents our prior for the parameters (Del Negro and Schorfheide, 2009, discusses the choice of priors for DSGE models and Müller, 2012, provides an easy way to assess their influence on the results). The discussion so far applies to any log-linearized DSGE model and closely follows An and Schorfheide (2007), Del Negro and Schorfheide (2010), and Herbst and Schorfheide (2015). The peculiarity of HANK models is that the state-space vector s_t is extremely large, making the Kalman filter and hence the evaluation of the likelihood $p(y_{1:T}|\theta)$ very costly.[10]

Since the posterior $p(\theta|y_{1:T})$ does not follow any known distribution, we need Monte Carlo methods in order to obtain draws from it and describe the results of our inference on θ (that is, tabulate the posterior mean, the 90 percent posterior coverage intervals, *et cetera*). The most standard Monte Carlo algorithm used for this purpose when estimating DSGE models and used in `Dynare`, is the *Random-Walk Metropolis Hastings* (RWMH). This algorithm is a so-called Markov Chain algorithm in that it produces a chain of draws from the posterior distribution $\{\theta^{(1)}, \ldots, \theta^{(j)}, \ldots, \theta^{(J)}\}$. Loosely speaking, the algorithm works as follows: in order to obtain the draw $\theta^{(j)}$, one takes the previous draw $\theta^{(j-1)}$, adds some randomness to generate a proposal θ^*, and then either accepts (that is, set $\theta^{(j)} = \theta^*$) or rejects (that is, set $\theta^{(j)} = \theta^{(j-1)}$) this proposal according to a formula that guarantees convergence of the chain to the desired ergodic distribution, that is,

[10] Herbst (2015) shows how the "Chandrasekhar Recursions" formulas can substantially reduce the computational burden of evaluating the likelihood. One issue with these formulas is that they are far less generous than standard formulas in terms of accommodating missing data, which is why we do not use them in this paper.

$p(\theta|y_{1:T})$ (again, see An and Schorfheide, 2007; Del Negro and Schorfheide, 2010; or Herbst and Schorfheide, 2015).

This is the algorithm used by almost all papers doing Bayesian estimation of DSGE models, including BBL, and for medium-sized models this algorithm has been shown to work reasonably well. It has a few downsides, however: 1) it is well known that RWMH may get stuck and fail to explore the entirety of the parameter space, especially in the presence of multi-modality (see, for instance, Herbst and Schorfheide, 2014); 2) it cannot be parallelized, since it is a Markov chain; and 3) one has to start from scratch for any new estimation, even if the changes in the estimation settings are relatively minor so that one would not expect a major change in the posterior distribution (e.g., adding one more quarter of data). These issues are particularly serious for HANK models because their posterior distribution is harder to evaluate. For instance, one approach to dealing with the first problem amounts to running very long chains, which increases the chances of visiting the entirety of the parameter space. Of course this approach is less appealing when computing $p(\theta^{(j)}|y_{1:T})$ is very costly. Similarly, the fact that the algorithm cannot be parallelized limits the extent to which one can take advantage of computer power to speed up the algorithm. While recent developments in Monte Carlo methods, such as Hamiltonian Monte Carlo (e.g., Duane et al., 1987; Neal et al., 2011; Stan Development Team, 2015, henceforth, HMC), have made Markov Chain methods more efficient and to some extent amenable to parallelization, the third problem—the fact that one has to start each estimation from scratch—makes SMC methods appealing. We describe these methods in the next section.

3.2 The Sequential Monte Carlo algorithm

To appreciate how and why Sequential Monte Carlo works, it may be useful to take a brief detour into the early history of Monte Carlo methods and discuss an approach called Importance Sampling (see Hammersley and Morton, 1954 for an early example and the textbooks mentioned in footnote 9 for a more modern treatment). Let's say we do not know how to draw from the posterior $p(\theta|y_{1:T})$, but we can draw very efficiently from a *proposal* distribution $q(\theta)$. For example, $q(\theta)$ could be a Gaussian with mean $\hat{\theta} = \text{argmax}_\theta \; p(\theta|y_{1:T})$, the peak of the posterior, and with variance proportional to the negative of the inverse of the numerical second derivative of the posterior evaluated at $\hat{\theta}$. Then we can obtain $\{\theta^{(1)}, \ldots, \theta^{(j)}, \ldots, \theta^{(J)}\}$ independent draws from $q(\theta)$ and assign to each of these draws a weight $W^{(j)} = w^{(j)}/(\frac{1}{J}\sum_{j=1}^{J} w^{(j)})$, where

$$w^{(j)} = p(y_{1:T}|\theta^{(j)})p(\theta^{(j)})/q(\theta^{(j)}) \propto p(\theta^{(j)}|y_{1:T}|)/q(\theta^{(j)}).$$

In other words, the idea behind Importance Sampling is to draw from $q(\theta)$ and then do a change of measure from $q(\cdot)$ to the so-called *target* distribution (the actual posterior) by reweighing these draws. Note that the denominator in (43) is irrelevant in the computation of $w^{(j)}$ since it does not depend on θ, and that the $W^{(j)}$ are in any case renormalized to sum up to J (the choice of J as the normalization constant, as opposed to the more conventional 1, is driven by numerical reasons). Given the *swarm of particles* $\{\theta^{(j)}, W^{(j)}\}_{j=1}^{J}$ produced by this approach, we can then approximate any object of interest $h(\theta)$ using the Monte Carlo average

$$\bar{h}_J = \frac{1}{J}\sum_{j=1}^{J} W_n^{(j)} h(\theta^{(j)}),$$

where, for instance, $h(\theta) = \theta$ if we want to compute the mean.

This may sound like a very reasonable approach except that the accuracy of this approximation does not just depend on J, which can be easily increased, but also on the effective particle sample size

$$\widehat{ESS} = J \Big/ \left(\frac{1}{J}\sum_{j=1}^{J} (W^{(j)})^2\right).$$

In other words, if $q(\theta)$ is a good proposal (in the example above, if the posterior is approximately Gaussian), then for J reasonably large Importance Sampling delivers a good approximation of the object of interest: all the weights $W^{(j)}$ will be similar in magnitude and the effective sample size $\widehat{ESS}$ will not be much lower than J. If it is not a good approximation, then most of the weights will be close to zero, and $\widehat{ESS}$ will be much lower than J. In this situation, Importance Sampling fails. When the posterior is irregular, as is the case for many DSGEs, coming up with a good (global) approximation is nearly impossible, and this may partly explain why in DSGE estimation these methods have been abandoned in favor of Markov Chain approaches such as RWMH.[11]

SMC brings Importance Sampling and the use of swarms of particles back into play for DSGE estimation thanks to two ideas. The first idea is that if we can pick the posterior we want to approximate, then the problem of choosing a suitable proposal becomes much easier. For instance, if the posterior is

$$p_n(\theta|y_{1:T}) = \frac{p(y_{1:T}|\theta)^{\phi_n} p(\theta)}{\int p(y_{1:T}|\theta)^{\phi_n} p(\theta) d\theta}, \tag{44}$$

with ϕ_n being a very small number, then the prior $p(\theta)$ is likely to work pretty well as a proposal: by construction, the target is almost the same as the proposal. Of course, $p_n(\theta|y_{1:T})$ constructed with ϕ_n close to zero is not what we want to approximate in the end. So we can increase ϕ_n progressively toward 1, and use the $n-1$ swarm as a proposal for the stage n target, making sure that at each stage n the target and the proposal remain reasonably close.

If the swarm of particles is still that generated by the prior, all this slicing into intermediate steps would amount to nothing: the prior is a poor proposal for the eventual posterior, and the effective sample size

[11] Importance Sampling-inspired approaches have remained very popular for filtering problems, such as the particle filter. See Fernández-Villaverde and Rubio-Ramírez, 2007.

will likely still be very low. But the second idea, which borrows from Markov Chain methods, comes to the rescue: from one stage to the other, particles can travel. Just like a single particle in RWMH travels around the posterior, and naturally tends to visit regions of the parameter space where the posterior places non-negligible mass, so can each of the particles in the swarm $\{\theta_n^{(j)}, W_n^{(j)}\}_{j=1}^J$. In other words, the particles can adapt as ϕ_n increases toward 1, so that in the end we have a good approximation of the posterior distribution.[12]

Formally, the SMC algorithm goes as follows:

Algorithm 1 (SMC Algorithm)**.**

1. **Initialization.** ($\phi_0 = 0$). Draw the initial particles from the prior: $\theta_1^{(j)} \overset{iid}{\sim} p(\theta)$ and $W_1^{(j)} = 1$, $j = 1, \ldots, J$.

2. **Recursion.** For $n = 1, \ldots, N_\phi$,

   (a) **Correction.** Re-weight the particles from stage $n-1$ by defining the incremental weights

   $$\tilde{w}_n^{(j)} = p(y_{1:T}|\theta_{n-1}^{(j)})^{\phi_n - \phi_{n-1}} \tag{45}$$

   and the normalized weights

   $$\tilde{W}_n^{(j)} = \frac{\tilde{w}_n^{(j)} W_{n-1}^{(j)}}{\frac{1}{J}\sum_{j=1}^J \tilde{w}_n^{(j)} W_{n-1}^{(j)}}, \quad j = 1, \ldots, J. \tag{46}$$

   (b) **Selection (Optional).** Resample the swarm of particles $\{\theta_{n-1}^{(j)}, \tilde{W}_n^{(j)}\}_{j=1}^J$ and denote resampled particles by $\{\hat{\theta}^{(j)}, W_n^{(j)}\}_{j=1}^J$, where $W_n^{(j)} = 1$ for all j.

   (c) **Mutation.** Propagate the particles $\{\hat{\theta}_i, W_n^{(j)}\}$ via N_{MH} steps of an MH algorithm with transition density $\theta_n^{(j)} \sim K_n(\theta_n|\hat{\theta}_n^{(j)}; \zeta_n)$ and stationary distribution $p_n(\theta|y_{1:T})$.[13]

[12] A little bit of history: In the statistics literature, Chopin (2002) showed how to adapt particle-filtering techniques to conduct posterior inference for a static parameter vector. John Geweke played an important role popularizing these techniques in economics (e.g., Durham and Geweke, 2014), and Creal (2007) was the first paper that applied SMC techniques to posterior inference in a DSGE model. Herbst and Schorfheide (2014) was a quite impactful paper, as it showed that a properly tailored SMC algorithm delivers more reliable posterior inference for the Smets and Wouters (2007) DSGE model with loose priors and a multimodal posterior distribution than the widely used RWMH algorithm. They also provide some convergence results for an adaptive version of the algorithm building on Chopin (2004).

[13] The transition kernel $K_n(\theta_n|\hat{\theta}_n; \zeta_n)$ needs to have the following invariance property:

$$p_n(\theta_n|y_{1:T}) = \int K_n(\theta_n|\hat{\theta}_n; \zeta_n) p_n(\hat{\theta}_n|y_{1:T}) d\hat{\theta}_n.$$

Thus, if $\hat{\theta}_n^{(j)}$ is a draw from the stage n posterior $p_n(\theta_n|y_{1:T})$, then so is $\theta_n^{(j)}$. The MH accept-reject probability ensures that this property is satisfied. In our application we follow Herbst and Schorfheide (2014) and Cai et al. (2021) in our choice of $K_n(\theta_n|\hat{\theta}_n; \zeta_n)$, but developments in MH algorithms, such as HMC, can be used to make this step, and hence the whole SMC algorithm, more efficient. Farkas and Tatar (2020) is an example of a paper that combines SMC with HMC.

3. For $n = N_\phi$ ($\phi_{N_\phi} = 1$) the final Importance Sampling approximation of $\mathbb{E}_\pi[h(\theta)]$ is given by

$$\bar{h}_{N_\phi,N} = \sum_{j=1}^{J} h(\theta_{N_\phi}^{(j)}) W_{N_\phi}^{(j)}. \tag{47}$$

Step 2a is the same as in Importance Sampling, where the proposal is the previous stage's posterior $p_{n-1}(\theta|y_{1:T})$ and the target is $p_n(\theta|y_{1:T})$. Step 2c is the Metropolis Hastings step, where each particle is given a chance to adapt to the new posterior. Step 2b needs some discussion. Its purpose is to make sure that if the weights of the particles in the swarm become very uneven, and effective particle sample size

$$\widehat{ESS}_n = J \Big/ \left(\frac{1}{J} \sum_{j=1}^{J} (\tilde{W}_n^{(j)})^2 \right)$$

falls below a threshold $\underline{J}$, a new swarm of particles is generated from the old swarm so that all the particles have the same weight.[14]

One aspect of the algorithm we have not yet discussed is the number of stages N_ϕ as well as the schedule $\{\phi_1, \ldots, \phi_{N_\phi}\}$. In estimating the Smets and Wouters (2007) model, Herbst and Schorfheide (2014) find that $N_\phi = 500$ and a schedule given by the function $\phi_n = (n/N_\phi)^{2.1}$ works well. The convexity of the schedule implies that ϕ_n increases very slowly at the beginning and faster at the end. Of course, it is far from obvious that whatever schedule works well for the Smets and Wouters (2007) model also works well for a HANK or any other DSGE model. In this respect, Cai et al. (2021) improve upon Herbst and Schorfheide (2014) by making the schedule $\{\phi_1, \ldots, \phi_{N_\phi}\}$ adaptive—that is, endogenous to the difficulty of the problem. Recall that the ESS measures, loosely speaking, the deterioration of the quality of the swarm $\{\theta_n^{(j)}, W_n^{(j)}\}_{j=1}^{J}$: if ESS is low, the swarm has essentially "lost" most of its particles as the weights have become very uneven. Adaptation is then achieved by setting at each stage $\phi_n = \phi$, where ϕ solves

$$\widehat{ESS}(\phi) - (1 - \alpha)\widehat{ESS}_{n-1} = 0,$$

where

$$\tilde{w}^{(j)}(\phi) = [p(y_{1:T}|\theta_{n-1}^{(j)})]^{\phi - \phi_{n-1}}, \quad \tilde{W}^{(j)}(\phi) = \frac{\tilde{w}^{(j)}(\phi) W_{n-1}^{(j)}}{\frac{1}{J}\sum_{j=1}^{J} \tilde{w}^{(j)}(\phi) W_{n-1}^{(j)}}, \quad \widehat{ESS}(\phi) = N \Big/ \left(\frac{1}{J} \sum_{j=1}^{J} (\tilde{W}_n^{(j)}(\phi))^2 \right).$$

The above formulas can be understood as follows. Pick a desired deterioration α of the effective sample size between stages $n - 1$ and n, and set ϕ_n so as to achieve exactly that deterioration (see Cai et al., 2021,

[14]Loosely speaking, particles with relatively large weight $\tilde{W}_n^{(j)}$—that is, that are in high posterior regions of the parameter space—are given an opportunity ro "procreate" (generate a number of children that is in expected values proportional $\tilde{W}_n^{(j)}$), while particles with relatively small weight ($\tilde{W}_n^{(j)} < 1$)—that is, that are in regions of the parameter space with very little mass—are "killed" with probability $1 - \tilde{W}_n^{(j)}$. There are many resampling schemes (see Liu, 2001, or Cappé et al., 2005). We use systematic resampling in the applications below.

for a more detailed description). The parameter α expresses the degree of "carefulness" of the researchers, bearing in mind that lower α's imply a longer estimation time (in light of the results in Cai et al., 2021, we choose $\alpha = 5$ percent in this application). Once α is chosen, the schedule becomes endogenous to the difficulty of the problem as measured by the deterioration of the ESS.

This section concludes with a description of some virtues of SMC. First, for a suitably large choice of the size of the swarm J, it is robust to irregular shapes of the posterior such as multi-modality, as shown in Herbst and Schorfheide (2014) and Cai et al. (2021), among others. This is because the initial swarm $\{\theta_0^{(j)}, W_0^{(j)}\}_{j=1}^J$ is drawn from the prior, and hence covers for large enough J any region of the parameter space where the prior places non-negligible mass. Hence, if the posterior has many modes, there ought to be some initial particles in the neighborhood of such modes. Second, most of the SMC steps, such as the computation of the incremental weights in Step 2a, and most important, the mutation step in Step 2c, can be parallelized. Third, the algorithm produces an approximation of the marginal likelihood as a by-product. In fact, using the definitions of $\tilde{w}_n^{(j)}$ and $\tilde{W}_{n-1}^{(j)}$ one can see that:

$$\frac{1}{J}\sum_{i=1}^{N} \tilde{w}_n^{(j)} \tilde{W}_{n-1}^{(j)} \approx \int \frac{p(y_{1:T}|\theta)^{\phi_n}}{p(y_{1:T}|\theta)^{\phi_{n-1}}} \left[\frac{p(y_{1:T}|\theta)^{\phi_{n-1}}}{\int p(y_{1:T}|\theta)^{\phi_{n-1}} p(\theta) d\theta} \right] d\theta = \frac{\int p(y_{1:T}|\theta)^{\phi_n} p(\theta) d\theta}{\int p(y_{1:T}|\theta)^{\phi_{n-1}} p(\theta) d\theta}. \tag{48}$$

This implies that the product $\prod_{n=1}^{N_\phi} \left(\frac{1}{J} \sum_{j=1}^{J} \tilde{w}_n^{(j)} W_{n-1}^{(j)} \right)$ approximates the marginal likelihood as long as the prior is proper ($\int p(\theta) d\theta = 1$), since all the terms cancel out except for $\int p(y_{1:T}|\theta) p(\theta) d\theta / \int p(\theta) d\theta$. Fourth, and perhaps most important for this application, the final swarm of particles $\{\theta_{N_\phi}^{(j)}, W_{N_\phi}^{(j)}\}_{j=1}^J$ can be reused, making recursive estimation of the model very convenient. We turn to this feature next.

3.3 Online estimation

Imagine running the SMC algorithm 1 and having a swarm of particles $\{\theta^{(j)}, W^{(j)}\}_{j=1}^J$ that approximates well the posterior $p(\theta|y_{1:T})$. Expression (45) in Step 2a of the algorithm can be generalized as

$$\tilde{w}_n^{(j)} = \frac{p_n(Y|\theta_{n-1}^{(j)})}{p_{n-1}(Y|\theta_{n-1}^{(j)})}, \tag{49}$$

where we now use the more generic notation Y for $y_{1:T}$ for reasons that will soon become apparent. Note that in (45) we considered the special case where the stage-n likelihood $p_n(Y|\theta) = p(Y|\theta)^{\phi_n}$.

Now imagine wanting to obtain the posterior for a different model $\tilde{p}(\cdot|\theta)$ (but with the same parameter vector θ) estimated on a different dataset $\tilde{Y}$:

$$\tilde{p}(\theta|\tilde{Y}) = \frac{\tilde{p}(\tilde{Y}|\theta) p(\theta)}{\int \tilde{p}(\tilde{Y}|\theta) p(\theta) d\theta}. \tag{50}$$

The simplest possible case is the one in which the model is the same ($\tilde{p}(\cdot|\theta) = p(\cdot|\theta)$), and the dataset has one more time-series observation ($\tilde{Y} = y_{1:T+1}$), but the algorithm can accommodate situations where the data has been revised or the model changed. Draws for the posterior $\tilde{p}(\theta|\tilde{Y})$ can be readily obtained from algorithm 1 after replacing expression (45) with expression (49) and using the stage-n likelihood function (see Cai et al., 2021):

$$\tilde{p}_n(\tilde{Y}|\theta) = \tilde{p}(\tilde{Y}|\theta)^{\phi_n} p(Y|\theta)^{1-\phi_n}. \tag{51}$$

In other words, we use the posterior distribution $p(\theta|Y)$ as a "bridge" to obtain the new posterior $\tilde{p}(\theta|\tilde{Y})$, as opposed to starting from the prior distribution. To the extent that the differences between $\tilde{p}(\tilde{Y}|\theta)$ and $p(Y|\theta)$ are not large, the swarm from $p(\theta|Y)$ should offer a fairly good starting point for the SMC algorithm.[15] This is the approach we use to estimate the BBL HANK model recursively. In particular, we start from the end-of-sample estimation $p(\theta|y_{1:T})$ and then proceed backward using formula (51) with $\tilde{Y} = y_{1:T-\tau}$ and $Y = y_{1:T-\tau+1}$, for $\tau = 1, ..., \bar{\tau}$. We should stress that doing the online recursive estimation backward or forward—that is, starting from $p(\theta|y_{1:T-\bar{\tau}})$ and using this as a bridge to obtain $p(\theta|y_{1:T-\bar{\tau}+1})$, and so on—should make no difference, as both procedures recover $p(\theta|y_{1:T-\tau})$.[16]

We conclude this section by highlighting some of the potentials of this approach besides the online estimation of HANK models. Mlikota and Schorfheide (2022) introduce the notion of "model tempering." If a model is very costly to estimate from scratch, one can save a lot of time by first estimating a coarser version of the model that is much cheaper to estimate (e.g., the linearized version of a nonlinear model) and then use that as a bridge to estimate the full model. Mlikota and Schorfheide (2022) use this approach to estimate a nonlinear model.

4 Results

This section presents the forecasting results and is divided into three parts. The first part describes the setup of the exercise, including the data. The second part discusses the estimates of the parameters, focusing on the differences between the original BBL results and those obtained using the SMC algorithm. The last part covers the forecasting horse race between BBL and SW.

[15] The initialization step in algorithm 1 needs to be modified so that the swarm $\{\theta^{(j)}, W^{(j)}\}_{j=1}^J$, possibly after a selection step 2b so that all the $W^{(j)}$'s equal 1, replaces the swarm drawn from the prior.

[16] In particular there is no sense in which the backward procedure introduces any hindsight bias: note that by the time that ϕ_n in (51) reaches 1, the posterior draws no longer condition on $Y = y_{1:T-\tau+1}$.

4.1 Setup

For our exercise we use the dataset made available by BBL online at https://github.com/BASEforHANK/HANK_BusinessCycleAndInequality as of June 2022, when we began this project. This dataset comprises the seven variables used by SW in the estimation of their model: the growth rates of per capita real output, consumption, investment, and wages, the logarithm of hours worked per capita, GDP deflator inflation, and the federal funds rate (during the zero-lower-bound period the authors use the shadow rate measure created by Wu and Xia, 2016). In their database, these variables are available at the quarterly frequency from 1954Q3 to 2019Q4. In addition, BBL estimate their model adding four variables that reflect various aspects of inequality and are not used in standard representative agent DSGE estimation. These are the wealth and income shares of the top 10 percent, estimates of tax progressivity constructed following Ferriere and Navarro (2018), and estimates of idiosyncratic income risk from Bayer et al. (2019). The top 10 percent shares are available annually from 1954 to 2019, the tax progressivity measure is available annually from 1954 to 2017, and the idiosyncratic income risk measure is available from 1983Q1 to 2013Q1.[17] The likelihood computation of the state-space model easily accommodates missing data.

BBL demean all the time series prior to estimation. While this is not standard practice in the DSGE estimation and forecasting literature and in central banks' practice (e.g., see Del Negro and Schorfheide, 2013; Cai et al., 2019), we follow BBL because adding a constant would imply introducing steady-state growth and inflation, therefore altering their model. We chose not to do this in order to remain as close as possible to BBL's specification. This choice has two implications. First, we have to use their dataset also for the forecasting exercise—that is, the demeaned data is what the model's forecasts are evaluated upon. Second, we estimate the competitor in the horse race—the SW model—also on demeaned data, which implies that we drop the constant from SW's measurement equations.[18]

The out-of-sample forecasting exercise begins in the first quarter of 2000 (in the notation of section 3, $T - \bar{\tau} = 2000Q1$) and ends in the last quarter of 2019 ($T = 2019Q4$), for a total of 80 periods. In order to avoid hindsight bias, for each period we re-estimate the model using only data available up to that period.[19] For each model $\mathcal{M}_m$ under consideration (BBL, SW), we then generate horizon-h mean forecasts

[17] See Bayer et al. (2022) for a more detailed description of the dataset.

[18] We should note that BBL have a representative agent version of their HANK model, which they show has a worse fit for the seven macro variables than the heterogeneous agents version in terms of marginal likelihood. Given that the actual SW model is available, we use it in the horse race as the alternative to the BBL HANK model. From the perspective of a central bank choosing whether to use a representative agent or a HANK model for predictions, arguably the choice is between SW and BBL.

[19] In the forecasting literature it is customary to perform so-called pseudo real time forecasting, where the data vintage available at time $T - \tau$ is used for estimation, as opposed to the revised data (here, the T vintage). The demeaning of the data, and the fact that there are no vintages for the inequality series, makes this pseudo real time exercise not possible.

$\mathbb{E}[y_{T-\tau+h}|y_{1:T-\tau}, \mathcal{M}_m]$ for the variables of interest using the state-space model consisting of equations (41) and (42), then compare these forecasts with actual outcomes $y_{T-\tau+h}$.[20] As discussed above in section 3, we estimate the model in period $T-\tau$ using the posterior distribution for $T-\tau+1$ as a bridge in the SMC algorithm. For the sake of robustness we start this process from two different posterior distributions $p(\theta|y_{1:T})$. One distribution consists of the draws made available by BBL on `GitHub` (which is the reason we do the online estimation backward, since these draws are only available for $T = 2019Q4$), and the other is based on an SMC estimation starting from the prior. For all SMC estimations we use a swarm of $J = 10k$ particles (we obtain nearly identical results when using $1k$ particles, at least in terms of RMSEs). The next section discusses these posterior estimates in some detail.

4.2 Estimation

This section presents the results from our estimation. Specifically, we discuss the prior distribution, which is the same as BBL's, and posteriors from the full sample SMC estimation using the eleven variables described in the previous section. Also, we show the posteriors from SMC estimation when using only seven aggregate variables and excluding data on inequality, tax progressivity estimates, and income risk estimates. For comparison, we present BBL's estimation results obtained from BBL's `GitHub` page as of June 2022.[21] As mentioned above, in order to make the estimation feasible, BBL calibrate, as opposed to estimate, several of the model's parameters. Table 1 shows the values of these calibrated parameters.

Priors

For parameters related to monetary policy, BBL impose normal distribution with a mean of 1.7 and standard deviation of 0.3 for θ_π, while imposing normal distribution with a mean of 0.13 and standard deviation of 0.05 for θ_Y. For the interest rate smoothing parameter ρ_R, they assume a beta distribution with parameters (0.5, 0.2).

Regarding fiscal policy, the debt-feedback parameter γ_B in the bond issuance rule is assumed to follow a gamma distribution with a mean of 0.10 and standard deviation of 0.08, which implies that the prior for the autocorrelation of government debt is 0.9. For the responsiveness of government debt to inflation and output growth, γ_π and γ_Y, they impose standard normal distributions. Similarly, they assume beta distributions with a mean of 0.5 and standard deviation of 0.2 for the autoregressive parameters in the tax rules, ρ_P, and ρ_τ. The feedback parameters for average tax rates, γ_Y^τ, and γ_B^τ, are assumed to follow standard normal distributions.

[20] In order to compute the expectation $\mathbb{E}[y_{T-\tau+h}|y_{1:T-\tau}, \mathcal{M}_m]$ using (41) and (42), only the filtered states $s_{T-\tau|T-\tau} = \mathbb{E}[s_{T-\tau}|y_{1:T-\tau}, \mathcal{M}_m]$ are needed, which are obtained from the Kalman filter.

[21] As mentioned, BBL made a few changes to their model and calibrated parameters since June 2022. Hence, these MH estimates do not replicate the results presented in the most recent version of their paper.

Table 1: Calibration

Par.	Value	Description
Households: Income process		
ρ_h	0.980	Persistence of labor productivity
σ_h	0.120	Std. dev. of labor productivity
ι	0.063	Trans. prod. from entrepreneurs to workers
ζ	1/3750	Trans. prod. from workers to entrepreneurs
Households: Financial frictions		
λ	0.095	Portfolio adj. prob.
$\bar{R}$	0.017	Borrowing premium
Households: Preferences		
β	0.984	Discount factor
ξ	4.000	Relative risk aversion
γ	2.000	Inverse of Frisch elasticity
Firms		
α	0.682	Share of labor
δ_0	0.022	Depreciation rate
$\bar{\eta}$	11.000	Elasticity of substitution
$\bar{\zeta}$	11.000	Elasticity of substitution
Government		
$\bar{\tau}^L$	0.175	Tax rate level
$\bar{\tau}^P$	0.12	Tax progressivity
$\bar{R}^b$	1	Gross nominal rate
$\bar{\pi}$	1	Steady-state inflation rate

For the structural shocks, BBL assume beta distributions with a mean of 0.5 and a standard deviation of 0.2 for the autocorrelation parameters and inverse-gamma distributions with a mean of 0.001 and a standard deviation of 0.02 for standard deviations of shocks. Finally, for idiosyncratic income risks, BBL impose a beta distribution with a mean of 0.7 and a standard deviation of 0.2 for autocorrelation parameters and a gamma distribution with a mean of 0.65 and a standard deviation of 0.03.

Regarding variable capital utilization, BBL assume a gamma distribution with a mean of 5.0 and standard deviation of 2.0 for $\delta_s = \delta_2/\delta_1$. Similarly, they impose a gamma distribution with a mean of 4.0 and standard deviation of 2.0 for ϕ, the parameter that governs investment adjustment costs. For the slopes of price and wage Phillips curves, κ_Y and κ_w, they adopt gamma priors with a mean of 0.10 and standard deviation of 0.03. The prior mean for these parameters implies that the average duration of price and wage is four quarters.

Posteriors

The posterior distributions using the full sample (T=2019Q4) are displayed in Tables 2 and 3. Column 5 of each table shows BBL's original posteriors they obtained using the RWMH algorithm (referred to as the MH estimation hereafter). Columns 6 and 7 show the posterior distributions we obtained via the SMC approach using eleven and seven variables, respectively (referreed to as the 11 and 7 var SMC estimations hereafter). In the 7 var SMC estimation we follow BBL and shut down income risk and tax progressivity shocks since we do not use the related data in the estimation.

The posteriors from the 11 and 7 var SMC estimations exhibit only small differences, which is consistent with the findings of BBL. Adding data on inequality to the estimation does not significantly affect the results for the parameters that govern the aggregate dynamics of the model. The investment adjustment cost is estimated to be higher in the 7 var SMC estimation, but otherwise the posterior distributions are close to each other.

Posteriors from the MH and SMC estimations are also broadly similar for many parameters, but exhibit differences for some others, which we discuss in the remainder of this section. Starting with the parameters of the monetary policy rule, posteriors from the SMC estimations imply a slightly higher interest rate inertia and lower sensitivities of the interest rate with respect to the inflation rate and output growth relative to those from the MH estimation. The interest rate smoothing parameter is 0.82 and 0.84 at the mean in the SMC estimations, while the posterior mean is 0.79 in the MH estimation. The Taylor rule coefficient on inflation is relatively low in the SMC estimations, with the 10 to 90 percentile range being from 1.53 to 1.80 in the 11 var estimation and 1.36 to 1.77 in the 7 var estimation. In contrast, the corresponding range is

Table 2: Prior and posterior distributions: policies and frictions

Par	Dist	Prior		Posterior		
		Mean	Std. Dev	BBL (MH)	BBL (SMC)	BBL (7 Var)
			Monetary policy			
ρ_R	Beta	0.50	0.20	0.785	0.818	0.841
				(0.754, 0.814)	(0.793, 0.842)	(0.819, 0.864)
σ_R	Inv-Gamma	0.10	2.00	0.243	0.206	0.210
				(0.224, 0.269)	(0.191, 0.220)	(0.195, 0.226)
θ_π	Normal	1.70	0.30	2.237	1.670	1.570
				(2.044, 2.424)	(1.532, 1.804)	(1.357, 1.773)
θ_Y	Normal	0.13	0.05	0.287	0.212	0.187
				(0.223, 0.361)	(0.158, 0.261)	(0.117, 0.258)
			Fiscal policy: deficit			
ρ_D	Beta	0.50	0.20	0.965	0.790	0.775
				(0.950, 0.980)	(0.760, 0.816)	(0.738, 0.811)
σ_D	Inv-Gamma	0.10	2.00	0.310	0.632	0.994
				(0.277, 0.342)	(0.548, 0.715)	(0.842, 1.159)
γ_B	Gamma	0.10	0.08	0.031	0.039	0.025
				(0.008, 0.047)	(0.021, 0.056)	(0.005, 0.043)
γ_π	Normal	0.00	1.00	-1.601	-2.739	-3.969
				(-1.778, -1.452)	(-3.091, -2.379)	(-4.529, -3.397)
γ_Y	Normal	0.00	1.00	-0.350	-0.736	-1.222
				(-0.418, -0.309)	(-0.826, -0.631)	(-1.405, -1.029)
			Fiscal policy: taxes			
ρ_τ	Beta	0.50	0.20	0.653	0.809	0.702
				(0.440, 0.961)	(0.731, 0.885)	(0.508, 0.886)
γ_B^τ	Normal	0.00	1.00	0.166	-1.765	-1.222
				(0.110, 0.217)	(-2.079, -1.387)	(-1.483, -0.937)
γ_Y^τ	Normal	0.00	1.00	-0.148	-0.329	1.152
				(-0.410, 0.038)	(-1.736, 1.228)	(-0.203, 2.507)
			Income risk			
ρ_s	Beta	0.50	0.20	0.663	0.917	-
				(0.606, 0.727)	(0.627, 0.995)	-
σ_s	Gamma	65.00	30.00	64.08	57.67	-
				(55.91, 71.06)	(51.87, 64.59)	-
			Frictions			
δ_s	Gamma	5.00	2.00	0.456	1.800	2.345
				(0.278, 0.631)	(1.419, 2.227)	(1.885, 2.796)
ϕ	Gamma	4.00	2.00	0.787	3.532	6.928
				(0.373, 1.244)	(2.899, 4.164)	(5.799, 8.066)
κ_Y	Gamma	0.10	0.03	0.111	0.116	0.114
				(0.094, 0.125)	(0.103, 0.128)	(0.099, 0.128)
κ_w	Gamma	0.10	0.03	0.112	0.126	0.113
				(0.095, 0.128)	(0.111, 0.142)	(0.096, 0.128)

Notes: The table shows prior and posterior distributions of the estimated parameters (The 10 and 90 percentiles of the distributions are in parentheses).

Table 3: Prior and posterior distributions: structural shocks

		Prior		Posterior		
Par	Dist	Mean	Std. Dev	BBL (MH)	BBL (SMC)	BBL (7 Var)
			Structural shocks			
ρ_A	Beta	0.50	0.20	0.954	0.976	0.984
				(0.925, 0.976)	(0.970, 0.982)	(0.976, 0.993)
σ_A	Inv-Gamma	0.10	2.00	0.162	0.078	0.062
				(0.133, 0.194)	(0.059, 0.094)	(0.041, 0.081)
ρ_Z	Beta	0.50	0.20	0.998	0.967	0.973
				(0.996, 0.999)	(0.958, 0.978)	(0.963, 0.983)
σ_Z	Inv-Gamma	0.10	2.00	0.569	0.616	0.612
				(0.526, 0.624)	(0.573, 0.662)	(0.569, 0.659)
ρ_Ψ	Beta	0.50	0.20	0.848	0.486	0.416
				(0.790, 0.904)	(0.422, 0.568)	(0.338, 0.495)
σ_Ψ	Inv-Gamma	0.10	2.00	3.814	12.25	24.29
				(2.820, 4.982)	(9.905, 14.77)	(20.58, 28.13)
ρ_{μ_p}	Beta	0.50	0.20	0.862	0.968	0.967
				(0.824, 0.907)	(0.955, 0.983)	(0.953, 0.980)
σ_{μ_p}	Inv-Gamma	0.10	2.00	1.563	1.410	1.452
				(1.404, 1.714)	(1.296, 1.520)	(1.316, 1.597)
ρ_{μ_w}	Beta	0.50	0.20	0.862	0.898	0.878
				(0.826, 0.907)	(0.867, 0.931)	(0.845, 0.911)
σ_{μ_w}	Inv-Gamma	0.10	2.00	6.142	4.732	5.079
				(5.385, 6.916)	(4.043, 5.292)	(4.399, 5.753)
ρ_P	Beta	0.50	0.20	0.961	0.975	-
				(0.943, 0.981)	(0.958, 0.990)	-
σ_P	Inv-Gamma	0.10	2.00	3.534	3.647	-
				(2.938, 4.192)	(3.011, 4.202)	-

Notes: The table shows prior and posterior distributions of the estimated parameters (The 10 and 90 percentiles of the distributions are in parentheses). Standard deviations are multiplied by 100 for readability.

Table 4: Prior and posterior distributions: policies and frictions (2000Q1 estimation)

Par	Dist	Prior		Posterior		
		Mean	Std. Dev	BBL (MH)	Backward from MH	Backward from SMC
				Monetary policy		
ρ_R	Beta	0.50	0.20	0.785 (0.754, 0.814)	0.763 (0.729, 0.797)	0.796 (0.762, 0.831)
σ_R	Inv-Gamma	0.10	2.00	0.243 (0.224, 0.269)	0.272 (0.244, 0.302)	0.243 (0.220, 0.264)
θ_π	Normal	1.70	0.30	2.237 (2.044, 2.424)	1.941 (1.692, 2.169)	1.566 (1.365, 1.746)
θ_Y	Normal	0.13	0.05	0.287 (0.223, 0.361)	0.198 (0.129, 0.270)	0.178 (0.105, 0.245)
				Fiscal policy: deficit		
ρ_D	Beta	0.50	0.20	0.965 (0.950, 0.980)	0.954 (0.918, 0.992)	0.824 (0.783, 0.860)
σ_D	Inv-Gamma	0.10	2.00	0.310 (0.277, 0.342)	0.424 (0.332, 0.505)	0.681 (0.546, 0.822)
γ_B	Gamma	0.10	0.08	0.031 (0.008, 0.047)	0.035 (0.008, 0.061)	0.028 (0.006, 0.048)
γ_π	Normal	0.00	1.00	-1.601 (-1.778, -1.452)	-2.0330 (-2.358, -1.707)	-2.898 (-3.402, -2.417)
γ_Y	Normal	0.00	1.00	-0.350 (-0.418, -0.309)	-0.435 (-0.562, -0.297)	-0.756 (-0.891, -0.615)
				Fiscal policy: taxes		
ρ_τ	Beta	0.50	0.20	0.653 (0.440, 0.961)	0.544 (0.346, 0.743)	0.718 (0.584, 0.853)
γ_B^τ	Normal	0.00	1.00	0.166 (0.110, 0.217)	0.0774 (-0.083, 0.212)	-1.386 (-1.738, -1.018)
γ_Y^τ	Normal	0.00	1.00	-0.148 (-0.410, 0.038)	-2.170 (-3.555,-0.913)	-0.050 (-1.482,1.336)
				Income risk		
ρ_s	Beta	0.50	0.20	0.663 (0.606, 0.727)	0.633 (0.520, 0.746)	0.615 (0.498, 0.729)
σ_s	Gamma	65.00	30.00	64.08 (55.91, 71.06)	52.93 (43.64, 61.43)	49.69 (42.10, 56.87)
				Frictions		
δ_s	Gamma	5.00	2.00	0.456 (0.278, 0.631)	0.474 (0.268, 0.655)	1.916 (1.335, 2.474)
ϕ	Gamma	4.00	2.00	0.787 (0.373, 1.244)	2.554 (1.774, 3.316)	3.820 (2.955, 4.647)
κ_Y	Gamma	0.10	0.03	0.111 (0.094, 0.125)	0.121 (0.104, 0.138)	0.128 (0.110, 0.144)
κ_w	Gamma	0.10	0.03	0.112 (0.095, 0.128)	0.097 (0.083, 0.113)	0.097 (0.083, 0.113)

Notes: The table shows prior and posterior distributions of the estimated parameters (The 10 and 90 percentiles of the distributions are in parentheses).

Table 5: Prior and posterior distributions: structural shocks (2000Q1 estimation)

		Prior		Posterior		
Par	Dist	Mean	Std. Dev	BBL (MH)	Backward from MH	Backward from SMC
				Structural shocks		
ρ_A	Beta	0.50	0.20	0.954 (0.925, 0.976)	0.932 (0.886, 0.984)	0.976 (0.964, 0.987)
σ_A	Inv-Gamma	0.10	2.00	0.162 (0.133, 0.194)	0.215 (0.146, 0.279)	0.085 (0.056, 0.117)
ρ_Z	Beta	0.50	0.20	0.998 (0.996, 0.999)	0.993 (0.990, 0.996)	0.962 (0.947, 0.977)
σ_Z	Inv-Gamma	0.10	2.00	0.569 (0.526, 0.624)	0.652 (0.580, 0.715)	0.645 (0.589, 0.701)
ρ_Ψ	Beta	0.50	0.20	0.848 (0.790, 0.904)	0.631 (0.543, 0.734)	0.438 (0.348, 0.526)
σ_Ψ	Inv-Gamma	0.10	2.00	3.814 (2.820, 4.982)	9.771 (6.856, 12.49)	15.64 (12.08, 18.91)
ρ_{μ_p}	Beta	0.50	0.20	0.862 (0.824, 0.907)	0.862 (0.815, 0.913)	0.849 (0.773, 0.937)
σ_{μ_p}	Inv-Gamma	0.10	2.00	1.563 (1.404, 1.714)	1.301 (1.121, 1.480)	1.277 (1.077, 1.461)
ρ_{μ_w}	Beta	0.50	0.20	0.862 (0.826, 0.907)	0.895 (0.863, 0.930)	0.917 (0.886, 0.950)
σ_{μ_w}	Inv-Gamma	0.10	2.00	6.142 (5.385, 6.916)	4.404 (3.840, 5.033)	4.173 (3.628, 4.717)
ρ_P	Beta	0.50	0.20	0.961 (0.943, 0.981)	0.949 (0.928, 0.971)	0.961 (0.940, 0.982)
σ_P	Inv-Gamma	0.10	2.00	3.534 (2.938, 4.192)	3.835 (3.093, 4.545)	3.686 (3.015, 4.341)

Notes: The table shows prior and posterior distributions of the estimated parameters (The 10 and 90 percentiles of the distributions are in parentheses). Standard deviations are multiplied by 100 for readability.

from 2.04 to 2.42 in the MH estimation. The coefficient on output growth is around 0.2 at the mean in the SMC estimations, while a bit higher at 0.29 in the MH estimation.

In the case of the fiscal policy parameters, differences between MH and SMC estimation are more pronounced. Regarding the bond issuance rule, the SMC estimations imply much less persistence of the structural deficit with an autoregressive coefficient of around 0.81 at the posterior mean, while the mean is 0.97 in the MH estimation. Also, posteriors from the SMC estimations imply a much stronger countercyclical response of government debt to the inflation rate and output growth. The elasticities of the bond issuance with respect to inflation and output growth are -2.74 and -0.74 at the posterior mean in the 11 var SMC estimation and -3.97 and -1.22 in the 7 var SMC estimation, as opposed to -1.60 and -0.35 in the MH estimation. Posteriors for parameters governing tax rules show even larger differences. While posteriors from the SMC estimations imply countercyclical tax rate responses with respect to the growth rate of government debt, the posterior from the MH estimation implies pro-cyclical responses. Also, tax rates are estimated to be more persistent in the SMC estimations than in the MH estimation.

Among the parameters governing the model's frictions, the posterior distributions of variable capital depreciation and investment adjustment cost parameters show significant differences. In the MH estimation, the capital depreciation parameter is 0.46 at the posterior mean. In contrast, the means for the same parameter are 1.80 and 2.35 in the 11 and 7 var SMC estimations, respectively. The posterior mean for the capital adjustment cost parameter is 3.53 and 6.93 in the two SMC estimations and only 0.79 in the MH estimation.

The posterior distributions for the rest of the parameters, including the parameters describing income risk, the slope of the price and wage Phillips curves, and the structural shocks, are broadly similar. The autocorrelation of the MEI shock is the only exception. In the SMC estimations, MEI shocks are estimated to be not very persistent with an autocorrelation of around 0.5 at the posterior mean, while in the MH estimation the posterior mean for this parameter is 0.85.

The differences between the MH and the SMC estimation results obviously lead to the question of which method is the most accurate. This is a nontrivial question to address, since doing so would involve repeated (independent) estimations of the HANK model as done, for instance, in Cai et al. (2021). This is computationally very costly. We therefore sidestep this issue entirely and use both the SMC and MH estimations in our forecasting comparison exercise. By this we mean that we obtain two different sets of backward bridge estimations using the approach described in section 3.3—one starting from the SMC and one starting from the MH draws. It turns out that the accuracy of the BBL model estimated using SMC is better than that using the MH draws. While this evidence is no proof that the SMC estimation is more reliable, it seems to point in that direction.

Finally, in Tables 4 and 5, we present the posterior distributions from the estimations using the data up to 2000Q1, which we obtain using the approach described in section 3. Columns 6 and 7 show the posterior distributions from the backward estimation starting from the 2019Q4 MH and SMC draws, respectively. For comparison, we also show the posteriors from the original 2019Q4 BBL estimation in column 4. The 2000Q1 posterior is close to the MH BBL posterior when the original MH draws were used as a starting point. Similarly, when using the full sample SMC estimation result as a starting point, the 2000Q1 results are close to the estimates obtained for the 2019Q4 SMC estimation.

4.3 Assessing HANK's out-of-sample forecasting accuracy

Figure 1 shows the results of the horse race between BBL and SW focusing on four variables of interest: output, consumption, investment growth, and the GDP deflator inflation. For each of these variables the figure displays the root mean square errors (RMSEs), expressed in percent, computed as

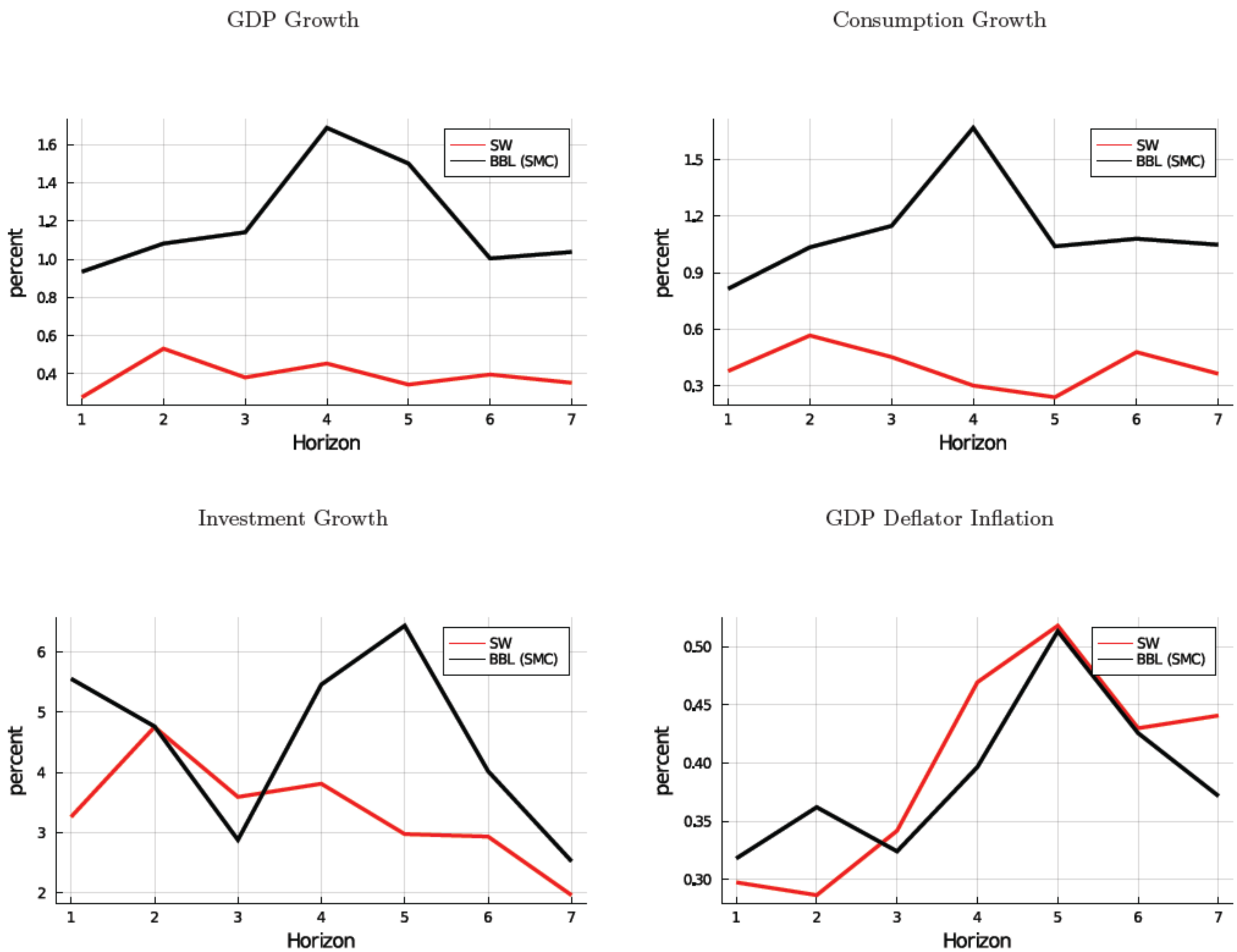
$$RMSE_{i,h,\mathcal{M}_m} = \frac{1}{\bar{\tau} - h + 1} \sum_{\tau=h}^{\bar{\tau}} \left(y_{i,T-\tau+h} - \mathbb{E}[y_{i,T-\tau+h} | y_{1:T-\tau}, \mathcal{M}_m]\right)^2 \tag{52}$$

where i indicates the variable being forecast, h is the forecast horizon, which ranges from one to seven quarters ahead, and $\mathcal{M}_m$ is the model. The model set is $\mathcal{M} = \{\text{BBL}, \text{SW}\}$, with the BBL RMSEs shown by the solid black line and the SW RMSEs by the solid red line. The BBL model is referred to as BBL (SMC) because it uses the posterior computed from the online estimation starting from the SMC draws, as opposed to the original MH draws from BBL (the SMC-based estimation performs better than the MH-based one, as shown later).

For the variables measuring real activity, in particular output and consumption growth, the results of the horse race are not very kind to the BBL model. This is especially true for consumption growth, where the RMSEs are roughly between about two (for both short and longer horizons) and six ($h = 4$) times larger for BBL. The differences in forecasting performance for consumption growth largely translate into similar differences in RMSEs for GDP growth, given that consumption represents the largest component of GDP. For investment growth, the forecasting accuracy of the two models is similar for shorter horizons, but is again worse for BBL for medium horizons. One piece of good news for the BBL model is that the RMSEs of GDP deflator inflation are comparable to those of SW for all forecast horizons.

The much worse forecasting performance for BBL compared to SW for consumption growth is particularly disappointing. The key difference between HANK and SW-type models is the following: in HANK models, the representative agent Euler equation, which determines consumption in standard DSGEs, is replaced with the aggregation of individual households' consumption policy functions. These consumption policy functions reflect inequality in both income and wealth: poor agents are hand-to-mouth, or close to

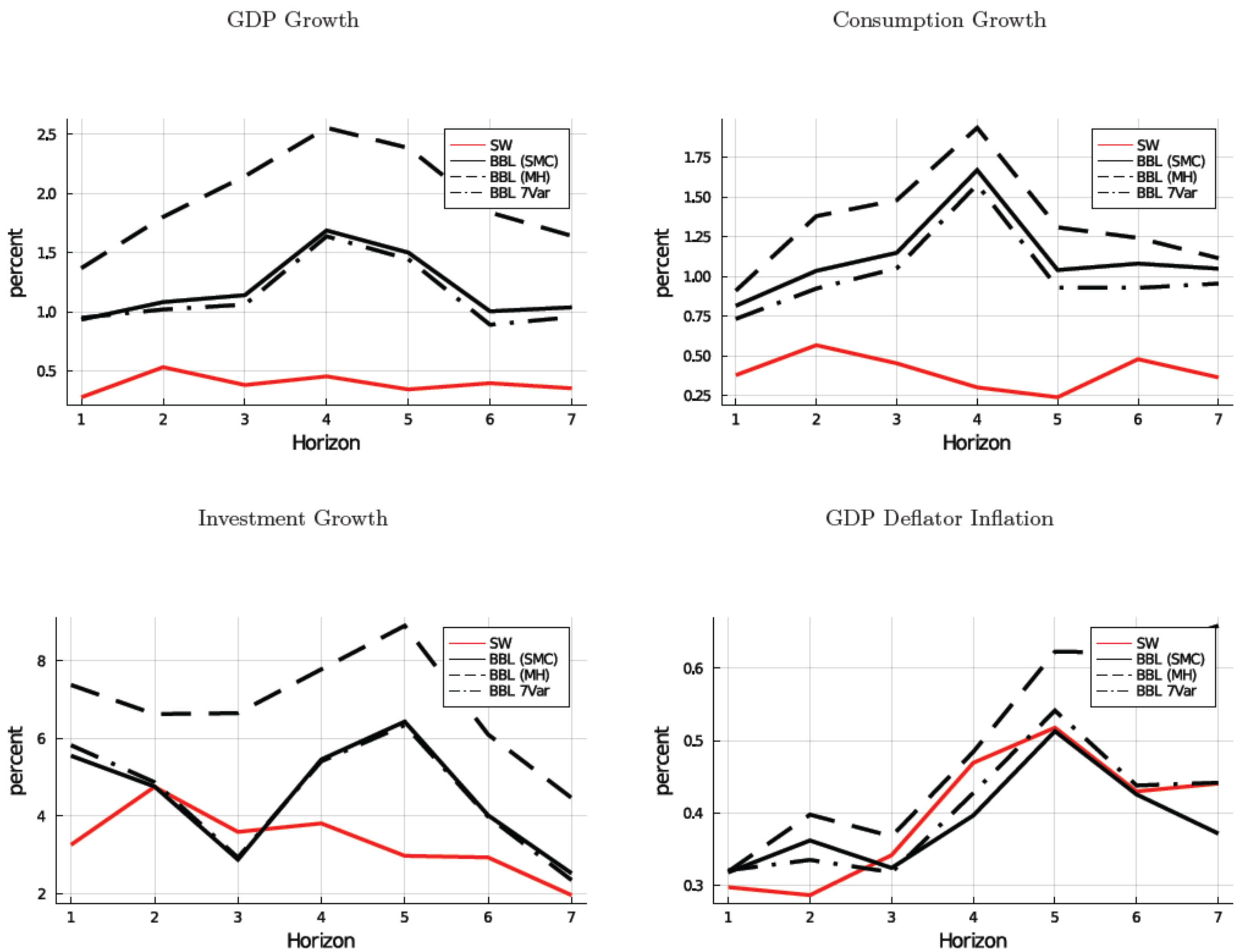
Figure 1: RMSEs: BBL vs SW



Note: The figure plots $RMSE_{i,h,\mathcal{M}_m}$ computed using expression (52) for the BBL (solid black lines) and the SW (solid red lines) models. The BBL model is referred to as BBL(SMC) since it uses the posterior computed from the online estimation starting from the SMC draws.

it, and have a high marginal propensity to consume out of income; meanwhile richer agents can substitute intertemporally and have low marginal propensities to consume. The BBL version we use to compute the RMSEs in figure 1 includes among the observables used in the estimation (and forecasting) those reflecting inequality, such as the top 10 percent income and wealth shares. One would have hoped that this much more realistic view of the world translated into a better quantitative understanding of the behavior of aggregate consumption and hence a better forecasting performance. This does not seem to be the case, at least for the

Figure 2: RMSEs: Robustness



Note: The figure plots $RMSE_{i,h,\mathcal{M}_m}$ computed using expression (52) for the seven-variable BBL model (BBL 7Var, dash-and-dotted black lines), the BBL model using the posterior computed from the online estimation starting from the MH draws, and referred to as BBL(MH) (dotted black lines), the BBL model using the posterior computed from the online estimation starting from the SMC draws, and referred to as BBL(SMC) (solid black lines), and the SW model (solid red lines).

BBL model.

Before discussing possible reasons for these findings, we show in figure 2 that the results are robust to using the results from 1) the online estimation starting from the Metropolis Hastings (MH) draws, which we refer to as BBL(MH) (dotted black lines), and 2) the model estimated using only the seven aggregate macro variables, and no measure of inequality, as observables (BBL 7Var, dash-and-dotted black lines). We find that the RMSEs obtained using the MH draws are uniformly worse than those obtained from the MH draws. We also find that the RMSEs for the eleven and the seven-variable BBL are almost indistinguishable from one another. This is somewhat disappointing from the perspective of the HANK literature, as it suggests that measures of inequality matter little for the dynamics of macroeconomic aggregates, at least for this model. The result is reminiscent of the findings in Chang et al. (2021), who use functional vector autoregressions to argue that there is limited feedback between inequality and aggregate macro time-series.

What are the possible reasons for these somewhat negative results? First, while BBL is *a priori* an ideal candidate for this forecasting comparison given that it incorporates SW's shocks and frictions, perhaps other HANK models may perform better than BBL from a forecasting point of view. Seen from this perspective, the results in this paper are an invitation to HANK modelers to use the methodology (and the code) described in this paper to see how well their model fares in terms of forecasting accuracy.

Second, the good (at least relative to VARs) forecasting performance of representative agent DSGEs à la SW was not achieved overnight, but resulted from a decade of advancement in modeling, crystallized in Christiano et al. (2005).[22] It may be that HANK models need to go through a similar process. There is also evidence (e.g., Del Negro et al., 2007) that some of the reasonable forecasting performance of representative agent DSGEs is due to features like habit persistence that 1) according to some may not have particularly strong micro-foundations, and 2) may be difficult to replicate in HANK models.

Finally, as mentioned in the introduction and discussed in section 4.2, the parameters in HANK affecting the model's steady state are calibrated, not estimated. This is for a computational reason: recomputing the steady state is extremely costly. But the estimated DSGE literature has shown that not estimating parameters, perhaps not too surprisingly, hurts the fit of DSGE models and their forecasting performance. If this is the reason why BBL forecasts are worse than those of SW, these findings pose a computational challenge to HANK researchers interested in estimation: finding ways of computing the steady state more efficiently and/or using estimation algorithms that do not require recomputing the steady state too many times.

5 Conclusion

This paper had two objectives. One was to provide a toolkit for efficient repeated estimation of HANK models that can be used by researchers at central banks and in academia. We argued that online estimation using

[22]There is a perception among macroeconomists that the reasonable forecasting performance of DSGEs is the result of hindsight: model features are chosen *ex post* so that these models produce reasonably good RMSEs. More than ten years of actual (*ex ante*) forecasting with DSGE models at the New York Fed arguably shows that this perception is unfounded (Cai et al., 2019).

Sequential Monte Carlo provides such a toolkit, and we explained how it works. The second objective was to "kick the tires" of HANK models by comparing the out-of-sample forecasting accuracy of a prominent example of such models, Bayer et al. (2022), to that of the Smets and Wouters (2007) model. HANK models did not fare too well: their forecasting performance for real activity variables, especially GDP and consumption growth, is notably inferior to that of SW. The results for consumption are particularly disappointing, given that the main difference between SW-type DSGEs and HANK models is the replacement of the representative agent Euler equation with the aggregation of individual households' consumption policy functions, which reflects inequality.

These findings should provide a motivation for more research on HANK models. First, leaving aside forecasting performance of HANK models, inequality is one of the critical issues of our times and features prominently in the transmission of policies. There are questions, such those surrounding the effect on growth and inflation of the government transfers during the COVID pandemic, that representative agent models simply cannot answer adequately. Kaplan et al. (2020) and Auclert et al. (2023) are recent examples of quantitative research based on HANK models that focuses on some of these salient policy issues. Second, since all models are misspecified, model diversity should play an important role for policymakers who use models to inform their decisions. Finally, the fact that the forecasting performance of HANK models can be improved is a just stimulus for further efforts, in terms of both modeling and making computations more efficient.

References

Ahn, SeHyoun, Greg Kaplan, Benjamin Moll, Thomas Winberry, and Christian Wolf, "When inequality matters for macro and macro matters for inequality," *NBER macroeconomics annual*, 2018, *32* (1), 1–75.

An, Sungbae and Frank Schorfheide, "Bayesian Analysis of DSGE Models," *Econometric Reviews*, 2007, *26* (2-4), 113–172.

Auclert, Adrien, Bence Bardóczy, Matthew Rognlie, and Ludwig Straub, "Using the sequence-space Jacobian to solve and estimate heterogeneous-agent models," *Econometrica*, 2021, *89* (5), 2375–2408.

_ **, Matthew Rognlie, and Ludwig Straub**, "The Trickling Up of Excess Savings," Technical Report, National Bureau of Economic Research 2023.

Bayer, Christian and Ralph Luetticke, "Solving heterogeneous agent models in discrete time with many idiosyncratic states by perturbation methods," *Quantitative Economics*, 2020, *11*, 1253–1288.

_ **, Benjamin Born, and Ralph Luetticke**, "Shocks, frictions, and inequality in US business cycles," 2022.

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

Cai, Michael, Marco Del Negro, Edward Herbst, Ethan Matlin, Reca Sarfati, and Frank Schorfheide, "Online estimation of DSGE models," *The Econometrics Journal*, 2021, *24* (1), C33–C58.

_ , _ , Marc P Giannoni, Abhi Gupta, Pearl Li, and Erica Moszkowski, "DSGE forecasts of the lost recovery," *International Journal of Forecasting*, 2019, *35* (4), 1770–1789.

Calvo, Guillermo, "Staggered Prices in a Utility Maximizing Framework," *Journal of Monetary Economics*, 1983, *12* (3), 383–398.

Cappé, Olivier, Eric Moulines, and Tobias Ryden, *Inference in Hidden Markov Models*, Springer Verlag, New York, 2005.

Chang, Minsu, Xiaohong Chen, and Frank Schorfheide, "Heterogeneity and aggregate fluctuations," Technical Report, National Bureau of Economic Research 2021.

Chopin, Nicolas, "A Sequential Particle Filter for Static Models," *Biometrika*, 2002, *89* (3), 539–551.

_ , "Central Limit Theorem for Sequential Monte Carlo Methods and its Application to Bayesian Inference," *Annals of Statistics*, 2004, *32* (6), 2385–2411.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans, "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 2005, *113*, 1–45.

Creal, Drew, "Sequential Monte Carlo Samplers for Bayesian DSGE Models," *Manuscript, University Chicago Booth*, 2007.

Del Negro, Marco and Frank Schorfheide, "Monetary Policy with Potentially Misspecified Models," *American Economic Review*, 2009, *99* (4), 1415–1450.

_ **and** _ , "Bayesian Macroeconometrics," in Herman K. van Dijk, Gary Koop, and John Geweke, eds., *Handbook of Bayesian Econometrics*, Oxford University Press, 2010.

_ **and** _ , "DSGE Model-Based Forecasting," in Graham Elliott and Allan Timmermann, eds., *Handbook of Economic Forecasting, Volume 2*, Elsevier, 2013.

_ , _ , Frank Smets, and Raphael Wouters, "On the Fit of New Keynesian Models," *Journal of Business and Economic Statistics*, 2007, *25* (2), 123 – 162.

Duane, Simon, Anthony D Kennedy, Brian J Pendleton, and Duncan Roweth, "Hybrid monte carlo," *Physics letters B*, 1987, *195* (2), 216–222.

Durham, Garland and John Geweke, "Adaptive Sequential Posterior Simulators for Massively Parallel Computing Environments," in Ivan Jeliazkov and Dale Poirier, eds., *Advances in Econometrics*, Vol. 34, Emerald Group Publishing Limited, West Yorkshire, 2014, chapter 6, pp. 1–44.

Edge, Rochelle and Refet Gürkaynak, "How Useful Are Estimated DSGE Model Forecasts for Central Bankers," *Brookings Papers of Economic Activity*, 2010, p. forthcoming.

Farkas, Mátyás and Balint Tatar, "Bayesian estimation of DSGE models with Hamiltonian Monte Carlo," Technical Report, IMFS Working Paper Series 2020.

Fernández-Villaverde, Jesús and Juan F Rubio-Ramírez, "Estimating macroeconomic models: A likelihood approach," *The Review of Economic Studies*, 2007, *74* (4), 1059–1087.

Ferriere, Axelle and Gaston Navarro, "The Heterogeneous Effects of Government Spending: Itâ€™ S All About Taxes," *FRB International Finance Discussion Paper*, 2018, (1237).

Gelman, Andrew, John B Carlin, Hal S Stern, and Donald B Rubin, *Bayesian data analysis*, Chapman and Hall/CRC, 1995.

Geweke, John, *Contemporary Bayesian econometrics and statistics*, John Wiley & Sons, 2005.

Greenwood, Jeremy, Zvi Hercowitz, and Gregory W. Huffman, "Investment, Capacity Utilization, and the Real Business Cycle," *The American Economic Review*, 1988, *78* (3), 402–417.

Hagedorn, Marcus, Iourii Manovskii, and Kurt Mitman, "Monetary Policy in Incomplete Market Models: Theory and Evidence," Technical Report, University of Pennsylvania Working Paper 2018.

Hammersley, John M and K William Morton, "Poor man's monte carlo," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1954, *16* (1), 23–38.

Herbst, Edward, "Using the "Chandrasekhar Recursions" for Likelihood Evaluation of DSGE Models," *Computational Economics*, 2015, *45* (4), 693–705.

_ **and Frank Schorfheide**, "Sequential Monte Carlo Sampling for DSGE Models," *Journal of Applied Econometrics*, 2014, *29* (7), 1073–1098.

_ **and** _ , *Bayesian Estimation of DSGE Models*, Princeton University Press, 2015.

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

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

_ **,** _ **, and** _ , "The great lockdown and the big stimulus: Tracing the pandemic possibility frontier for the US," Technical Report, National Bureau of Economic Research 2020.

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

Lee, Donggyu, "Quantitative easing and inequality," Technical Report, mimeo 2021.

Liu, Jun S, *Monte Carlo Strategies in Scientific Computing*, Springer Verlag, New York, 2001.

Mlikota, Marko and Frank Schorfheide, "Sequential Monte Carlo With Model Tempering," *arXiv preprint arXiv:2202.07070*, 2022.

Müller, Ulrich K, "Measuring prior sensitivity and prior informativeness in large Bayesian models," *Journal of Monetary Economics*, 2012, *59* (6), 581–597.

Neal, Radford M et al., "MCMC using Hamiltonian dynamics," *Handbook of markov chain monte carlo*, 2011, *2* (11), 2.

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

Smets, Frank and Raf Wouters, "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 2007, *97* (3), 586 – 606.

Stan Development Team, "Stan: A C++ library for probability and sampling, version 2.8. 0," 2015.

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

Wu, Jing Cynthia and Fan Dora Xia, "Measuring the macroeconomic impact of monetary policy at the zero lower bound," *Journal of Money, Credit and Banking*, 2016, *48* (2-3), 253–291.