CENTER FOR ADVANCED RESEARCH IN FINANCE

CARF Working Paper

CARF-F-535

Estimating a Behavioral New Keynesian Model with the Zero Lower Bound

Yasuo Hirose
Keio University

Hirokuni Iiboshi
Tokyo Metropolitan University

Mototsugu Shintani
The University of Tokyo

Kozo Ueda
Waseda University

March 2022

CARF is presently supported by Nomura Holdings, Inc., Sumitomo Mitsui Banking Corporation, The Dai-ichi Life Insurance Company, Limited, The Norinchukin Bank, MUFG Bank, Ltd. and Ernst & Young ShinNihon LLC. This financial support enables us to issue CARF Working Papers.

Estimating a Behavioral New Keynesian Model with the Zero Lower Bound*

Yasuo Hirose † Hirokuni Iiboshi‡ Mototsugu Shintani§ Kozo Ueda¶

March 2022

Abstract

We estimate a New Keynesian model incorporating two notable features: bounded rationality and the zero lower bound on the nominal interest rate. Our Bayesian estimation of a fully nonlinear model shows that the model with bounded rationality better fits the US data than its rational expectations counterpart and that both households and firms exhibit a substantial degree of bounded rationality. Moreover, we demonstrate that bounded rationality expands a parameter region in which the model can be estimated and weakens the power of forward guidance.

Keywords: Behavioral DSGE model; Bounded rationality; Zero lower bound; Bayesian inference; Nonlinear DSGE model

JEL Classification: C13, C63, E43, E52, E70

*We thank Takashi Kano, Taisuke Nakata, Takayuki Tsuruga, and the conference and seminar participants at the University of Tokyo for their useful comments and suggestions. The authors are grateful for financial support from the Japan Society for the Promotion of Science (Hirose: 19K01560, 21H04397; Iiboshi, Shintani, and Ueda: 19H01491) and the Zengin Foundation for Studies on Economics and Finance (Ueda).

†Keio University (E-mail: yhirose@econ.keio.ac.jp)

‡Tokyo Metropolitan University (E-mail: iiboshi@tmu.ac.jp)

§The University of Tokyo (E-mail: shintani@e.u-tokyo.ac.jp)

¶Waseda University (E-mail: kozo.ueda@waseda.jp)

1 Introduction

A deviation from the rational expectations (RE) assumption likely changes the implications of dynamic macroeconomic models. Gabaix (2020) proposes a tractable approach to incorporating bounded rationality (BR) in an otherwise standard New Keynesian (NK) model to analyze its implications for macroeconomic policy. He demonstrates that bounded rationality considerably changes the properties of the model. However, the results are based on a linearized model with calibrated parameters.

This study addresses the following two questions: whether the BR assumption improves the fit of a model to US macroeconomic time series, including the virtually zero interest rate period, and to what extent the estimated degree of BR changes implications for macroeconomic analyses. The primary goal of this paper is to provide quantitative answers to these two questions. To this end, we estimate a NK model under BR incorporated with the zero lower bound (ZLB) on the nominal interest rate in a fully nonlinear and stochastic setting using Bayesian methods. The model follows from Gabaix (2020), who introduces cognitive myopia (discounting) to capture BR. Agents are partially myopic, shrinking their expectations about the future state of the economy toward its steady state. For a better fit to the data, we extend Gabaix (2020)'s model by incorporating habit persistence in consumption preferences, a stochastic trend in technological progress, and monetary policy smoothing.

To quantify the effects of BR, considering nonlinearity in the model, particularly the ZLB, is of crucial importance. It is important not only because the ZLB has constrained actual monetary policy since the Great Recession, but also because BR can alter the quantitative implications of the ZLB. Richter and Throckmorton (2015) argue that, in the economy with the ZLB, monetary policy needs to be more aggressive against inflation than is suggested by the Taylor principle to find a nonlinear RE solution. This result comes from the fact that a negative shock to the economy is reinforced by the contractionary effects of the ZLB. However, as Gabaix (2020) indicates, BR remarkably weakens the contractionary effects of the ZLB. It is important to note that incorporating BR expands a parameter region in which

the model can be estimated and enables us to obtain parameter estimates that would not be obtained in the estimation of its RE counterpart. Moreover, BR can provide a favorable implication for the power of forward guidance, which is important, especially when monetary policy is constrained by the ZLB. It has been well known that, under the RE, standard NK models grossly over-predict the effect of forward guidance (the so-called forward guidance puzzle).[1] In this regard, Gabaix (2020) shows that BR undermines an expectation channel through which forward guidance works and dramatically weakens its effect. Thus, in this paper, we quantitatively evaluate the power of forward guidance in the estimated BR model. Because we employ a fully nonlinear model, the effect of forward guidance is state dependent. We demonstrate how the power of forward guidance can change, depending on whether or not the ZLB constraint binds.

Our estimation results are threefold. First, the BR model better fits the US data than its RE counterpart. This outcome occurs because the BR model can more effectively replicate the ZLB episodes in the aftermath of the global financial crisis. The estimated parameters on cognitive discounting suggest a substantial degree of BR; that is, both households and firms discount the future more, by around 15%, than those under the RE.

Second, given the estimated parameters, we demonstrate that BR considerably expands the region of the model's parameter space in which the policy function iteration converges to an equilibrium solution. More specifically, with the estimated BR parameters, we can obtain convergence for all the positive values for monetary policy coefficients on inflation and output; whereas these coefficients are required to be larger in the nonlinear RE model with the ZLB than those satisfying the Taylor principle in the linearized RE model without the ZLB. This finding suggests that the estimation of an RE model would be constrained by a requirement for convergence, which might, in turn, yield an upward bias in the estimates of the monetary policy coefficients. On the other hand, the convergence requirement cannot

[1] See Del Negro, Giannoni, and Patterson (2015) and McKay, Nakamura, and Steinsson (2016). Nakata et al. (2019) discuss optimal monetary policy when the power of forward guidance is attenuated because of discounting.

be a constraint for the estimation of the BR model and is unlikely to generate such a bias. We also find that firms' cognitive discounting parameter plays a more important role in expanding the convergence region than that of households.

Third, we evaluate the power of forward guidance in the estimated BR model and compare it with that in its RE counterpart by adding a one-period-ahead monetary policy news shock as a proxy for forward guidance. The simulation analysis shows that BR substantially weakens the effect of forward guidance on inflation and output. Besides, our fully nonlinear approach enables us to find that the effect of forward guidance is further weakened when there is hardly any room for an interest rate cut.

To estimate a nonlinear NK model incorporating the ZLB, we employ the approach that is used in Iiboshi, Shintani, and Ueda (forthcoming), which consists of two useful methods. First, to solve a (bounded) rational expectations equilibrium in the model with the ZLB, we use the *time iteration with linear interpolation* (TL) method (see Richter, Throckmorton, and Walker 2014). Second, to estimate the model, we use the *sequential Monte Carlo squared* (SMC^2) method (see Chopin, Jacob, and Papaspiliopoulos 2013 and Herbst and Schorfheide 2015). Although the TL and SMC^2 methods have many advantages, they are still computationally intensive. To facilitate the computations, we use a simple NK model with BR, which abstracts capital formation and wage stickiness.

A handful of studies have attempted to estimate NK models with BR in the form of cognitive discounting based on Gabaix (2020), including Gabaix (2018), Andrade, Coredeiro, and Lambais (2019), Ilabaca, Meggiorini, and Milani (2020), Afsar et al. (2020). Gabaix (2018), which is an earlier version of Gabaix (2020), estimates a behavioral NK model using a Bayesian technique and documents the significant degree of BR. However, he notes that his result is preliminary and further well-identified empirical work is necessary. Our paper is different from these studies in that the model is fully nonlinear and incorporates the ZLB. Our paper is also related to empirical studies on cognitive discounting, which are based on a partial, rather than general, equilibrium framework. This strand of the literature includes

Gali and Gertler (1999) and Linde (2005) on the Phillips curve, Fuhrer and Rudebusch (2004) on the Euler equation, and Ganong and Noel (2017) on micro-level cognitive discounting.

Our analysis, which solves and estimates a nonlinear NK model with the ZLB, is closely related to previous studies by Gust et al. (2017), Plante, Richter, and Throckmorton (2018), and Iiboshi, Shintani, and Ueda (forthcoming). The most important difference is that their models are based on rational expectations, whereas our model incorporates bounded rationality.

The remainder of this paper is structured as follows. After Section 2 describes our behavioral model, Section 3 explains our estimation methods. Section 4 presents the estimation results and empirical consequences of BR. Section 5 concludes.

2 Behavioral Model

In the the model economy, there are behavioral households, behavioral firms, and a central bank. The firms consist of monopolistically competitive intermediate-good producers facing price stickiness and a perfectly competitive final-good producer. The central bank sets the nominal interest rate following a Taylor-type monetary policy rule that is constrained by the ZLB. The economy is subject to three types of exogenous shocks: a discount factor (preference) shock, a technology shock, and a monetary policy shock. For a better fit to the data, the model incorporates habit persistence in consumption preferences, a stochastic trend in output, and monetary policy smoothing.

Following Gabaix (2020), we introduce cognitive discounting for behavioral agents. In comparison with the RE case, the agents further discount their expectations for k period ahead by a factor M^k; that is, $\mathbb{E}_t^{BR}[X_{t+k}] = M^k \mathbb{E}_t[X_{t+k}]$, where $\mathbb{E}_t^{BR}$ and $\mathbb{E}_t$ represent expectation operators in period t under the BR and RE assumptions, respectively, and X_t is a state vector with zero mean.[2] Moreover, as in Gabaix (2020), we assume that

[2]Gabaix (2020) refers to M as an aggregate-level attention parameter, differentiating it from an individual's cognitive discounting factor $\bar{m}$. In his benchmark model and our model, M equals $\bar{m}$.

behavioral firms discount their expectations differently from behavioral households; that is, their cognitive discount factor is denoted by M^f.

2.1 Households

Each behavioral household i ($\in [0,1]$) maximizes the intertemporal utility function

$$\mathbb{E}_t^{BR}\left[\sum_{j=0}^{\infty}\beta^j Z_{t+j}^b\left\{\frac{(C_{i,t+j}-hC_{t+j-1})^{1-\sigma}}{1-\sigma}-\chi\frac{(A_{t+j})^{1-\sigma}\,l_{i,t+j}^{1+\omega}}{1+\omega}\right\}\right], \tag{1}$$

subject to the budget constraint $C_{i,t}+B_{i,t}/P_t \leq W_t l_{i,t}+R_{t-1}B_{i,t-1}/P_t+T_t$, where $C_{i,t}$, $l_{i,t}$, P_t, W_t, R_t, and T_t represent consumption, labor services, the aggregate price level, real wage, the nominal rate of return, and the lump-sum transfer, respectively, in period t. In addition, $B_{i,t}$ is the holding of one-period riskless bonds at the end of period t. Parameter $\beta(\in (0,1))$ is the subjective discount factor; $\sigma(>0)$ is the inverse of the intertemporal elasticity of substitution of consumption; $h(\in [0,1))$ measures the degree of external habit persistence in consumption preferences; $\omega(>0)$ is the inverse of the labor supply elasticity; and $\chi(>0)$ is the scale parameter for the disutility from working. As in Erceg, Guerrieri, and Gust (2006), we allow preferences for leisure to shift with the nonstationary technology level, A_t, to ensure the existence of a balanced growth path. Finally, Z_t^b represents a shock to the discount factor (preference) and follows a first-order autoregressive (AR(1)) process:

$$\log(Z_t^b)=\rho_b\log(Z_{t-1}^b)+\epsilon_t^b, \tag{2}$$

where $\rho_b \in [0,1)$ and $\epsilon_t^b \sim i.i.d.\ N(0,\sigma_b^2)$.

The first-order conditions for optimal decisions on consumption, labor supply, and bond-

holding are identical across households and therefore can be written without subscript i:

$$\Lambda_t A_t^{\sigma} = \left(C_t/A_t - hC_{t-1}/A_{t-1}\left(A_{t-1}/A_t\right)\right)^{-\sigma}, \tag{3}$$

$$\Lambda_t W_t = A_t^{1-\sigma}\chi l_t^{\omega}, \tag{4}$$

$$1 = \beta \mathbb{E}_t^{BR}\left[\frac{\Lambda_{t+1}A_{t+1}^{\sigma}}{\Lambda_t A_t^{\sigma}}\left(\frac{A_{t+1}}{A_t}\right)^{-\sigma}\frac{Z_{t+1}^b}{Z_t^b}\frac{R_t}{\pi_{t+1}}\right], \tag{5}$$

where Λ_t denotes the Lagrange multiplier and $\pi_t = P_t/P_{t-1}$ is the (gross) inflation rate. As is assumed in Gabaix (2020), the agents anchor on the correct steady state. Subtracting the steady state value (on the balanced growth path) from both sides of equation (5), we obtain the behavioral household's Euler equation:

$$(\Lambda_t A_t^{\sigma})A_t^{-\sigma}Z_t^b - \beta(\Lambda A^{\sigma})\gamma_a^{-\sigma}A_t^{-\sigma}\frac{R}{\pi} = \mathbb{E}_t^{BR}\left[\beta(\Lambda_{t+1}A_{t+1}^{\sigma})A_{t+1}^{-\sigma}Z_{t+1}^b\frac{R_t}{\pi_{t+1}} - \beta(\Lambda A^{\sigma})\gamma_a^{-\sigma}A_t^{-\sigma}\frac{R}{\pi}\right]$$

$$\Leftrightarrow (\Lambda_t A_t^{\sigma})A_t^{-\sigma}Z_t^b - (\Lambda A^{\sigma})A_t^{-\sigma} = M\mathbb{E}_t\left[\beta(\Lambda_{t+1}A_{t+1}^{\sigma})A_{t+1}^{-\sigma}Z_{t+1}^b\frac{R_t}{\pi_{t+1}} - (\Lambda A^{\sigma})A_t^{-\sigma}\right]$$

$$\Leftrightarrow 1 - \frac{\Lambda A^{\sigma}}{\Lambda_t A_t^{\sigma}}\frac{1}{Z_t^b} = M\mathbb{E}_t\left[\beta\frac{\Lambda_{t+1}A_{t+1}^{\sigma}}{\Lambda_t A_t^{\sigma}}\left(\frac{A_{t+1}}{A_t}\right)^{-\sigma}\frac{Z_{t+1}^b}{Z_t^b}\frac{R_t}{\pi_{t+1}} - \frac{\Lambda A^{\sigma}}{\Lambda_t A_t^{\sigma}}\frac{1}{Z_t^b}\right], \tag{6}$$

where the variables without subscript t denote their steady-state values and M represents the behavioral households' cognitive discounting parameter measuring attention to the future.[3] Note that we use the steady-state relationship for the real interest rate; that is, $R/\pi = \gamma_a^{\sigma}/\beta(\geq 1)$, where γ_a is the stady-state growth rate of A_t.

2.2 Firms

The representative final-good firm produces output Y_t under perfect competition by choosing a combination of intermediate inputs $Y_{f,t}$ so as to maximize its profit $P_t Y_t - \int_0^1 P_{f,t}Y_{f,t}\,df$

[3] In the present paper, we extend the linearized model of Gabaix (2020) to its non-linear version by assuming $\mathbb{E}_t^{BR}[\cdot_{t+1}] = M\mathbb{E}_t[\cdot_{t+1}]$. Although this assumption greatly simplifies our analysis, it does not necessarily ensure consistency with the sparse behavioral dynamic programming that serves as the basis of Gabaix (2020) in a nonlinear setting.

subject to the Dixit–Stiglitz aggregator $Y_t = \left\{ \int_0^1 Y_{f,t}^{\frac{\varepsilon-1}{\varepsilon}} df \right\}^{\frac{\varepsilon}{\varepsilon-1}}$, where $P_{f,t}$ is the price of intermediate good f and $\varepsilon(>1)$ represents the elasticity of substitution among intermediate goods.

Each monopolistically competitive intermediate-good firm $f(\in [0,1])$ produces one kind of differentiated goods $Y_{f,t}$ by choosing a cost-minimizing labor input $l_{f,t}$, given the real wage W_t, subject to the production function:

$$Y_{f,t} = A_t l_{f,t}, \tag{7}$$

where A_t represents the technology level and the log of A_t follows the non-stationary stochastic process:

$$\log A_t = \log \gamma_a + \log A_{t-1} + \mu_t^a, \tag{8}$$

where γ_a denotes the steady-state (gross) rate of technological change and μ_t^a is the technology shock, which follows a stationary AR(1) process:

$$\mu_t^a = \rho_a \mu_{t-1}^a + \epsilon_t^a, \tag{9}$$

where $\rho_a \in [0,1)$ and $\epsilon_t^a \sim i.i.d.\ N(0, \sigma_a^2)$.

Then, each intermediate-good firm f maximizes its firm value by setting the optimal price $P_{f,t}$ in period t in the presence of a Rotemberg-type price adjustment cost:

$$\mathbb{E}_t^{BR} \left[\sum_{j=0}^{\infty} \beta^j \frac{\Lambda_{t+j} Z_{t+j}^b}{\Lambda_t Z_t^b} \left(\frac{P_{f,t+j}}{P_{t+j}} - \frac{W_{t+j}}{A_{t+j}} - \frac{\phi}{2} \left(\frac{P_{f,t+j}}{P_{f,t+j-1}} - \pi \right)^2 \right) Y_{f,t+j} \right] \tag{10}$$

subject to the final-good firm's demand curve $Y_{f,t} = (P_{f,t}/P_t)^{-\varepsilon} Y_t$, where ϕ is the parameter for the adjustment cost.

Combining the first-order condition with equation (4) and the aggregate production function $Y_t = A_t l_t$ yields

$$1-\phi\left(\pi_t-\pi\right)\pi_t-\varepsilon\left[1-\chi\frac{\left(Y_t/A_t\right)^{\omega}}{\Lambda_t A_t^{\sigma}}-\frac{\phi}{2}\left(\pi_t-\pi\right)^2\right]$$
$$+\beta\mathbb{E}_t^{BR}\left[\frac{\Lambda_{t+1}A_{t+1}^{\sigma}}{\Lambda_t A_t^{\sigma}}\left(\frac{A_{t+1}}{A_t}\right)^{1-\sigma}\frac{Z_{t+1}^b}{Z_t^b}\phi\left(\pi_{t+1}-\pi\right)\pi_{t+1}\frac{Y_{t+1}/A_{t+1}}{Y_t/A_t}\right]=0,$$

or

$$1-\phi\left(\pi_t-\pi\right)\pi_t-\varepsilon\left[1-\chi\frac{\left(Y_t/A_t\right)^{\omega}}{\Lambda_t A_t^{\sigma}}-\frac{\phi}{2}\left(\pi_t-\pi\right)^2\right]$$
$$+\beta M^f\mathbb{E}_t\left[\frac{\Lambda_{t+1}A_{t+1}^{\sigma}}{\Lambda_t A_t^{\sigma}}\left(\frac{A_{t+1}}{A_t}\right)^{1-\sigma}\frac{Z_{t+1}^b}{Z_t^b}\phi\left(\pi_{t+1}-\pi\right)\pi_{t+1}\frac{Y_{t+1}/A_{t+1}}{Y_t/A_t}\right]=0, \tag{11}$$

where M^f denotes the behavioral firms' cognitive discounting parameter.

The final-good market clearing condition is given by

$$Y_t=C_t+\phi\left(\pi_t-\pi^*\right)^2Y_t/2. \tag{12}$$

Because the technology level A_t is nonstationary, as specified by equation (8), we rewrite the equilibrium conditions in terms of stationary variables detrended by A_t: $y_t = Y_t/A_t$, $c_t = C_t/A_t$, $\lambda_t = \Lambda_t A_t^{\sigma}$, so that we can derive a nonstochastic steady state for the detrended variables.

The detrended equilibrium conditions and the steady-state relationships are shown in Appendix A.1 and A.2, respectively.

2.3 Central Bank

Without the ZLB constraint, the central bank sets the nominal interest rate R_t in response to the deviations of inflation, detrended output, and output growth from their steady-state

values, as in Gust et al. (2017). More specifically, its monetary policy rule is given by

$$R_t = (R_{t-1})^{\rho_r} \left(R \left(\frac{\pi_t}{\pi} \right)^{\psi_\pi} \left(\frac{y_t}{y} \right)^{\psi_y} \left(\frac{Y_t}{Y_{t-1}\gamma_a} \right)^{\psi_{\Delta y}} \right)^{1-\rho_r} e^{\epsilon_t^r}, \tag{13}$$

where $\rho_r (\in [0, 1))$ captures interest rate smoothing, ψ_π, ψ_y, and $\psi_{\Delta y} (> 0)$ represent the degrees of monetary policy responses to the target variables, and $\epsilon_t^r \sim i.i.d.N(0, \sigma_r^2)$ is the monetary policy shock.

With the ZLB constraint, the nominal interest rate R_t cannot be less than one, so that the monetary policy rule is described as

$$R_t = \max[R_t^*, 1], \tag{14}$$

where R_t^* denotes the notional interest rate that the central bank would set, according to the monetary policy rule specified as equation (13) in the absence of the ZLB. We consider the following two specifications for R_t^*, depending on the choice of the lagged interest rate in (13).

Nominal Rate Model

In the first specification, the observed nominal interest rate enters as the lagged interest rate:

$$R_t^* = (R_{t-1})^{\rho_r} \left(R \left(\frac{\pi_t}{\pi} \right)^{\psi_\pi} \left(\frac{y_t}{y} \right)^{\psi_y} \left(\frac{Y_t}{Y_{t-1}\gamma_a} \right)^{\psi_{\Delta y}} \right)^{1-\rho_r} e^{\epsilon_t^r}. \tag{15}$$

This specification has been employed, for example, by Aruoba, Cuba-Borda, and Schorfheide (2018). Since R_t^* depends on the nominal interest rate in the previous period, we refer to the model with this specification as the *nominal rate model*.

Notional Rate Model

In the second specification, the notional interest rate appears as the lagged interest rate:

$$R_t^* = \left(R_{t-1}^*\right)^{\rho_r} \left(R \left(\frac{\pi_t}{\pi}\right)^{\psi_\pi} \left(\frac{y_t}{y}\right)^{\psi_y} \left(\frac{Y_t}{Y_{t-1}\gamma_a}\right)^{\psi_{\Delta y}} \right)^{1-\rho_r} e^{\epsilon_t^r}. \tag{16}$$

This specification has been employed by Gust et al. (2017) and Plante, Richter, and Throckmorton (2018). We refer to the model with this specification as the *notional rate model*.

The notional rate model differs from the nominal rate model only in the choice of the lagged interest rate in the interest smoothing. With this change, however, the notional rate model induces a stronger commitment to the zero-interest-rate policy in the future than does the nominal rate model. Once the notional rate takes a value below one due to adverse shocks to the economy and depends on the past notional rate, it is more likely to be below one in the future periods, implying that the central bank will keep the policy rate at zero for longer periods. As a result, the central bank can compensate for its inability to lower the policy rate below zero by promising to continue the zero-interest-rate policy.

3 Methodology

This section outlines how we solve and estimate the model with the ZLB in a fully nonlinear and stochastic setting. We then describe the data and prior specifications used for estimation. For details about the solution and estimation methods, see Appendix B.1 and B.2, respectively.

3.1 Model Solution

To solve for the policy functions satisfying the detrended equilibrium conditions, we employ a type of policy function iteration called the *time iteration with linear interpolation* (TL) method. The TL method has been recommended by Richter, Throckmorton, and Walker

(2014) as it is well balanced in terms of computational speed and approximation accuracy, compared with other alternative methods. The TL method begins with a time iteration for policy functions until the equilibrium conditions are satisfied at every node. Then, a standard linear interpolation is used to interpolate the policy functions for future variables. In comparison with the projection method using the Chebyshev polynomial basis, the linear interpolation is expected to perform better when the policy functions exhibit kinks due to the ZLB.

In the subsequent analysis, we investigate how the BR assumption can expand the region of the model's parameter space in which the policy function iteration converges to an equilibrium solution. We regard the solution as convergent if the following three conditions are met: (i) the policy functions converge to a tolerance size of 10^{-5} within 200 iterations; (ii) the ZLB binds fewer than 50 percent of the nodes in the state space, following Richter and Throckmorton (2015); and (iii) the policy functions do not jump during the time iteration to guarantee contraction mapping. Otherwise, the solution is classified as non-convergent.

3.2 Estimation

To estimate the nonlinear model with the ZLB, we employ the SMC^2 method. The "squared" means that the sequential Monte Carlo (SMC) algorithm is used for two objectives. First, for each draw of parameters, it evaluates the likelihood of a nonlinear model by generating particles that represent the states of endogenous variables. This part is often referred to as the particle filter. Second, it approximates the posterior distribution of parameters by sampling the draws of parameters as particles.

This method comprises the following four steps. In Step 1 (initialization), we draw N_θ particles for parameters θ. We then repeat Steps 2 to 4 below for N_ϕ stages. In Step 2 (correction), given θ, we compute the likelihood $p(\mathbf{Y}_t|\theta)$ and normalized weight $\tilde{W}$. In Step 3 (selection), we resample θ together with unnormalized weight W, which is based on θ in the previous stage and $\tilde{W}$ in the previous step. Then, in Step 4 (mutation), we propagate θ

and W using the Metropolis–Hastings algorithm.

Specifically, in Step 2, we adopt the tempered particle filter (TPF) proposed by Herbst and Schorfheide (2019) to approximate the likelihood $p(\mathbf{Y}_t|\theta)$. The TPF can further equalize particle weights by gradually reducing the measurement error variances, which enables us to make a much more accurate particle approximation of the likelihood than the conventional bootstrap particle filter (BPF). Moreover, the TPF can approximate the likelihood accurately even if the number of particles is relatively small.[4] This feature enables us to estimate our model with a considerably smaller number of particles, as specified below, than those in the previous studies that estimate nonlinear NK models with the ZLB using the BPF.[5]

We set the numbers of particles for states and parameters as $N_S = 10,000$ and $N_\theta = 1,200$, respectively. The number of stages for approximating the posterior distribution of parameters is set at $N_\phi = 10$. Following Herbst and Schorfheide (2019), the number of TPF stages is adaptively determined in the algorithm and results in at most three to four in our estimation. By utilizing parallel computing, a single estimation takes about a week using a workstation with a 32-core processor (Intel Xeon E5-2698v3).

3.3 Data

The model is estimated using three quarterly time series for the US from 1983:Q1 to 2019:Q4: the per capita real GDP growth rate ($100\Delta \log GDP_t$), the inflation rate of the GDP implicit price deflator ($100\Delta \log PGDP_t$), and the (quarterly) federal funds rate (FF_t). The federal funds rate was strictly above zero even when the Fed conducted the virtually zero-interest policy from 2009 to 2015, the minimum value being 0.0175% on a quarterly basis. To regard such a very small value as binding at zero, we replace the quarterly federal funds rate data that are smaller than or equal to 0.05% with zeros.[6] In Figure 1, the solid lines show these

[4]Herbst and Schorfheide (2019) demonstrate that, for a given level of accuracy, the TPF's 4,000 particles are comparable to 40,000 particles of the conventional BPF in estimating a small-scale NK model.

[5]Gust et al. (2017) generate 500,000 particles for a medium-scale NK model. Plante, Richter, and Throckmorton (2018) adopt 40,000 particles for a small-scale NK model.

[6]Gust et al. (2017) make no adjustment regarding the effective lower bound, whereas Plante, Richter, and Throckmorton (2018) set the lower bound equal to the minimum federal funds rate of 0.017%. We set

three time series.

The observation equations that relate the data to model variables are given by

$$\begin{pmatrix} 100\Delta \log GDP_t \\ 100\Delta \log PGDP_t \\ FF_t \end{pmatrix} = \begin{pmatrix} 100 \log \left(\frac{y_t}{y_{t-1}} \gamma_a e^{\mu^a_{t+1}} \right) \\ 100 \log \pi_t \\ 100 \log R_t \end{pmatrix} + u_t,$$

where $u_t = (u_t^y, u_t^\pi, u_t^R)'$ are measurement errors and $u_t \sim N(0, \Sigma_u)$.

3.4 Prior Specifications

Before estimation, we fix some parameter values. The curvature parameter σ in the households' utility function is fixed at 1.5. This choice is necessary because Gabaix (2020) points out an identification issue between the cognitive discount factor M and this σ, in that both parameters affect the degree of intertemporal substitution in the Euler equation. Since one of the primary objectives of this paper is to estimate the behavioral parameter M, we fix σ. Moreover, we fix the subjective discount factor β at 0.998 to ensure the contraction mapping in solving the nonlinear model for all the possible draws of parameters. Further, we fix $\chi = 1$ and $\varepsilon = 6$ to avoid identification problems.

All the other parameters are estimated.[7] For convenience, we transform some of the parameters as follows. Instead of estimating the adjustment cost parameter ϕ, we define the slope of the Phillips curve $\kappa = (\varepsilon - 1)\left[\omega + \sigma/(1 - h/\gamma_a)\right]/(\phi\pi)$ and estimate it. Also, rather than estimating the steady-state growth rate γ_a and inflation rate π in gross terms, we estimate these rates in net percentage terms: $\bar{a} = 100 \log \gamma_a$ and $\bar{\pi} = 100 \log \pi$.

Table 1 shows the prior distributions of parameters. Whereas most of the priors are in

the lower bound at 0.05%, which is slightly larger than those in the previous studies. This choice is needed to correctly capture the virtually zero-interest rate periods in our estimation with the relatively small size of the measurement error on the nominal interest rate.

[7] We conduct the identification tests of Iskrev (2010), Komunjer and Ng (2011), and Qu and Tkachenko (2012) and confirm that all estimated parameters are locally identified at the prior mean specified below. See also Andrade, Coredeiro, and Lambais (2019) for discussion about identification in a behavioral NK model.

line with those in previous studies, such as Smets and Wouters (2007) and Iiboshi, Shintani, and Ueda (forthcoming), the cognitive discount factors, M and M^f, are specific in our empirical analysis. For these priors, we impose the beta distribution with mean 0.85 and 0.80, respectively, and the common standard deviation 0.05. The prior mean values follows from calibration in Gabaix (2020). As for the steady-state rates of output growth and inflation, $\bar{a}$ and $\bar{\pi}$, the prior mean is set equal to the corresponding sample mean.

The sizes of measurement errors for the three observables ($100\Delta \log GDP_t$, $100\Delta \log PGDP_t$, and FF_t) are respectively set at 6.25%, 6.25%, and 0.25% of their sample variances in the data. These sizes are much smaller than the 25% in Gust et al. (2017) and the 10% in Plante, Richter, and Throckmorton (2018). In particular, we set the very small measurement error for the federal funds rate so that its filtered estimate can well track the data during the zero interest rate periods.

4 Empirical Consequences

4.1 Estimation Results

Table 2 shows the posterior distributions of parameters and the marginal likelihoods for the following four models: the BR notional rate model (i.e., the BR model, where the notional rate depends on its lagged value), the BR nominal rate model (i.e., the BR model, where the notional rate depends on the lagged (observed) nominal interest rate), the RE notional rate model, and the RE nominal rate model.

The marginal likelihood for the BR notional rate model is the highest, followed by that for the BR nominal rate model. The marginal likelihoods then drop substantially for the remaining two RE models. Between the two RE models, the nominal rate model exhibits a better fit to the data than the notional rate model. Thus, in what follows, we mainly analyze the estimation results for the BR notional rate model and compare them with those for its RE counterpart.

In the BR notional model, M and M^f are both close to 0.85. In his working paper, Gabaix (2018) obtains the estimates of M and M^f around 0.7 and 0.9, respectively. Ilabaca, Meggiorini, and Milani (2020) obtain a larger degree of BR (i.e., 0.7 for M and 0.4 for M^f). Other parameter estimates are basically not considerably different between the BR and RE models. However, we find that the habit persistence parameter h is somewhat smaller in the BR model (0.36) than in the RE model (0.63). Moreover, the interest rate smoothing parameter ρ_r is larger in the BR model (0.77) than in the RE model (0.47).

Figure 1 compares the filtered median estimates (dashed lines) of output growth, inflation, and the nominal interest rate, which are based on the estimated BR notional rate model, with the corresponding data (solid lines). The shaded areas show the 90% credible intervals of the filtered series.[8] The filtered median estimates trace the actual fluctuations well; however, the filtered inflation rate failed to replicate the huge drop in the actual rate when the global financial crisis was triggered by the collapse of Lehman Brothers in 2008:Q3 and around the periods when the policy rate was raised from zero in 2016:Q1.

Monetary Policy during the Zero Interest Rate Period

As addressed above, in terms of the fit of the model to the data, the BR notional rate model performs the best, followed by the BR nominal rate model. By contrast, between the two RE models, the notional rate model performs considerably worse than the nominal rate model. This result is noteworthy in that Gust et al. (2017) and Plante, Richter, and Throckmorton (2018) assume the notional rate model in their estimation of RE models, which corresponds to the worst performing model of the four considered in our analysis. One important difference from Gust et al. (2017) and Plante, Richter, and Throckmorton (2018) is that our observation period includes the period after the zero-rate policy was lifted in 2015, while in Gust et al. (2017) and Plante, Richter, and Throckmorton (2018), the observation period ends in 2014:Q1 and 2014:Q2, respectively. Thus, a good model needs

[8]The filtered estimates are computed using the posterior mean estimates of parameters.

to explain not only the continuation of the zero-rate policy from 2009 to 2015, but also the termination of this policy and the subsequent increases in the policy rate from 2016.

To understand why the BR notional rate model empirically performs the best and why its RE counterpart is the worst, we conduct two exercises. First, we compare the sequences of log-likelihood increments $\log p(\mathbf{y}_t \mid \mathbf{Y}_{1:t-1})$ for the two models in panel (a) of Figure A.2 and plot the filtered mean estimates of the nominal and notional interest rates after 2008 in panel (b).[9] Second, we calculate the expected duration of the zero interest rate based on the four models as shown in Figure 3.

Regarding the first exercise, panel (a) of Figure A.2 shows that, for both models, the likelihood increments tend to deteriorate during the zero interest rate period (shaded area), compared with those during the other periods. This outcome implies that it is more difficult to predict the observables during the period, especially when either entering or leaving the zero interest rate is endogenously determined. In particular, the likelihood increments for the RE notional rate model substantially decrease several times during the zero interest rate period, which coincides with the periods when the filtered estimate of the nominal interest rate is larger than zero as shown in panel (b). As we have argued, the notional rate model exhibits a stronger commitment to keeping the zero-rate policy (i.e., forward guidance) than does the nominal rate model, once the notional rate turns to be negative. As the so-called forward guidance puzzle indicates, such a commitment boosts the economy too much under the RE assumption, which is at odds with the data. Thus, to mitigate this commitment effect, the notional RE model is prone to predicting positive interest rates, despite the zero-rate policy, which decreases the likelihood considerably. By contrast, the BR model substantially weakens the power of forward guidance, as we discuss in the next subsection. Thus, the predicted (filtered) notional rate can fall below zero, inducing the zero interest rate as observed in the data. Consequently, the likelihood for the BR model improves.

[9]The filtered mean estimates are the averages of the 120 means of state particles, which are based on 120 parameter particles from the posterior distribution. See Appendix C for the filtered series of output growth, inflation, and the nominal interest rate for the full sample period.

As the second exercise in understanding the difference in model performance, we calculate the duration of ZLB spells as in Gust et al. (2017) and Iiboshi, Shintani, and Ueda (forthcoming). In Figure 3, panel (a) plots the cumulative distribution functions for the average duration of staying at the zero interest rate implied by the BR notional rate model and its RE counterpart, while panel (b) shows the right tails of the histograms for the distributions.[10] According to panel (a), which is based on the RE notional rate model, the median duration of ZLB spells is nearly two quarters, which is almost the same as that in Gust et al. (2017). By contrast, the BR notional rate model can replicate a longer duration, whose median is around four quarters, and thus improve a fit to the data.

4.2 Quantitative Implications of Estimated BR Model

In this subsection, we discuss the quantitative implications of our estimated BR model, following the analyses in Gabaix (2020) regarding the Taylor principle and forward guidance.

4.2.1 Region for Convergence

Regarding the Taylor principle, Gabaix (2020) argues that even when monetary policy is passive, equilibrium can be determinate if the degree of BR is strong enough. He analytically derives the condition for determinacy in a simplified linear model. Here we complement his analysis by investigating the parameter region in which the policy function iteration converges to an equilibrium solution of our nonlinear model with the ZLB. Appendix A.3 shows the determinacy condition for the log-linearized version of our model.

It should be emphasized that convergence does not necessarily imply determinacy (i.e., holding the Taylor principle). Particularly in a nonlinear RE model, Richter and Throckmorton (2015) demonstrate that, when the ZLB constrains the economy, monetary policy needs to be more aggressive than suggested by the Taylor principle in order to obtain a stable

[10] The average duration of ZLB spells are calculated from the simulated series of the nominal interest rate using 120 parameter particles from the posterior distribution. For each draw of parameters, 100 series are simulated with a sample size of 1,000 observations.

solution of the nonlinear model. In this regard, our study contributes to the literature by studying a difference in the parameter regions for convergence between the BR and RE models, whereas Richter and Throckmorton (2015) indicate a difference between the parameter region for convergence and that for holding the Taylor principle based on a RE model.

In panel (a) of Figure 4, the light gray (plus dark gray) area depicts the convergence region for monetary policy parameters, ψ_π and ψ_y, which respectively represent the responses to inflation and detrended output, given the other parameters fixed at its posterior mean estimates of the BR notional model. The dark gray area in the figure shows the same region for the RE case with $M = M^f = 1$. When $M = M^f = 1$, the monetary policy responses are required to be larger than those satisfying the Taylor principle in order to guarantee convergence, as is consistent with the argument in Richter and Throckmorton (2015). With the estimated M and M^f, however, all the positive values for ψ_π and ψ_y lead to convergence.

In the figure, the large filled black circle represents the posterior mean estimates of ψ_π and ψ_y in the BR notional rate model. This circle lies at the boundary of the convergence region for the RE case. This result suggests that, if the model were estimated under the RE assumption, this posterior mean would not be obtained, taking account of the dispersion of the posterior distribution. In other words, the estimation of the RE model would be constrained by a requirement for convergence, which might, in turn, yield an upward bias in the estimates of ψ_π and ψ_y. By contrast, in the estimation of the BR model, the convergence requirement is much less stringent and therefore unlikely to generate such a bias.

To demonstrate which BR parameter, M^f or M, plays a crucial role in expanding the region for convergence, panels (b) and (c) show how the convergence regions expand by changing each of the two BR parameters, while keeping the other fixed at the posterior mean estimate. Panel (b) indicates that, with the estimate of M^f, convergence occurs for almost all the positive values of ψ_π and ψ_y even if $M = 1$. Thus, the estimated M^f explains almost all the expansion of the convergence region shown in panel (a). Moreover, with M fixed at the estimated value, panel (c) demonstrates that the reduction in M^f by small increments

(0.025) from $M^f = 1$ (RE case) to $M^f = 0.925$ remarkably expands the convergence region.

4.2.2 Power of Forward Guidance

Gabaix (2020) argues that the forward guidance policy becomes much less powerful under BR because the intertemporal substitution for households and firms is weakened with $M, M^f < 1$. Here, we quantitatively investigate this argument using our estimated model. Gabaix (2020) simulates his linear model by adding T-period-ahead monetary policy news shocks that are announced at period 0 and materialize as an interest rate cut in period T and quantifies the impact on inflation at time 0. While he considers up to the 80-period-ahead news shock, it is not feasible for our nonlinear model to incorporate news shocks with such a long anticipation horizon because of the increased number of state variables. Thus, in what follows, we add the one-period-ahead monetary policy news shock in period 0. Because this shock will materialize in the next period after the announcement, we offset the shock in period 1 to eliminate the effect of an actual interest cut and to isolate the effect of forward guidance.

Figure 5 depicts the impulse response functions (IRFs) of output, inflation, and the nominal interest rate to the one-period-ahead monetary policy news shock of -2.5% in period 0, followed by the offsetting shock of 2.5% in period 1. IRFs are computed using the following three models: (1) the estimated BR notional rate model (solid line); (2) the estimated BR notional rate model with M and M^f replaced by 1(dotted line); and (3) the estimated RE notional rate model (dashed line). The parameters are fixed at the posterior mean estimates for each model except for the second model in dotted line. Because all the models are nonlinear, the IRFs are state dependent. Thus, we calculate the IRFs conditional on three historical episodes, 1988:Q4, 2010:Q1, and 2015:Q4. The federal funds rate was well above zero in 1988:Q4, whereas 2010:Q1 is the period in which it was the lowest and virtually bounded at zero. In 2015:Q4, the Federal Reserve commenced the policy normalization process by raising the federal funds rate from its effective lower bound.

A comparison of the solid and dotted lines in Figure 5 indicates that the responses of output and inflation are indeed smaller in the estimated BR model than those in the estimated BR model with M and M^f replaced by 1, as is consistent with the argument of Gabaix (2020). Thus, the estimated degree of the BR considerably decreases the effects of forward guidance. While the differences between the two models are remarkable, irrespective of different historical episodes, the effects of forward guidance in both models are somewhat weakened in 2010:Q1 and 2015:Q4, when there was hardly any room for a reduction in the policy rate. This result occurs because the expected future paths of the nominal interest rate are constrained by the ZLB, which impairs the stimulus effects of forward guidance.

The dashed lines in Figure 5 demonstrate that the estimated RE model can generate the weak effects of forward guidance as with the estimated BR model, albeit its poor empirical performance presented in Section 4.1. As shown in Table 2, the interest rate smoothing parameter ρ_r is much smaller in the RE model (0.47) than in the BR model (0.77). This smaller estimate of ρ_r implies that monetary policy is less history dependent and thus mitigates the effects of anticipated future policy (Woodford 2003).

5 Concluding Remarks

This paper has estimated an NK model incorporated with BR and the ZLB in a fully nonlinear and stochastic setting. Our Bayesian estimation results show that BR improves the fit of the model to US data and that the estimated parameters on cognitive discounting indicate a substantial degree of BR. We have found that BR considerably expands the model's parameter region where a solution converges, which enables us to obtain parameter estimates that would not be obtained in the estimation of RE models. Given the estimated parameters, we have demonstrated that the power of forward guidance is weakened by incorporating BR and is further weakened when monetary policy is constrained by the ZLB.

The following two extensions are left for future work. First, whereas we employ a dy-

namically richer model than that analyzed in Gabaix (2020), the model can be extended to a medium-scale model with capital accumulation and sticky wages to assess more empirically relevant implications of BR. Second, although we have focused on the implications of our estimated BR model for monetary policy, it is worth investigating how and to what extent an estimated BR model can alter quantitative implications for fiscal policy. For instance, Gabaix (2020) shows that lump-sum transfers from the government to households, which are ineffective under the RE assumption, can be effective in a BR model. It is meaningful to empirically examine the effect of lump-sum transfers because such a fiscal policy option has been adopted worldwide since the COVID-19 pandemic.

References

[1] Andrade, Joaquim, Pedro Cordeiro, and Guilherme Lambais (2019). "Estimating a Behavioral New Keynesian Model." draft, https://arxiv.org/abs/1912.07601.

[2] Aruoba, S. Borağan, Pablo Cuba-Borda, and Frank Schorfheide (2018). "Macroeconomic Dynamics Near the ZLB: A Tale of Two Countries." *Review of Economic Studies*, 85, 87–118.

[3] Afsar, Atahan, Richard Jaimes, Jose Dago, and Edgar Silgado-Gomez (2020). "Reconciling Empirics and Theory: The Behavioral Hybrid New Keynesian Model." draft.

[4] Chopin, Nicolas, P. E. Jacob, and O. Papaspiliopoulos (2013). "SMC2: An Efficient Algorithm for Sequential Analysis of State Space Models." *Journal of the Royal Statistical Society B*, 75(3), 397–426.

[5] Del Negro, Marco, Marc P. Giannoni, and Christina Patterson (2015). "The Forward Guidance Puzzle." Federal Reserve Bank of New York Staff Reports, 574.

[6] Fuhrer, Jeffrey C., and Glenn D. Rudebusch (2004). "Estimating the Euler Equation for Output." *Journal of Monetary Economics*, 51(6), 1133–1153.

[7] Gabaix, Xavier (2018). "A Behavioral New Keynesian Model." draft, http://pages.stern.nyu.edu/˜xgabaix/papers/brNK.pdf.

[8] Gabaix, Xavier (2020). "A Behavioral New Keynesian Model." *American Economic Review*, 110(8), 2271–2327.

[9] Gali, Jordi, and Mark Gertler (1999). "Inflation Dynamics: A Structural Econometric Analysis." *Journal of Monetary Economics*, 44(2), 195–222.

[10] Ganong, Peter, and Pascal Noel (2019). "Consumer Spending during Unemployment: Positive and Normative Implications." *American Economic Review*, 109 (7), 2383–2424.

[11] Gust, Christopher, Edward Herbst, David López-Salido, and Matthew E. Smith (2017). "The Empirical Implications of the Interest-Rate Lower Bound." *American Economic Review*, 107(7), 1971–2006.

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

[13] Herbst, Edward, and Frank Schorfheide (2015). *Bayesian Estimation of DSGE Models*. Princeton University Press: Princeton, NJ.

[14] Herbst, Edward, and Frank Schorfheide (2019). "Tempered Particle Filtering." *Journal of Econometrics*, 210(1), 26–44.

[15] Iiboshi, Hirokuni, Mototsugu Shintani, and Kozo Ueda (forthcoming). "Estimating a Nonlinear New Keynesian Model with the Zero Lower Bound for Japan." *Journal of Money, Credit and Banking*.

[16] Ilabaca, Francisco, Greta Meggiorini, and Fabio Milani (2020). "Bounded Rationality, Monetary Policy, and Macroeconomic Stability." *Economics Letters*, 186, 108522.

[17] Iskrev, Nikolay (2010). "Local Identification in DSGE Models." *Journal of Monetary Economics*, 57(2), 189–202.

[18] Komunjer, Ivana, and Serena Ng (2011). "Dynamic Identification of Dynamic Stochastic General Equilibrium Models." *Econometrica*, 79(6), 1995–2032.

[19] Linde, Jesper (2005). "Estimating New-Keynesian Phillips Curves: A Full Information Maximum Likelihood Approach." *Journal of Monetary Economics*, 52(6), 1135–1149.

[20] McKay, Alisdair, Emi Nakamura, and Jon Steinsson (2016). "The Power of Forward Guidance Revisited." *American Economic Review*, 106(10), 3133–3158.

[21] Nakata, Taisuke, Ryota Ogaki, Sebastian Schmidt, and Paul Yoo (2019). "Attenuating the Forward Guidance Puzzle: Implications for Optimal Monetary Policy." *Journal of Economic Dynamics and Control*, 105, 90–106.

[22] Plante, Michael, Alexander W. Richter, and Nathaniel A. Throckmorton (2018). "The Zero Lower Bound and Endogenous Uncertainty." *Economic Journal*, 128, 1730–1757.

[23] Qu, Zhongjun, and Denis Tkachenko (2012). "Identification and Frequency Domain Quasi-Maximum Likelihood Estimation of Linearized Dynamic Stochastic General Equilibrium Models." *Quantitative Economics*, 3(1), 95–132.

[24] Richter, Alexander, and Nathaniel Throckmorton (2015). "The Zero Lower Bound: Frequency, Duration, and Numerical Convergence." *B.E. Journal of Macroeconomics*, 15(1), 1–26.

[25] Richter, Alexander W., Nathaniel A. Throckmorton, and Todd B. Walker (2014). "Accuracy, Speed and Robustness of Policy Function Iteration." *Computational Economics*, 44, 445–476.

[26] Smets, Frank, and Raf Wouters (2007). "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." *American Economic Review*, 90, 586–606.

[27] Woodford, Michael (2003). "Optimal Interest-Rate Smoothing." *Review of Economic Studies*, 70(4), 861–886.

Table 1: Prior Distributions

Parameter	Description	Mean	S.D.	Dstribution
M	Households' cognitive discounting factor	0.85	0.05	Beta
M^f	Firms' cognitive discounting factor	0.80	0.05	Beta
h	Consumption habit	0.50	0.20	Beta
$\bar{a}$	Steady-state growth rate	0.42	0.50	Normal
ω	Inverse of labor supply elasticity	2.00	0.75	Normal
κ	Slope of the Phillips curve	0.05	0.006	Normal
$\bar{\pi}$	Target inflation	0.50	0.05	Normal
ρ_r	Interest rate smoothing	0.60	0.10	Beta
ψ_π	Monetary policy response to inflation	2.00	0.50	Normal
ψ_y	Monetary policy response to output	0.25	0.10	Normal
$\psi_{\Delta y}$	Monetary policy response to output growth	0.25	0.10	Normal
ρ_a	Persistence of technology shock	0.60	0.10	Beta
ρ_b	Persistence of discount factor shock	0.60	0.10	Beta
σ_a	Standard deviation of technology shock	0.50	5.00	Inv. Gamma
σ_b	Standard deviation of discount factor shock	0.50	5.00	Inv. Gamma
σ_r	Standard deviation of monetary policy shock	0.20	5.00	Inv. Gamma

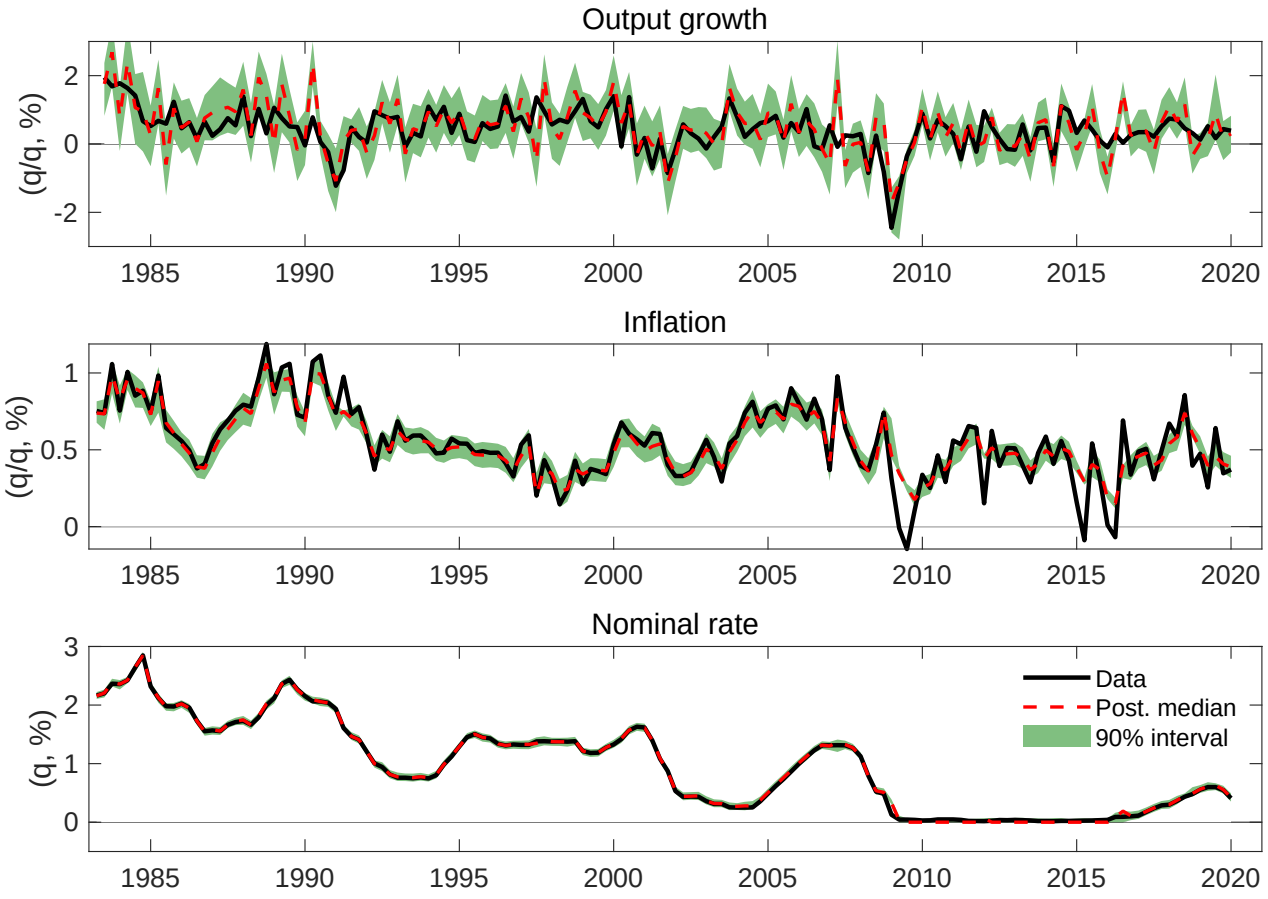
Table 2: Posterior Distributions and Marginal Likelihoods

	BR Notional Rate Model		BR Nominal Rate Model	
Parameter	Mean	90% interval	Mean	90% interval
M	0.856	[0.8097, 0.8861]	0.861	[0.8051, 0.8905]
M^f	0.838	[0.8152, 0.8864]	0.816	[0.799, 0.8352]
h	0.364	[0.3275, 0.436]	0.484	[0.3879, 0.5825]
$\bar{a}$	0.418	[0.3496, 0.4441]	0.436	[0.4271, 0.4701]
ω	2.245	[1.6866, 3.29]	1.964	[1.6411, 3.2491]
κ	0.046	[0.0411, 0.0536]	0.062	[0.05, 0.0667]
$\bar{\pi}$	0.492	[0.4738, 0.519]	0.437	[0.405, 0.5938]
ρ_r	0.774	[0.7434, 0.7895]	0.716	[0.6881, 0.7688]
ψ_π	2.016	[1.9383, 2.1569]	1.656	[0.9883, 1.959]
ψ_y	0.232	[0.1382, 0.2773]	0.241	[0.1647, 0.3585]
$\psi_{\Delta y}$	0.252	[0.1908, 0.4163]	0.150	[0.0962, 0.2619]
ρ_a	0.443	[0.3891, 0.6471]	0.619	[0.547, 0.6641]
ρ_b	0.688	[0.6546, 0.7775]	0.683	[0.6059, 0.8197]
σ_a	0.415	[0.3877, 0.4821]	0.467	[0.4167, 0.5331]
σ_b	0.800	[0.521, 0.8995]	0.776	[0.6271, 0.8562]
σ_r	0.204	[0.1333, 0.2835]	0.311	[0.1898, 0.3944]
$\log p(\mathbf{Y})$	-93.97		-105.06	

	RE Notional Rate Model		RE Nominal Rate Model	
Parameter	Mean	90% interval	Mean	90% interval
h	0.625	[0.5082, 0.6878]	0.509	[0.3339, 0.5555]
$\bar{a}$	0.387	[0.3746, 0.4089]	0.412	[0.4083, 0.4136]
ω	2.075	[1.602, 2.9562]	2.263	[2.0028, 2.3355]
κ	0.031	[0.0255, 0.0407]	0.045	[0.03, 0.0467]
$\bar{\pi}$	0.441	[0.411, 0.4953]	0.567	[0.5222, 0.5759]
ρ_r	0.472	[0.4046, 0.5962]	0.643	[0.5921, 0.8164]
ψ_π	2.052	[1.9636, 2.2152]	2.493	[2.2071, 2.8187]
ψ_y	0.165	[0.1027, 0.1992]	0.207	[0.1685, 0.2431]
$\psi_{\Delta y}$	0.352	[0.2968, 0.3816]	0.239	[0.1755, 0.2583]
ρ_a	0.561	[0.5591, 0.5621]	0.727	[0.5885, 0.7659]
ρ_b	0.611	[0.5446, 0.7328]	0.607	[0.5801, 0.7373]
σ_a	0.511	[0.4066, 0.7047]	0.388	[0.3197, 0.4261]
σ_b	0.728	[0.5028, 0.8535]	0.633	[0.6024, 0.7276]
σ_r	0.264	[0.1846, 0.307]	0.226	[0.203, 0.2545]
$\log p(\mathbf{Y})$	-277.23		-175.53	

Notes: This table reports the posterior mean and 90 percent credible intervals of estimated parameters in each model, based on $1,200$ particles from the final importance sampling in the SMC^2 algorithm. In the table, $\log p(\mathbf{Y})$ represents the SMC-based approximation of the log marginal likelihood.

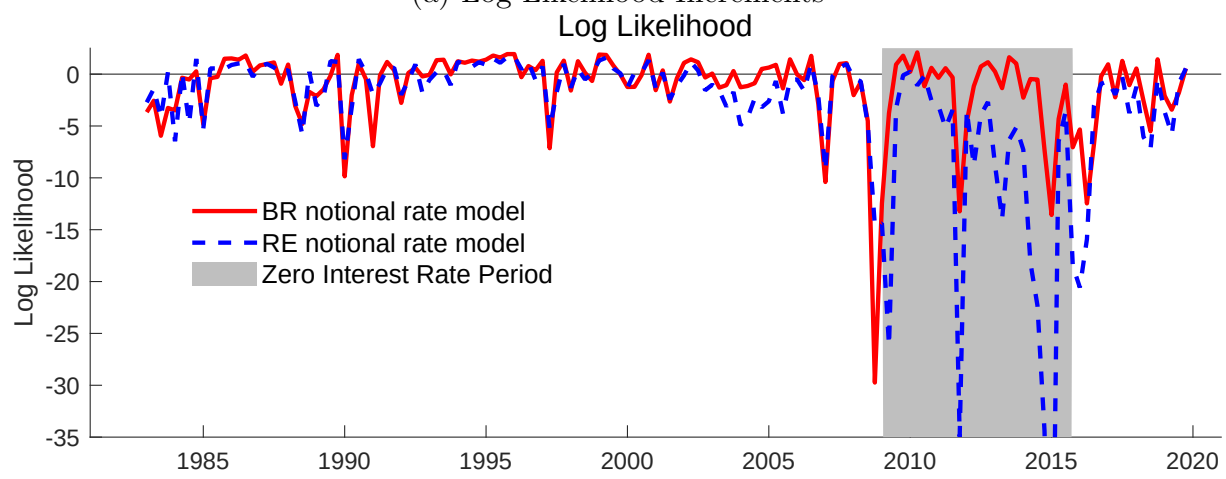
Figure 1: Filtered Series of Observables



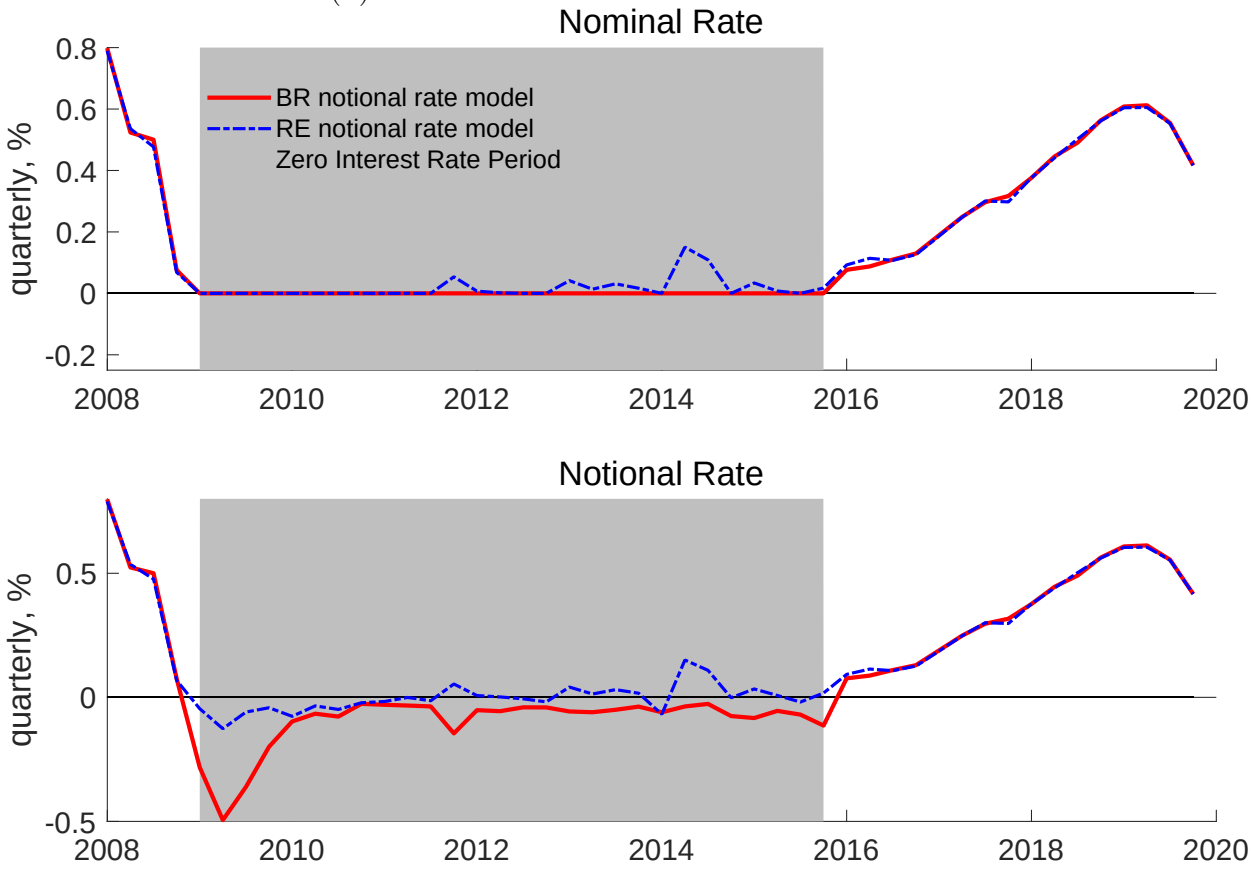
Notes: This figure compares the filtered median estimates (dashed lines) of output growth, inflation, and the nominal interest rate, based on the estimated BR notional rate model with the corresponding data (solid lines). The shaded areas show the 90% credible intervals of the filtered estimates.

Figure 2: Series of Log-Likelihood Increments and Filtered Nominal/Notional Rates

(a) Log-Likelihood Increments

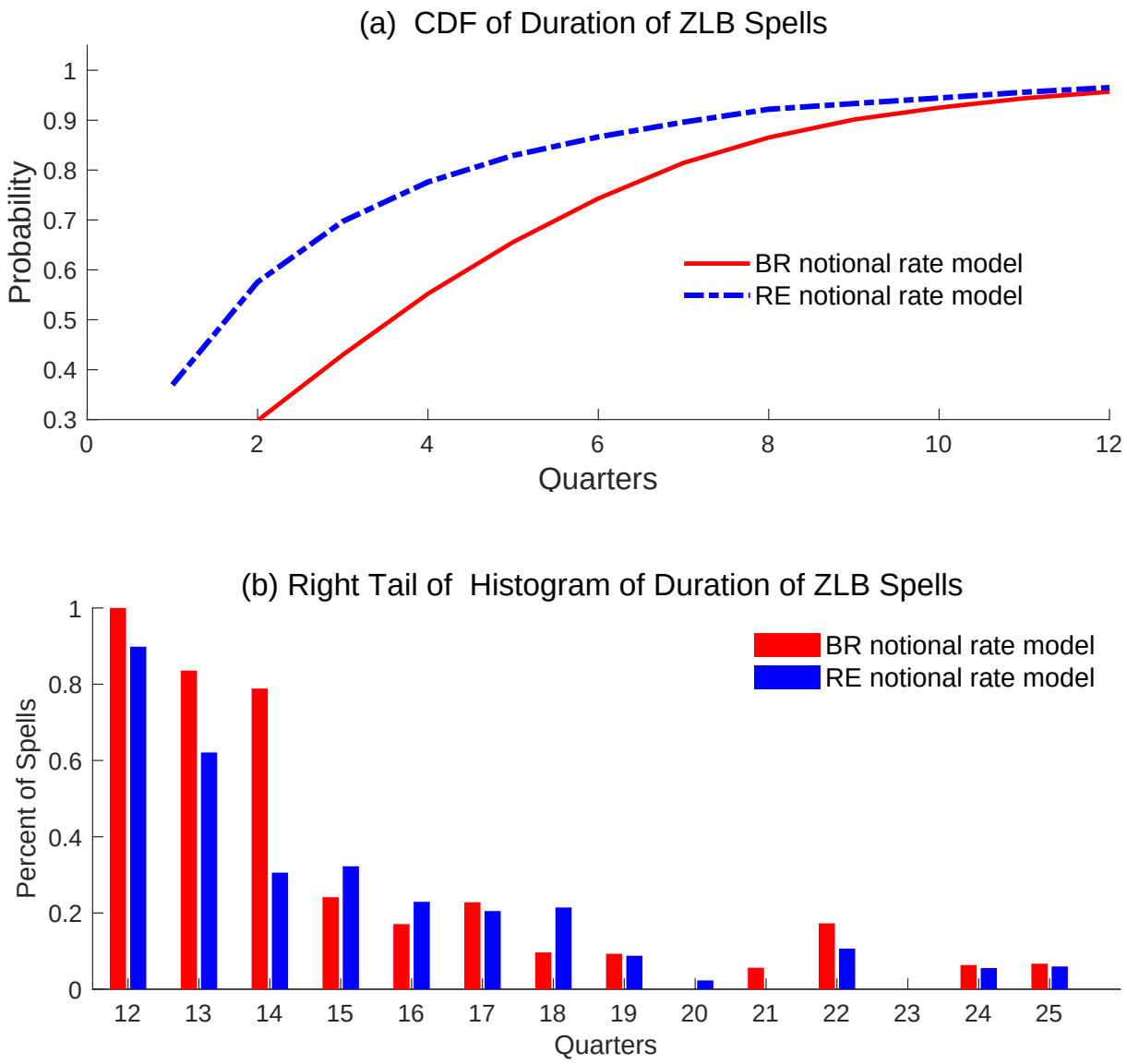


(b) Nominal and Notional Rates After 2008



Notes: In this figure, panel (a) compares the sequence of log-likelihood increments $\log p(\mathbf{y}_t \mid \mathbf{Y}_{1:t-1})$ for the BR notional rate model with that for the RE notional rate model. Panel (b) plots the filtered mean estimates of the nominal and notional interest rates after 2008, based on the two models.

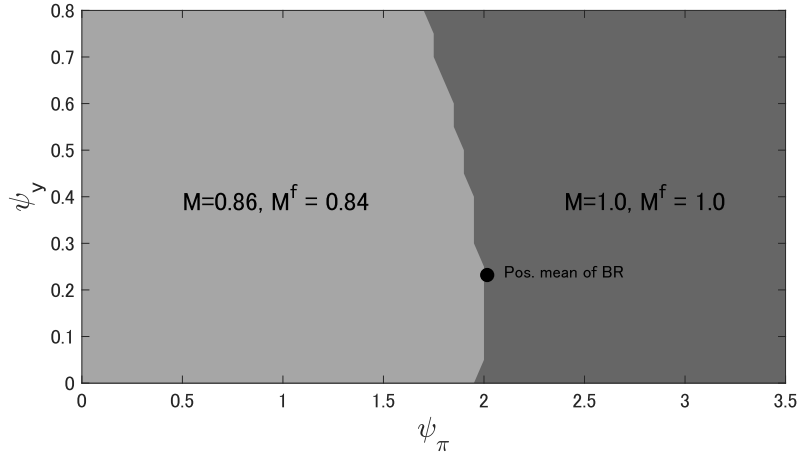
Figure 3: Duration of Being at the Zero Lower Bound



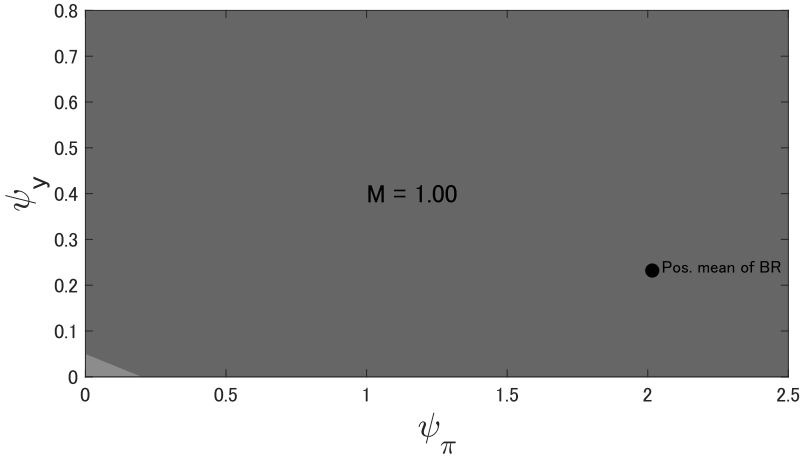
Notes: In this figure, panel (a) plots the cumulative distribution functions for the average duration of the ZLB spells implied by the BR notional rate model and its RE counterpart. Panel (b) shows the right tails of the histograms for the distributions.

Figure 4: Regions for Convergence

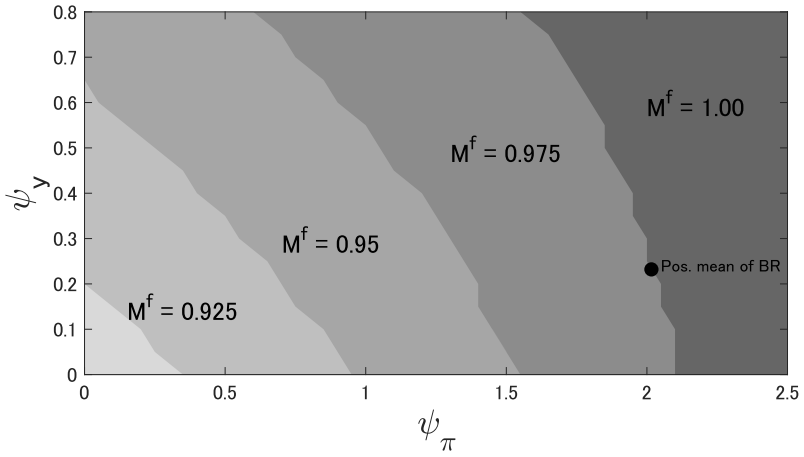
(a) Effect of BR



(b) Effect of M, given $M^f = 0.84$

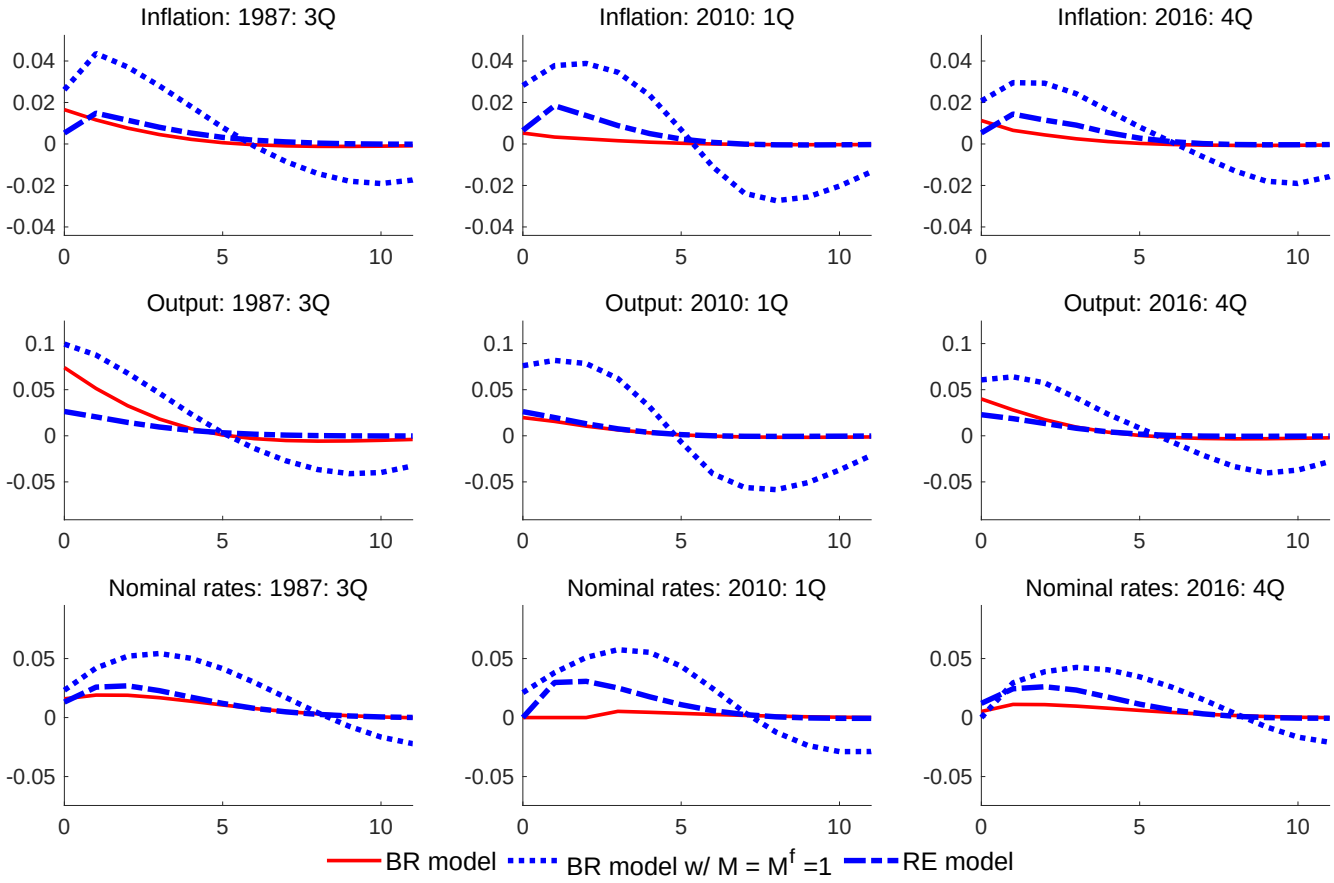


(c) Effect of M^f, given $M = 0.86$



Notes: This figure shows the convergence regions in (ψ_π, ψ_y)-space for different values of M and M^f, given the posterior mean estimates of the other parameters in the BR notional rate model.

Figure 5: Power of Foward Guidance: Impulse Responses to a Monetary Policy News Shock



Note: This figure shows the impulse responses of output, inflation, and the nominal interest rate to the one-period-ahead monetary policy news shock of -2.5% in period 0, followed by the offsetting shock of 2.5% in period 1, based on the BR notional rate model (solid line), the same model but with $M = M^f = 1$ (dotted line), and the RE notional rate model (dashed line), given the posterior mean estimates of parameters for each model.

Appendix to "Estimating a Behavioral New Keynesian Model with the Zero Lower Bound"

Yasuo Hirose Hirokuni Iiboshi Mototsugu Shintani Kozo Ueda

March 2022

A Details of the Model

A.1 Detrended Equilibrium Conditions

Because the technology level A_t is nonstationary, we rewrite the equilibrium conditions in terms of stationary variables detrended by A_t: $y_t = Y_t/A_t$, $c_t = C_t/A_t$, $\lambda_t = \Lambda_t A_t^{\sigma}$.

$$\lambda_t = \left[c_t - h c_{t-1} \left(\frac{1}{\gamma_a e^{\mu_t^a}} \right) \right]^{-\sigma}, \tag{17}$$

$$1 - \frac{\lambda}{\lambda_t} \frac{1}{Z_t^b} = M \mathbb{E}_t \left[\beta \frac{\lambda_{t+1}}{\lambda_t} \left(\gamma_a e^{\mu_{t+1}^a} \right)^{-\sigma} \frac{Z_{t+1}^b}{Z_t^b} \frac{R_t}{\pi_{t+1}} - \frac{\lambda}{\lambda_t} \frac{1}{Z_t^b} \right], \tag{18}$$

$$\begin{aligned} &1 - \phi \left(\pi_t - \pi \right) \pi_t - \varepsilon \left(1 - \chi \frac{(y_t)^{\omega}}{\lambda_t} - \frac{\phi}{2} \left(\pi_t - \pi \right)^2 \right) \\ &+ \beta M^f \mathbb{E}_t \left[\frac{\lambda_{t+1}}{\lambda_t} \left(\gamma_a e^{\mu_{t+1}^a} \right)^{1-\sigma} \frac{Z_{t+1}^b}{Z_t^b} \phi \left(\pi_{t+1} - \pi \right) \pi_{t+1} \frac{y_{t+1}}{y_t} \right] = 0, \end{aligned} \tag{19}$$

$$y_t = c_t + \phi \left(\pi_t - \pi \right)^2 y_t / 2, \tag{20}$$

$$R_t = \max[R_t^*, 1], \tag{21}$$

where, in the nominal rate model,

$$R_t^* = (R_{t-1})^{\rho_r} \left(R \left(\frac{\pi_t}{\pi} \right)^{\psi_\pi} \left(\frac{y_t}{y} \right)^{\psi_y} \left(\frac{y_t}{y_{t-1}} e^{\mu_{t+1}^a} \right)^{\psi_{\Delta y}} \right)^{1-\rho_r} e^{\epsilon_t^r}, \tag{22}$$

or, in the notional rate model,

$$R_t^* = \left(R_{t-1}^*\right)^{\rho_r} \left(R \left(\frac{\pi_t}{\pi}\right)^{\psi_\pi} \left(\frac{y_t}{y}\right)^{\psi_y} \left(\frac{y_t}{y_{t-1}} e^{\mu_{t+1}^a}\right)^{\psi_{\Delta y}} \right)^{1-\rho_r} e^{\epsilon_t^r}. \tag{23}$$

An equilibrium is given by the sequences $\{y_t, c_t, \lambda_t, \pi_t, R_t, R_t^*\}$ satisfying the detrended equilibrium conditions.

A.2 Steady State

At the steady state, given π, we have

$$y = \left[\frac{\varepsilon - 1}{\varepsilon \chi} \left(1 - \frac{h}{\gamma_a}\right)^{-\sigma} \right]^{\frac{1}{\sigma+\omega}}, \tag{24}$$

$$c = y, \tag{25}$$

$$\lambda = \frac{\varepsilon}{\varepsilon - 1} \chi y^{\omega}, \tag{26}$$

$$R = \frac{\gamma_a^{\sigma} \pi}{\beta} (\geq 1), \tag{27}$$

$$R^* = R. \tag{28}$$

A.3 Linearized System of Equations and Condition for Equilibrium Determinacy

Log-linearizing the detrended equilibrium conditions, rearranging the resulting equations, and omitting the structural shocks yields the following three-equation system:

$$\tilde{y}_t = \frac{h}{\gamma_a + Mh} \tilde{y}_{t-1} + \frac{M\gamma_a}{\gamma_a + Mh} \mathbb{E}_t \tilde{y}_{t+1} - \frac{\gamma_a - h}{(\gamma_a + Mh)\,\sigma} \left(\tilde{R}_t - \mathbb{E}_t \pi_{t+1} \right), \tag{29}$$

$$\tilde{\pi}_t = \beta M^f \gamma_a^{1-\sigma} \mathbb{E}_t \tilde{\pi}_{t+1} + \frac{\varepsilon - 1}{\phi \pi} \left(\omega + \frac{\sigma}{1 - h/\gamma_a} \right) \tilde{y}_t, \tag{30}$$

$$\tilde{R}_t = \rho_r \tilde{R}_{t-1} + (1-\rho_r)\left[\psi_\pi \tilde{\pi}_t + \psi_y \tilde{y}_t + \psi_{\Delta y}(\tilde{y}_t - \tilde{y}_{t-1})\right]. \tag{31}$$

To derive a condition for equilibrium determinacy analytically, we consider the case where $h = 0$ and $\psi_{\Delta y} = 0$.

Then, the system can be written as the following matrix form:

$$A \begin{pmatrix} \tilde{R}_t \\ \mathbb{E}_t \tilde{\pi}_{t+1} \\ \mathbb{E}_t \tilde{y}_{t+1} \end{pmatrix} = B \begin{pmatrix} \tilde{R}_{t-1} \\ \tilde{\pi}_t \\ \tilde{y}_t \end{pmatrix}, \tag{32}$$

where

$$A \equiv \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\sigma^{-1} & \sigma^{-1} & 1 \end{pmatrix},$$

$$B \equiv \begin{pmatrix} 0 & \psi_\pi & \psi_y \\ 0 & (\beta M^f \gamma_a^{1-\sigma})^{-1} & -\kappa \\ 0 & 0 & M^{-1} \end{pmatrix},$$

with $\kappa = \left[\frac{\varepsilon-1}{\phi\pi}(\omega+\sigma)\right]^{-1}$

The equilibrium is determinate if and only if the number of the eigenvalues of the matrix $A^{-1}B$ that are outside the unit circle is two. Since

$$A^{-1}B = \begin{pmatrix} 0 & \psi_\pi & \psi_y \\ 0 & (\beta M^f \gamma_a^{1-\sigma})^{-1} & -\kappa \\ 0 & \sigma^{-1}\psi_\pi - \sigma^{-1}(\beta M^f \gamma_a^{1-\sigma})^{-1} & \sigma^{-1}\psi_y + \sigma^{-1}\kappa + M^{-1} \end{pmatrix},$$

the characteristic equation is

$$f(\lambda) = \left\{(\beta M^f \gamma_a^{1-\sigma})^{-1} - \lambda\right\}\left\{\sigma^{-1}\psi_y + \sigma^{-1}\kappa + M^{-1} - \lambda\right\} + \left\{\sigma^{-1}\psi_\pi - \sigma^{-1}(\beta M^f \gamma_a^{1-\sigma})^{-1}\right\} \cdot \kappa$$
$$= \lambda^2 - \left\{(\beta M^f \gamma_a^{1-\sigma})^{-1} + \sigma^{-1}\psi_y + \sigma^{-1}\kappa + M^{-1}\right\}\lambda + \left\{\sigma^{-1}\psi_y + M^{-1}\right\}(\beta M^f \gamma_a^{1-\sigma})^{-1} + \sigma^{-1}\psi_\pi\kappa.$$

The two eigenvalues of the matrix are outside the unit circle if $f(1) > 0$; that is,

$$f(1) = 1 - (\beta M^f \gamma_a^{1-\sigma})^{-1} - \sigma^{-1}\kappa - M^{-1} + (\beta M M^f \gamma_a^{1-\sigma})^{-1} + \sigma^{-1}\psi_\pi\kappa + \sigma^{-1}\psi_y\left\{(\beta M^f \gamma_a^{1-\sigma})^{-1} - 1\right\}$$
$$> 0.$$

If $\gamma_a = 1$, this becomes

$$f(1) = 1 - (\beta M^f)^{-1} - \sigma^{-1}\kappa - M^{-1} + (\beta M M^f)^{-1} + \sigma^{-1}\psi_\pi\kappa + \sigma^{-1}\psi_y\left\{(\beta M^f)^{-1} - 1\right\}$$
$$\Leftrightarrow (\sigma/\kappa)f(1) = \psi_\pi - 1 + \psi_y\frac{1-\beta M^f}{\kappa\beta M^f} + \frac{\sigma}{\kappa}\left\{1 - (\beta M^f)^{-1} - M^{-1} + (\beta M M^f)^{-1}\right\} > 0.$$

Thus, the condition for equilibrium determinacy is given by

$$\psi_\pi - 1 + \psi_y\frac{1-\beta M^f}{\kappa\beta M^f} + \frac{\sigma}{\kappa}\frac{1}{\beta M M^f}(1-M)\left(1-\beta M^f\right) > 0. \tag{33}$$

This condition indicates that the determinacy region of the model's parameter space expands as either M or M^f decreases.

B Methodology

B.1 Model Solution

The detrended equilibrium conditions contain two endogenous state variables (output y_{t-1} and the notional nominal interest rate R^*_{t-1}) and three exogenous shocks (the technology shock μ^a_t, the discount factor shock Z^b_t, and the monetary policy shock ϵ^r_t). We solve for

policy functions that depend on these state variables. More specifically, we compute the time-invariant policy functions for output y_t and inflation π_t:

$$(y_t, \pi_t)' = \Phi(z_{t-1}, \epsilon_t),$$

where $z_{t-1} = (y_{t-1}, R^*_{t-1})'$ and $\epsilon_t = (\mu^a_t, Z^b_t, \epsilon^r_t)'$. The TL method locally approximates the policy functions at each node in the state space (z_{t-1}, ϵ_t). We discretize five grid points on each of the continuous state variables; i.e., $3,125 (= 5^5)$ nodes in total.

The policy function iteration in the TL method takes the following steps. Let $i \in \{0, \cdots, I\}$ denote the iterations of the algorithm and $n \in \{1, \cdots, N\}$ denote the nodes in the policy function $\Phi(z_{t-1}, \epsilon_t)$.

1. For $i = 0$, we make an initial conjecture of the policy function $\Phi^0(z_{t-1}, \epsilon_t)$ from a solution for a linearized version of the model without the ZLB using Sims' (2002) `gensys` algorithm.

2. For each iteration $i \in \{1, \cdots, I\}$ and node $n \in \{1, \cdots, N\}$, we execute the following procedures.

   (a) Given y^{i-1}_t and π^{i-1}_t obtained by $\Phi^{i-1}(z_{t-1}, \epsilon_t)$, solve for endogenous variables $\{c_t, R_t, R^*_t\}'$ from the equilibrium conditions including the ZLB.

   (b) Calculate the future variables $\{\mathbb{E}_t y_{t+1}, \mathbb{E}_t \pi_{t+1}\}$ by applying a piecewise linear interpolation to the policy function $\Phi^{i-1}(z_t, \epsilon_{t+1})$. Then, substitute the future variables into the equilibrium conditions. We employ the Gauss-Hermite integration to approximate the conditional expectations with three nodes per shock, following Gust et al. (2017).

   (c) Use a nonlinear solver such as Sims' `csolve` to find the policy function $\Phi^i(z_{t-1}, \epsilon_t)$ that minimizes the errors in the model's intertemporal equations.

3. Define $\texttt{maxdist} = \texttt{max}(\,|y^i_{t,n} - y^{i-1}_{t,n}|,\, |\pi^i_{t,n} - \pi^{i-1}_{t,n}|)$. Repeat Step 2 until the policy function converges, say, to $\texttt{maxdist} < 10^{-4}$, for all nodes, n.

We classify the solution as convergent when the following conditions are met: (i) the policy function converges before iteration i reaches $I = 1,000$; (ii) the ZLB constrains fewer than 50 percent of the nodes in the state space, following Richter and Throckmorton (2015); and (iii) the maximum of the differences in policy functions between iterations $i-1$ and i is less than 1.5 times the maximum of the differences in policy functions between iterations $i-2$ and $i-1$ (contraction mapping). Otherwise, we classify the solution as non-convergent.

B.2 Estimation Procedures

To generate draws from the posterior distribution of parameters θ in our nonlinear model, we use the SMC^2 sampler instead of the widely-used particle filter Metropolis-Hastings (PFMH) algorithm. While the PFMH algorithm requires serialization in generating draws and hence is not amenable to parallel computing, the SMC^2 can make use of parallelization to shorten the computation time dramatically. Moreover, SMC^2 can approximate the posterior distribution more accurately by working independent sampling efficiently. In approximating the likelihood, we incorporate the tempered particle filter (TFP) proposed by Herbst and Schorfheid (2019) into the SMC^2, instead of the so-called bootstrap particle filter (BSPF) employed by Gust et al. (2017).

SMC^2 consists of two parts: parameter sampling and likelihood evaluation. In what follows, we first explain the algorithm for SMC-based parameter sampling, following Herbst and Schorfheide (2015) and Fernández-Villaverde, Rubio-Ramirez, and Schorfheide (2016). Then we turn to the TFP algorithm for likelihood evaluation, following Herbst and Schorfheid (2019). Notice that both algorithms share a similarity in the use of tempering stages and schedule to improve approximation accuracy.

B.2.1 Algorithm for Parameter Sampling

Let ϕ_n for $n = 0, \cdots, N_\phi$ denote a sequence that slowly increases from zero to one. We define a sequence of bridge distributions, $\{\pi_n(\theta)\}_{n=0}^{N_\phi}$ as

$$\pi_n(\theta) = \frac{[p(\mathbf{Y}|\theta)]^{\phi_n} p(\theta)}{\int [p(\mathbf{Y}|\theta)]^{\phi_n} p(\theta) d\theta}, \quad \text{for } n = 0, \cdots, N_\phi, \ \phi_n \uparrow 1,$$

where $p(\theta)$ and $p(\mathbf{Y}|\theta)$ are the prior density and likelihood function, respectively. We adopt the likelihood tempering approach that generates the bridge distributions, $\{\pi_n(\theta)\}_{n=0}^{N_\phi}$, by taking the power transformation of $p(\mathbf{Y}|\theta)$ with parameter ϕ_n (i.e., $[p(\mathbf{Y}|\theta)]^{\phi_n}$). The bride distribution converges to the target posterior distribution as $\phi_n \to 1$. The tempering schedule $\{\phi_n\}_{n=0}^{N_\phi}$ is determined by $\phi_n = (n/N_\phi)^\lambda$, where λ is a parameter that controls the shape of the tempering schedule. Following Herbst and Schorfheide (2015), we set $\lambda = 2$.

Let θ_n^i for $i = 0, \cdots, N_\theta$ denote the particles of a parameter vector in stage n. In the SMC^2, the posterior draws of parameters are generated from the following steps.

1. **Initialization Step**:

   (a) Set the initial stage as $n = 0$ and draw the initial particle of parameters θ_0^i from the prior distribution $p(\theta)$.

   (b) Set a weight for each particle in the initial stage as $W_0^i = 1$ for $i = 1, \cdots, N_\theta$.

   Then, for $n = 1, \cdots, N_\phi$, repeat Steps 2 to 4.

2. **Correction Step**: Calculate the normalized weight, $\tilde{W}_n^i$, for each particle as

$$\tilde{W}_n^i = \frac{\tilde{w}_n^i W_{n-1}^i}{\frac{1}{N}\sum_{i=1}^N \tilde{w}_n^i W_{n-1}^i} \text{ for } i = 1, \cdots, N_\theta,$$

   where $\tilde{w}_t^i$ is an incremental weight defined as

$$\tilde{w}_n^i = [p(\mathbf{Y}|\theta_{n-1}^i)]^{\phi_n - \phi_{n-1}}.$$

The likelihood $\hat{p}(\mathbf{Y}|\theta)$ is approximated by the particle filter explained in the next subsection.

The correction step is a classic importance sampling step in which particle weights are updated to reflect the stage n distribution, $\pi_n(\theta)$. Because this step does not change the particle value, we can skip this step only by calculating the power transformation of $p(\mathbf{Y}|\theta)$ with parameter ϕ_n.

3. **Selection (Resampling) Step**:

   (a) Calculate the effective particle sample size $\widehat{ESS_n}$, which is defined as

   $$\widehat{ESS_n} = N_\theta / \left(\frac{1}{N_\theta} \sum_{i=1}^{N_\theta} (\tilde{W}_n^i)^2 \right).$$

   (b) If $\widehat{ESS_n} < N_\theta/2$, then resample particles $\{\hat{\theta}_n^i\}$ by a multinomial resampling procedure and set $W_n^i = 1$.

   (c) Otherwise, let $\hat{\theta}_n^i = \theta_{n-1}^i$ and $W_n^i = \tilde{W}_n^i$.

4. **Mutation Step**: Propagate the particles $\{\hat{\theta}_n^i,\, W_n^i\}$ via the random walk MH algorithm with the proposal distribution

   $$\vartheta|\hat{\theta}_n^i \sim N\left(\hat{\theta}_n^i,\, c_n^2 \Sigma_{\hat{\theta}_n}\right),$$

   where $\Sigma_{\hat{\theta}_n}$ denotes the covariance matrix of the particles of parameters $\{\hat{\theta}_n^i\}$ in stage n and c_n is a scaling factor. To keep the acceptance rate around 25%, we set the scaling factor c_n for $n > 2$ as

   $$c_n = c_{n-1} f(A_{n-1}),$$

   where A_{n-1} represents the acceptance rate in the mutation step in stage $n-1$ and

function $f(A_{n-1})$ is given by

$$f(A_{n-1}) = 0.95 + 0.10\frac{e^{16(A_{n-1}-0.25)}}{1+e^{16(A_{n-1}-0.25)}}.$$

5. In the final stage of $n = N_\phi$, the final importance sampling approximation of the marginal likelihood is obtained as

$$P_{SMC}(Y) = \prod_{n=1}^{N_\phi}\left(\frac{1}{N_\theta}\sum_{i=1}^{N_\theta}\tilde{w}_n^i W_{n-1}^i\right),$$

which converges almost surely to $p(Y)$ as $N_\theta \to \infty$.

Similarly, we can calculate the particle approximation of the posterior estimator, $E_\pi[h(\theta)]$, as

$$h_{N_\phi, N_\theta} = \sum_{i=1}^{N_\theta} h(\theta_{N_\phi}^i) W_{N_\phi}^i,$$

where $h(\cdot)$ is a function of any interest such as impulse response functions.

B.2.2 Algorithm of the Tempered Particle Filter (TPF)

Given the parameter vector θ, the solution for our nonlinear model and the observation equations can be written as the following state-space form:

$$s_t = \Phi(s_{t-1},\ \epsilon_t,\ \theta), \quad \epsilon_t \sim N(0,\ \Sigma_\epsilon),$$
$$\mathbf{y}_t = \Psi(s_t,\ \theta) + u_t, \quad u_t \sim N(0,\ \Sigma_u),$$

where $s_t = \left(y_t, c_t, \pi_t, R_t, R_t^*, \mu_t^a, Z_t^b\right)'$ denotes a vector of endogenous variables, $\epsilon_t = \left(\epsilon_t^a,\ \epsilon_t^b,\ \epsilon_t^r\right)'$ an exogenous shock vector, $\mathbf{y}_t = (100\Delta \log GDP_t, 100\Delta \log PGDP_t, FF_t)'$ is a vector of observables, and $u_t = (u_t^y, u_t^\pi, u_t^R)'$ is a vector of measurement errors. ϵ_t and u_t follow a multivariate normal distributions with covariance matrices Σ_u and Σ_ϵ, respectively.

The structure of the TPF algorithm is basically the same as the generic BSPF algorithm.

The only difference is the inclusion of the tempering iteration step outlined in Step 2 (b) below, in which the measurement error variances are reduced iteratively. In this step, a tempering schedule $\phi_{n,t}$ and the number of stages N_t^{ϕ} are introduced as similar to the parameter sampling in the previous subsection, but they are adaptively determined in the algorithm to improve both accuracy and efficiency of approximation.

Let s_t^j and ϵ_t^j for $j = 0, \cdots, N_S$ respectively denote the particles of endogenous variables and exogenous shocks in period t.

1. For period $t = 0$, draw the particles of exogenous shocks ϵ_0^j from $N(0,\, \Sigma_\epsilon)$ and generate the initial particles of endogenous variables $s_{0|0}^j$ from the nonlinear solution $\Phi(\bar{s},\ \epsilon_0^j,\ \theta)$, where $\bar{s}$ is the ergodic mean of endogenous variables.

2. For $t = 1, \cdots, T$, repeat Steps (a) and (b).

   (a) **Particle Initialization Step**:

      i. Draw the particles of exogenous shocks ϵ_t^j from $N(0,\, \Sigma_\epsilon)$ and generate the forecasts of endogenous variables $s_{t|t-1}^j$ from the nonlinear solution $\Phi(s_{t-1|t-1},\ \epsilon_t^j,\ \theta)$.

      ii. Approximate the predictive density of observables $\mathbf{y}_t$ by

$$p(\mathbf{y}_t|\mathbf{Y}_{1:t-1},\, \theta) = \frac{1}{N_S}\sum_{j=1}^{N_S} w_t^j,$$

      where w_t^j is the normal predictive density of particle j based on the measurement equations $\mathbf{y}_t = \Psi(s_t, \theta) + u_t$, $u_t \sim N(0,\, \Sigma_u)$. This density is given by

$$w_t^j = (2\pi)^{-N_\mathbf{y}/2}|\Sigma_u|^{-1/2}\exp\left\{ -\frac{1}{2}\left(\mathbf{y}_t - \Psi(s_{t|t-1}^j, \theta)\right)'\Sigma_u^{-1}\left(\mathbf{y_t} - \Psi(s_{t|t-1}^j, \theta)\right)\right\},$$

      where $N_\mathbf{y}$ is the dimension of $\mathbf{y}_t$; i.e., the number of observables.

      iii. Resample the particles $\{s_{t|t-1}^j\}$ from a multinomial distribution with proba-

bility $w_t^j / \Sigma w_t^j$ and regard the resulting particles as $\{s_{t|t}^j\}$.

(b) **Tempering Iteration Step**:

Conduct the algorithm below and approximate the likelihood increment:

$$p(\mathbf{y}_t \,|\, \mathbf{Y}_{1:t-1}) = \prod_{n=1}^{N_t^\phi} \left(\frac{1}{N_s} \sum_{i=1}^{N_S} w_t^{i,n} W_t^{i,n-1} \right).$$

3. The particle approximation of the likelihood is obtained as

$$p(\mathbf{Y}|\theta) = \prod_{t=1}^{T} p(\mathbf{y}_t \,|\, \mathbf{Y}_{1:t-1}).$$

Algorithm for Tempering Iteration Step

Until $\phi_{n,t} = 1$, repeat Steps 1 to 3.

1. **Correction Step**:

(a) Calculate the normalized weight $\tilde{W}_t^{j,n}$ for each particle as

$$\tilde{W}_t^{j,n}(\phi_{n,t}) = \frac{w_t^{j,n} W_t^{j,n-1}}{\frac{1}{N_S} \sum_{i=1}^{N_S} w_t^{j,n} W_t^{j,n-1}} \text{ for } j = 1, \cdots, N_S,$$

where $w_t^{j,n}$ is an incremental weight derived from likelihood tempering with $\phi_{n,t} - \phi_{n-1,t}$:

$$w_t^{j,n} = [p(\mathbf{y_t}|\mathbf{Y}_{1:t}, \theta)]^{\phi_{n,t} - \phi_{n-1,t}}.$$

(b) Calculate the inefficiency ratio $\text{InEff}(\phi_n)$ from

$$\text{InEff}(\phi_n) = \frac{1}{N_S} \sum_{j=1}^{N_S} \left(\tilde{W}_t^{j,n}(\phi_{n.t}) \right)^2.$$

(c) Let r^* denotes target inefficiency ratio, say $r^* = 3$. If $\text{InEff}(\phi_n) > r^*$, then set $\phi_{n,t}$ to the value that solves $\text{InEff}(\phi_n) = r^*$ and set $\tilde{W}_t^{j,n} = \tilde{W}_t^{j,n}(\phi_{n.t})$. Otherwise,

set $\phi_{n,t} = 1$, $N_t^{\phi} = n$, and $\tilde{W}_t^{j,n} = \tilde{W}_t^{j,n}(1)$.

2. **Selection Step**:

   Resample the particles $\{s_t^{i,n-1}, \epsilon_t^{j,n-1}, s_{t-1}^{j,N_{t-1}^{\phi}}, \tilde{W}_t^{j,n-1}\}$ from a multinomial distribution with probability $\tilde{W}_t^{j,n-1}$ and regard the resulting particles as $\{\hat{s}_t^{i,n}, \hat{\epsilon}_t^{j,n}, s_{t-1}^{j,N_{t-1}^{\phi}}, W_t^{j,n}\}$. Then, set $W_t^{i,n} = 1$ for $j = 1, \cdots, N_S$.

3. **Mutation Step**. Propagate the particles $\{\hat{s}_t^{i,n}, \hat{\epsilon}_t^{j,n}, s_{t-1}^{j,N_{t-1}^{\phi}}, W_t^{j,n}\}$ to obtain $\{s_t^{i,n}, \epsilon_t^{j,n}, s_{t-1}^{j,N_{t-1}^{\phi}}, W_t^{j,n}\}$ via the random walk MH algorithm with the proposal distribution:

$$e_t^j \sim N\left(\hat{\epsilon}_t^{j,n,l-1},\, c_n^2 I_n\right),$$

   where l is the iteration number in the MH algorithm and the number of iteration is set at two in our estimation. c_n is a scaling factor for the identity matrix I_n in stage n. To keep the acceptance rate around 40%, we set the scaling factor c_n for $n > 2$ as

$$c_n = c_{n-1} f(A_{n-1}),$$

   where A_{n-1} represents the acceptance rate in the mutation step in stage $n-1$ and function $f(A_{n-1})$ is given by

$$f(A_{n-1}) = 0.95 + 0.10 \frac{e^{20(A_{n-1}-0.4)}}{1 + e^{20(A_{n-1}-0.4)}}.$$

B.3 Comparison of the TPF and the BSPF

Herbst and Schorfheide (2019) conduct a Monte Carlo experiment to compare the performance of the TPF and the BSPF in estimating a linear DSGE model. We follow their experiments but use our nonlinear model with the ZLB to show the advantages of the TPF.

Figure A.1 shows the distributions of the approximated likelihood based on three different settings of the particle filters. For each setting, the particle approximation of the likelihood

is computed for 100 times, given the same parameters and data. While the solid blue line is the distribution of the approximated likelihood based on the BSPF using 39,600 particles, the red and black dotted lines are based on the TPF using 10,800 particles with the target inefficiency ratios $r^* = 2$ and $r^* = 3$, respectively. Even with the much smaller number of particles, the variance of the distribution based on the TPF with $r^* = 3$ is on par with that based on the BSPF.

Moreover, the distribution based on the BSPF is located leftward compared with the other two. This is consistent with the finding in Herbst and Schorfheide (2019), who demonstrate that the BSPF causes a downward bias in the approximated likelihood from its true value but that the TPF does not.

C Filtered Series for Full Sample Period

Figure A.2 shows the filtered mean estimates of the nominal interest rate, the notional rate, output growth, and inflation for the full sample period from 1983Q1 to 2019Q4, based on the for models: the BR notional and nominal rate models and their RE counterparts. Other than the ZLB period (gray shaded area), the filtered estimates of the nominal and notional rates are almost the same across the four models because of the small measurement errors for the nominal interest rate series. Regarding the estimates of output growth and inflation, some differences arise among the four models. However, these differences do not necessarily cause remarkable differences in the likelihood increments, as shown in Figure 2 of the paper.

Figure A.1: Comparison of the Tempered Particle Filter and the Bootstrap Particle Filter

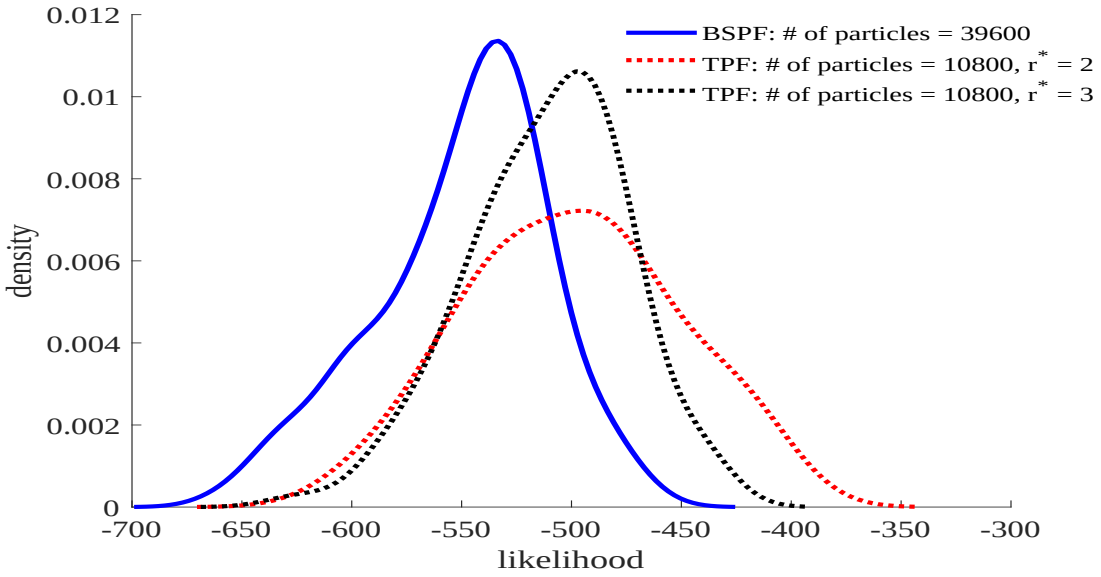


Figure A.2: Comparison of Fitted Series