Technical Appendix to "The Empirical Implications of the Interest-Rate Lower Bound"*

Christopher Gust
Federal Reserve Board

Edward Herbst
Federal Reserve Board

David López-Salido
Federal Reserve Board

Matthew E. Smith†
Hutchin Hill Capital

January 2017

Abstract

This technical appendix discusses in detail the solution and estimation of the nonlinear model described in "The Empirical Implications of the Interest-Rate Lower Bound."

*The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or any other person associated with the Federal Reserve System.

†This paper represents the personal work product and views of Matthew E. Smith and have neither been reviewed nor endorsed by Hutchin Hill Capital, LP, its affiliates, any funds advised by them, or any of their officers, directors, employees, agents, partners, principals, owners, or shareholders (collectively, "Hutchin Hill Parties"). No representation or warranty is made by Matthew E. Smith or any Hutchin Hill Party about the completeness or accuracy of any information provided herein, and such parties expressly disclaim all liability for any of the information contained in these materials. These materials have been prepared for educational use only and may not be relied upon for any other purpose (including without limitation any commercial purpose). These materials do not constitute an offer, or the solicitation of an offer, to buy or sell any security or financial instrument or investment advisory services. No Hutchin Hill Party shall be liable or have any obligation with respect to these slides or any reliance thereon.

1 Introduction

This appendix describes the equilibrium conditions of the model discussed in our paper, "The Empirical Implications of the Interest Rate Lower Bound." It uses those equilibrium conditions to characterize the solution as time-invariant functions. We approximate the nonlinear solution using a computationally efficient algorithm that is easily parallelizable and well-suited for handling an occasionally binding constraint. With this algorithm, we then describe the particle filtering algorithm for estimating the likelihood of a nonlinear state space system, highlighting the modifications made to facilitate estimation, including its parallelization. Finally, we lay out the particle filter Metropolis-Hastings algorithm used in conjuction with the particle filter to elicit draws from the posterior distribution.

2 Equilibrium Conditions

In a symmetric equilibrium, optimization by the firms in the economy implies:

$$[\frac{\pi_t}{\widetilde{\pi}_{t-1}}-1]\frac{\pi_t}{\widetilde{\pi}_{t-1}} = \beta E_t\left\{\frac{\Lambda_{t+1}}{\Lambda_t}\left[\frac{\pi_{t+1}}{\widetilde{\pi}_t}-1\right]\frac{\pi_{t+1}}{\widetilde{\pi}_t}\frac{Y_{t+1}}{Y_t}\right\} + \frac{\varepsilon_p}{\varphi_p}\left\{mc_t - \frac{\varepsilon_p - 1}{\varepsilon_p}\right\}, \tag{1}$$

$$(1-\alpha)mc_t = \frac{W_t N_t}{P_t Y_t}, \tag{2}$$

$$P_t r_t^k = \frac{\alpha}{1-\alpha}\frac{W_t N_t}{u_t \overline{K}_t}, \tag{3}$$

where the indexation term for price changes is given by:

$$\widetilde{\pi}_{t-1} = \overline{\pi}^a \pi_{t-1}^{1-a}. \tag{4}$$

The aggregate production function is given by:

$$Y_t = \left(u_t \overline{K}_t\right)^{\alpha}\left(Z_t N_t\right)^{1-\alpha}. \tag{5}$$

Household optimization for consumption, bond holdings, and wages implies:

$$\Lambda_t = \left[C_t - \gamma C_{t-1}\right]^{-1} - \gamma\beta E_t\left[C_{t+1} - \gamma C_t\right]^{-1}, \tag{6}$$

$$\Lambda_t = \beta R_t \eta_t E_t\left\{\Lambda_{t+1}\pi_{t+1}^{-1}\right\}, \tag{7}$$

$$\left[\frac{\pi_{w,t}}{\widetilde{\pi}_{w,t}}-1\right]\frac{\pi_{w,t}}{\widetilde{\pi}_{w,t}} = \beta E_t\left\{\left[\frac{\pi_{w,t+1}}{\widetilde{\pi}_{w,t+1}}-1\right]\frac{\pi_{w,t+1}}{\widetilde{\pi}_{w,t+1}}\right\} + N_t\Lambda_t\varepsilon_w\varphi_w^{-1}\left\{\psi_L\frac{N_t{}^{\sigma_L}}{\Lambda_t} - \frac{\varepsilon_w - 1}{\varepsilon_w}\frac{W_t}{P_t}\right\}. \tag{8}$$

In the above, wage inflation is defined by:

$$\pi_{w,t} = \frac{W_t}{W_{t-1}} \tag{9}$$

and the indexation term for wage changes is given by:

$$\widetilde{\pi}_{w,t} = G_Z\overline{\pi}^{a_w}\left(\exp\left(\epsilon_{Z,t}\right)\pi_{t-1}\right)^{1-a_w}. \tag{10}$$

A household's optimal choice of physical capital, investment, and utilization imply:

$$q_t = \beta E_t \left\{ \frac{\Lambda_{t+1}}{\Lambda_t} \left[r^k_{t+1} u_{t+1} - a(u_{t+1}) + (1-\delta) q_{t+1} \right] \right\}, \tag{11}$$

$$1 = q_t \mu_t \left(1 - \frac{\varphi_I}{2} \left(\frac{I_t}{G_Z I_{t-1}} - 1 \right)^2 - \varphi_I \left(\frac{I_t}{G_Z I_{t-1}} - 1 \right) \frac{I_t}{G_Z I_{t-1}} \right) + \tag{12}$$

$$\beta \varphi_I E_t \left\{ q_{t+1} \frac{\Lambda_{t+1}}{\Lambda_t} \mu_{t+1} \left(\frac{I_{t+1}}{G_Z I_t} - 1 \right) \frac{I^2_{t+1}}{G_Z I^2_t} \right\},$$

$$r^k_t = r^k \exp\left(\sigma_a (u_t - 1)\right), \tag{13}$$

where r^k denotes the rental cost of capital in the non-stochastic steady state and the utilization cost is given by:

$$a(u_t) = \frac{r^k}{\sigma_a} \{\exp\left(\sigma_a(u_t - 1)\right) - 1\}. \tag{14}$$

The capital stock evolves according to:

$$\overline{K}_{t+1} = (1-\delta)\overline{K}_t + \mu_t \left\{ 1 - \frac{\varphi_I}{2} \left(\frac{I_t}{G_Z I_{t-1}} - 1 \right)^2 \right\} I_t. \tag{15}$$

The central bank's desired or notional interest rate is given by:

$$\ln\left(\frac{R^N_t}{R}\right) = \rho_R \ln\left(\frac{R^N_{t-1}}{R}\right) + (1-\rho_R)\left[\gamma_\pi \ln\left(\frac{\pi_t}{\overline{\pi}}\right) + \gamma_x x^g_t + \gamma_g \ln\left(\frac{Y_t}{G_Z Y_{t-1}}\right)\right] + \epsilon_{R,t}, \tag{16}$$

where $R = \beta^{-1} G_Z \overline{\pi}$ denotes the steady state nominal rate. The output gap is given by:

$$x^g_t = \alpha \ln(u_t) + (1-\alpha)\left(\ln(N_t) - \ln(N)\right), \tag{17}$$

where N denotes the non-stochastic steady state value of hours worked. The nominal interest rate satisfies the zero lower bound (ZLB) constraint so that:

$$R_t = \max\left(1, R^N_t\right). \tag{18}$$

The economy's resource constraint is:

$$C_t + I_t + G_t + \frac{\varphi_p}{2}\left[\frac{\pi_t}{\widetilde{\pi}_{t-1}} - 1\right]^2 Y_t + a(u_t)\overline{K}_t = Y_t \tag{19}$$

The economy's disturbances evolve according to:

$$Z_t = Z_{t-1} G_Z \exp\left(\epsilon_{Z,t}\right), \quad \epsilon_{Z,t} \sim iid \ \mathrm{N}\left(0, \sigma^2_Z\right), \tag{20}$$

$$\ln(\eta_t) = \rho_\eta \ln(\eta_{t-1}) + \epsilon_{\eta,t}, \quad \epsilon_{\eta,t} \sim iid \ \mathrm{N}\left(0, \sigma^2_\eta\right), \tag{21}$$

$$\ln(\mu_t) = \rho_\mu \ln(\mu_{t-1}) + \epsilon_{\mu,t}, \quad \epsilon_{\mu,t} \sim iid \ \mathrm{N}\left(0, \sigma^2_\mu\right), \tag{22}$$

$$\ln(g_t) = (1-\rho_g)\ln g + \rho_g \ln(g_{t-1}) + \epsilon_{g,t}, \quad \epsilon_{g,t} \sim iid \ \mathrm{N}\left(0, \sigma^2_g\right), \tag{23}$$

where $g_t = \frac{1}{1 - \frac{G_t}{Y_t}}$ and g denotes its value in the non-stochastic steady state. The monetary policy disturbance satisfies $\epsilon_{R,t} \sim iid \ \mathrm{N}\left(0, \sigma^2_R\right)$.

2.1 Stationary Representation of Equilibrium Conditions

The random walk in the technology shock, expression (20), implies that some of the real variables will be non-stationary. The system of equations can be transformed so that it becomes stationary. The transformed variables are denoted as the lower case of a variable. So, $\mathrm{y}_t = \frac{Y_t}{Z_t}$, $c_t = \frac{C_t}{Z_t}$, $i_t = \frac{I_t}{Z_t}$, $w_t = \frac{W_t}{P_t Z_t}$, $\overline{k}_{t+1} = \frac{\overline{K}_{t+1}}{Z_t}$, and $\lambda_t = Z_t \Lambda_t$. It is also convenient to define $G_{Z,t} = G_Z \exp(\epsilon_{Z,t})$.

The stationary versions of the price and wage inflation equations are:

$$\left[\frac{\pi_t}{\widetilde{\pi}_{t-1}} - 1\right] \frac{\pi_t}{\widetilde{\pi}_{t-1}} = \beta E_t \left\{ \frac{\lambda_{t+1}}{\lambda_t} \left[\frac{\pi_{t+1}}{\widetilde{\pi}_t} - 1\right] \frac{\pi_{t+1}}{\widetilde{\pi}_t} \frac{\mathrm{y}_{t+1}}{\mathrm{y}_t} \right\} + \frac{\varepsilon_p}{\varphi_p} \left\{ mc_t - \frac{\varepsilon_p - 1}{\varepsilon_p} \right\}, \tag{24}$$

$$\left[\frac{\pi_{w,t}}{\widetilde{\pi}_{w,t}} - 1\right] \frac{\pi_{w,t}}{\widetilde{\pi}_{w,t}} = \beta E_t \left\{ \left[\frac{\pi_{w,t+1}}{\widetilde{\pi}_{w,t+1}} - 1\right] \frac{\pi_{w,t+1}}{\widetilde{\pi}_{w,t+1}} \right\} + N_t \lambda_t \varepsilon_w \varphi_w^{-1} \left\{ \psi_L \frac{N_t^{\sigma_L}}{\lambda_t} - \frac{\varepsilon_w - 1}{\varepsilon_w} w_t \right\}, \tag{25}$$

where the price and wage indexation terms are given by equations (4) and (10).

The stationary representations for the consumption Euler equation and its associated Lagrange multiplier can be written as:

$$\lambda_t = V_{\lambda,t} \equiv \beta \eta_t R_t E_t \{ \frac{\lambda_{t+1}}{G_{Z,t+1}} \pi_{t+1}^{-1} \}, \tag{26}$$

$$\lambda_t = \left[c_t - \gamma \frac{c_{t-1}}{G_{Z,t}} \right]^{-1} - \gamma \beta V_{c,t}, \tag{27}$$

where $V_{c,t}$, is defined as:

$$V_{c,t} \equiv E_t \left\{ \frac{1}{G_{Z,t+1}} \left[c_{t+1} - \gamma \frac{c_t}{G_{Z,t+1}} \right]^{-1} \right\}. \tag{28}$$

The stationary representations of the optimal conditions for capital and investment supply are given by:

$$q_t = V_{q,t} \equiv \beta E_t \left\{ \frac{\lambda_{t+1}}{\lambda_t G_{Z,t+1}} \left[r_{t+1}^k u_{t+1} - a(u_{t+1}) + (1-\delta) q_{t+1} \right] \right\}, \tag{29}$$

$$1 = q_t \mu_t \left(1 - \varphi_I (\frac{i_t}{i_{t-1}} \exp(\epsilon_{Z,t}) - 1) \frac{i_t}{i_{t-1}} \exp(\epsilon_{Z,t}) \right) + V_{i,t}, \tag{30}$$

where $V_{i,t}$ is defined as:

$$V_{i,t} \equiv \beta \varphi_I E_t \left\{ q_{t+1} \mu_{t+1} \frac{\lambda_{t+1}}{\lambda_t} \left(\frac{i_{t+1}}{i_t} \exp(\epsilon_{Z,t+1}) - 1 \right) \left(\frac{i_{t+1}}{i_t} \right)^2 \exp(\epsilon_{Z,t+1}) \right\} - q_t \mu_t \frac{\varphi_I}{2} (\frac{i_t}{i_{t-1}} exp(\epsilon_{Z,t}) - 1)^2. \tag{31}$$

The stationary representation for the optimal utilization of capital is unchanged and given by equation (13). The expression for the utilization cost is also unchanged and given by equation (14).

After solving for output, the stationary representation for the resource constraint is:

$$\mathrm{y}_t = A_{y,t}^{-1} \left(c_t + i_t + a(u_t) \frac{\overline{k}_t}{G_{Z,t}} \right), \tag{32}$$

where $A_{y,t}$ is given by:

$$A_{y,t} = \frac{1}{g_t} - \frac{\varphi_p}{2} \left[\frac{\pi_t}{\widetilde{\pi}_{t-1}} - 1 \right]^2. \tag{33}$$

The stationary representation for the real wage is given by:

$$w_t = \frac{\pi_{w,t} w_{t-1}}{G_{Z,t}\pi_t}. \tag{34}$$

The stationary representation of the production function can be expressed to determine hours worked:

$$N_t = \left(\frac{u_t \overline{k}_t}{G_{Z,t}}\right)^{\frac{\alpha}{\alpha-1}} \mathrm{y}_t^{\frac{1}{1-\alpha}}. \tag{35}$$

Real marginal cost and the rental rate of capital are given by:

$$mc_t = \frac{w_t N_t}{(1-\alpha)\mathrm{y}_t}, \tag{36}$$

$$r_t^k = \frac{\alpha}{1-\alpha}\frac{G_{Z,t} w_t N_t}{u_t \overline{k}_t}. \tag{37}$$

The stationary representation of the capital accumulation equation is:

$$\overline{k}_{t+1} = (1-\delta)\frac{\overline{k}_t}{G_{Z,t}} + \mu_t \left[1 - \frac{\varphi_I}{2}(\frac{i_t}{i_{t-1}} \exp\left(\epsilon_{Z,t}\right) - 1)^2\right] i_t. \tag{38}$$

The notional rate is given by:

$$\ln\left(\frac{R_t^N}{R}\right) = \rho_R \ln\left(\frac{R_{t-1}^N}{R}\right) + (1-\rho_R)\left[\gamma_\pi \ln(\frac{\pi_t}{\overline{\pi}}) + \gamma_x x_t^g + \gamma_g \ln\left(\frac{\mathrm{y}_t \exp\left(\epsilon_{Z,t}\right)}{\mathrm{y}_{t-1}}\right)\right] + \epsilon_{R,t} \tag{39}$$

where the output gap, x_t^g, is defined in equation (17) and the nominal rate is defined in equation (18).

2.2 Steady State

In the non-stochastic steady state, equation (31) implies $V_i = 0$ and equation (30) implies $q = 1$. Similarly, in the non-stochastic steady state, $\pi = \widetilde{\pi} = \overline{\pi}$, $\pi_w = G_Z\overline{\pi}$, and $u = 1$. The other steady state relationships include:

$$mc = \frac{\varepsilon_p - 1}{\varepsilon_p}, \tag{40}$$

$$mc = \frac{wN}{(1-\alpha)\mathrm{y}}, \tag{41}$$

$$\lambda c = \left(1 - G_Z^{-1}\gamma\right)^{-1}\left(1 - \beta G_Z^{-1}\gamma\right), \tag{42}$$

$$R = \beta^{-1} G_Z \overline{\pi}, \tag{43}$$

$$\frac{i}{\overline{k}} = 1 - G_Z^{-1}(1-\delta), \tag{44}$$

$$\mathrm{y} = \left(G_Z^{-1}\overline{k}\right)^\alpha N^{1-\alpha}, \tag{45}$$

$$r^k = \beta^{-1} G_Z - 1 + \delta, \tag{46}$$

$$\varepsilon_w \psi_L \frac{N^{\sigma_L}}{\lambda c} = (\varepsilon_w - 1)\frac{w}{c}, \tag{47}$$

$$r^k = \frac{\alpha}{1-\alpha}\frac{\omega N}{\overline{k}/G_Z}, \tag{48}$$

$$\mathrm{y} = g\left(c + i\right). \tag{49}$$

From equations (46) and (48) it follows that:

$$\overline{k} = \frac{\alpha}{1-\alpha} \frac{G_Z \omega N}{(\beta^{-1} G_Z - 1 + \delta).}$$

Using equation (41) to replace wN yields the following expression for the capital-output ratio:

$$\frac{\overline{k}}{\mathrm{y}} = \frac{(\epsilon_p - 1)}{\epsilon_p} \frac{\alpha G_Z}{(\beta^{-1} G_Z - 1 + \delta)}.$$

The steady state capital to output ratio and equation (44) can be combined to determine the investment-output ratio:

$$\frac{i}{\mathrm{y}} = \left(1 - G_Z^{-1}(1-\delta)\right) \frac{\overline{k}}{\mathrm{y}} = \frac{(\epsilon_p - 1)}{\epsilon_p} \frac{\left(1 - G_Z^{-1}(1-\delta)\right) \alpha G_Z}{(\beta^{-1} G_Z - 1 + \delta)}.$$

From equation (49) it follows:

$$\frac{c}{\mathrm{y}} = \frac{1}{g} - \frac{i}{\mathrm{y}},$$

where g is a parameter fixed to match the sample average of the ratio of government spending to output, i.e., $g = \frac{1}{1-\frac{G}{Y}}$.

Combining equations (47) and (42) yields:

$$\psi_L (1 - \frac{\gamma}{G_Z}) N^{1+\sigma_L} \frac{c}{\mathrm{y}} = \frac{\varepsilon_w - 1}{\varepsilon_w} \frac{\omega N}{\mathrm{y}} (1 - \frac{\gamma}{G_Z} \beta).$$

Using expression (41) to rewrite the above expression gives:

$$\psi_L (1 - \frac{\gamma}{G_Z}) N^{1+\sigma_L} \frac{c}{\mathrm{y}} = \frac{\varepsilon_w - 1}{\varepsilon_w} mc(1-\alpha)(1 - \frac{\gamma}{G_Z} \beta).$$

This expression can be used to determine steady state hours:

$$N = \left[\frac{\frac{\varepsilon_w - 1}{\varepsilon_w} mc(1-\alpha)(1 - \frac{\gamma}{G_Z}\beta)}{\psi_L (1 - \frac{\gamma}{G_Z}) \frac{c}{\mathrm{y}}} \right]^{\frac{1}{1+\sigma_L}}. \tag{50}$$

3 Solution Algorithm

This section discusses the characterization of the model's solution and then how to approximate it.

3.1 Characterizing the Solution

The model's equilibrium conditions are written as time-invariant functions that depend on the minimum state vector, $(\mathbb{X}_{t-1}, \tau_t)$, where

$$\mathbb{X}_{t-1} = \left(\overline{k_t}, c_{t-1}, i_{t-1}, w_{t-1}, R^N_{t-1}, \pi_{t-1}, \mathrm{y}_{t-1}, \right), \tag{51}$$

$$\tau_t = (\eta_t, \mu_t, \epsilon_{Z,t}, \epsilon_{R,t}, g_t). \tag{52}$$

We denote the endogenous state vector, $\mathbb{X}_{t-1}$, as depending on date $t-1$ variables, because $\overline{k}_t$ is determined at date $t-1$.

It is convenient to define the time invariant functions, $V_\pi\left(\mathbb{X}_{t-1},\tau_t\right) \equiv V_{\pi,t}$ and $V_w\left(\mathbb{X}_{t-1},\tau_t\right) \equiv V_{w,t}$ using equations (24) and (25):

$$V_{\pi,t} = \beta E_t \left\{ \frac{\lambda_{t+1}}{\lambda_t} \left[\frac{\pi_{t+1}}{\widetilde{\pi}_t} - 1 \right] \frac{\pi_{t+1}}{\widetilde{\pi}_t} \frac{\mathrm{y}_{t+1}}{\mathrm{y}_t} \right\} + \frac{\varepsilon_p}{\varphi_p} \left\{ mc_t - \frac{\varepsilon_p - 1}{\varepsilon_p} \right\}, \tag{53}$$

$$V_{w,t} = \beta E_t \left\{ \left[\frac{\pi_{w,t+1}}{\widetilde{\pi}_{w,t+1}} - 1 \right] \frac{\pi_{w,t+1}}{\widetilde{\pi}_{w,t+1}} \right\} + N_t \lambda_t \varepsilon_w \varphi_w^{-1} \left\{ \psi_L \frac{N_t^{\sigma_L}}{\lambda_t} - \frac{\varepsilon_w - 1}{\varepsilon_w} w_t \right\}. \tag{54}$$

Given these two functions, equations (24) and (25) can be rewritten as:

$$\pi\left(\mathbb{X}_{t-1},\tau_t\right) = \frac{\widetilde{\pi}_{t-1}}{2} \left(1 + \sqrt{1 + 4V_{\pi,t}}\right) \tag{55}$$

$$\pi_w\left(\mathbb{X}_{t-1},\tau_t\right) = \frac{\widetilde{\pi}_{w,t}}{2} \left(1 + \sqrt{1 + 4V_{w,t}}\right) \tag{56}$$

where $\pi\left(\mathbb{X}_{t-1},\tau_t\right) \equiv \pi_t$ and $\pi_w\left(\mathbb{X}_{t-1},\tau_t\right) \equiv \pi_{w,t}$.

More broadly, the decision rules for the remainder of the endogenous state variables in $\mathbb{X}_{t-1}$ can be determined given the vector of functions:

$$\begin{aligned} V\left(\mathbb{X}_{t-1},\tau_t\right) = &\left(V_\pi\left(\mathbb{X}_{t-1},\tau_t\right), V_w\left(\mathbb{X}_{t-1},\tau_t\right), V_c\left(\mathbb{X}_{t-1},\tau_t\right), V_i\left(\mathbb{X}_{t-1},\tau_t\right),\right. \\ &\left. V_\lambda\left(\mathbb{X}_{t-1},\tau_t\right), V_q\left(\mathbb{X}_{t-1},\tau_t\right), V_u\left(\mathbb{X}_{t-1},\tau_t\right)\right), \end{aligned} \tag{57}$$

where $V_u\left(\mathbb{X}_{t-1},\tau_t\right) \equiv u_t$ is a function determining the equilibrium value for the utilization of capital. Using these functions, we determine the remaining endogenous variables as follows. The Lagrange multiplier on the household's budget constraint, $\lambda\left(\mathbb{X}_{t-1},\tau_t\right)$, can be determined using equation (26) and $V_\lambda\left(\mathbb{X}_{t-1},\tau_t\right)$. Equation (27) then determines the decision rule for consumption, $c\left(\mathbb{X}_{t-1},\tau_t\right)$ as:

$$c\left(\mathbb{X}_{t-1},\tau_t\right) = \gamma \frac{c_{t-1}}{G_{Z,t}} + \left(\lambda_t + \gamma\beta V_{c,t}\right)^{-1}. \tag{58}$$

Taking $V_{i,t}$ as given and applying the quadratic equation to expression (30), equilibrium investment is given by:

$$i\left(\mathbb{X}_{t-1},\tau_t\right) = \frac{i_{t-1}}{2\exp\left(\epsilon_{Z,t}\right)} \left[1 + \sqrt{1 + \frac{4}{\varphi_I}\left(1 - \frac{1 - V_{i,t}}{q_t \mu_t}\right)}\right]. \tag{59}$$

Given the functions for c_t, i_t, π_t and the vector of functions, $V\left(\mathbb{X}_{t-1},\tau_t\right)$, equilibrium output, $\mathrm{y}\left(\mathbb{X}_{t-1},\tau_t\right)$, is determined using expression (32). Equation (34) determines the real wage, $w\left(\mathbb{X}_{t-1},\tau_t\right)$, using the functions for π_t and $\pi_{w,t}$, while equation (35) determines hours worked, $N\left(\mathbb{X}_{t-1},\tau_t\right)$ using the functions for y_t and u_t. Equations (36)-(39) can be used to determine the functions, $mc\left(\mathbb{X}_{t-1},\tau_t\right)$, $r^k\left(\mathbb{X}_{t-1},\tau_t\right)$, $k\left(\mathbb{X}_{t-1},\tau_t\right) \equiv \overline{k}_{t+1}$, and $R^N\left(\mathbb{X}_{t-1},\tau_t\right)$. Accordingly, given $V\left(\mathbb{X}_{t-1},\tau_t\right)$, we can define the vector of functions that define the decision rules, $\mathbb{X}_t = g_X\left(\mathbb{X}_{t-1},\tau_t\right)$, where

$$\begin{aligned} g_X\left(\mathbb{X}_{t-1},\tau_t\right) = &\left(k\left(\mathbb{X}_{t-1},\tau_t\right), c\left(\mathbb{X}_{t-1},\tau_t\right), i\left(\mathbb{X}_{t-1},\tau_t\right), w\left(\mathbb{X}_{t-1},\tau_t\right),\right. \\ &\left. R^N\left(\mathbb{X}_{t-1},\tau_t\right), \pi\left(\mathbb{X}_{t-1},\tau_t\right), \mathrm{y}\left(\mathbb{X}_{t-1},\tau_t\right)\right). \end{aligned} \tag{60}$$

We use the functions, $V\left(\mathbb{X}_{t-1},\tau_t\right)$, to determine the decisions rule though we do not approximate $V\left(\mathbb{X}_{t-1},\tau_t\right)$, directly, because they inherit a kink associated with the zero lower bound constraint on the nominal rate. Instead, we follow the methodology described in Gust, López-Salido, and Smith (2012), which builds on Christiano and Fisher (2000). Specifically, we approximate functions, $V_{l,i}\left(\mathbb{X}_{t-1},\tau_t\right)$, that are smoother and easier to approximate by specifying:

$$V_l\left(\mathbb{X}_{t-1},\tau_t\right) = V_{l,1}\left(\mathbb{X}_{t-1},\tau_t\right)\mathbb{I}\left(\mathbb{X}_{t-1},\tau_t\right) + V_{l,2}\left(\mathbb{X}_{t-1},\tau_t\right)\left(1 - \mathbb{I}\left(\mathbb{X}_{t-1},\tau_t\right)\right). \tag{61}$$

for $l \in \{\pi, w, c, i, \lambda, q, u\}$ and $j = 1, 2$ and where $\mathbb{I}(\mathbb{X}_{t-1}, \tau_t)$ is defined by:

$$\begin{aligned} \mathbb{I}(\mathbb{X}_{t-1}, \tau_t) &= 1 \text{ if } R(\mathbb{X}_{t-1}, \tau_t) > 0 && (62) \\ &= 0 \text{ otherwise.} && (63) \end{aligned}$$

In the above, $R(\mathbb{X}_{t-1}, \tau_t) = \max\left(1, R_1^N(\mathbb{X}_{t-1}, \tau_t)\right)$ where $R_1^N(\mathbb{X}_{t-1}, \tau_t)$ denotes the value of the notional rate derived from evaluating the functions $V_{l,1}(\mathbb{X}_{t-1}, \tau_t)$ and using expression (39). (For each variable, we use $j = 1$ to denote the function associated with the regime with a positive nominal rate and $j = 2$ to denote that function associated with the ZLB regime; similarly, $g_{X,j}(\mathbb{X}_{t-1}, \tau_t)$ denotes the vector of regime-specific decision rules.)

The functions, $V_{l,j}(\mathbb{X}_{t-1}, \tau_t)$, satisfy the residual functions, $\nu_{l,j}(\mathbb{X}_{t-1}, \tau_t)$, for $l \in \{\pi, w, c, i, \lambda, q, u\}$ and $j = 1, 2$:

$$\begin{aligned} \nu_{\lambda,1}(\mathbb{X}_{t-1}, \tau_t) &= V_{\lambda,1}(\mathbb{X}_{t-1}, \tau_t) - \beta\eta_t R_t E_t\{\frac{\lambda(\mathbb{X}_t, \tau_{t+1})}{G_{Z,t+1}}\pi^{-1}(\mathbb{X}_t, \tau_{t+1})\} = 0, \quad (64) \\ \nu_{\lambda,2}(\mathbb{X}_{t-1}, \tau_t) &= V_{\lambda,2}(\mathbb{X}_{t-1}, \tau_t) - \beta\eta_t E_t\{\frac{\lambda(\mathbb{X}_t, \tau_{t+1})}{G_{Z,t+1}}\pi^{-1}(\mathbb{X}_t, \tau_{t+1})\} = 0, \end{aligned}$$

$$\nu_{c,j}(\mathbb{X}_{t-1}, \tau_t) = V_{c,j}(\mathbb{X}_{t-1}, \tau_t) - E_t\left\{\frac{1}{G_{Z,t+1}}\left[c(\mathbb{X}_t, \tau_{t+1}) - \gamma\frac{c_j(\mathbb{X}_{t-1}, \tau_t)}{G_{Z,t+1}}\right]^{-1}\right\} = 0, \quad (65)$$

$$\nu_{q,j}(\mathbb{X}_{t-1}, \tau_t) = V_{q,j}(\mathbb{X}_{t-1}, \tau_t) - \beta E_t\left\{\frac{\widetilde{\lambda}_j}{G_{Z,t+1}}\left[\widetilde{r}^k(\mathbb{X}_t, \tau_{t+1}) + (1-\delta)q(\mathbb{X}_t, \tau_{t+1})\right]\right\} = 0, \quad (66)$$

$$\begin{aligned} \nu_{i,j}(\mathbb{X}_{t-1}, \tau_t) = V_{i,j}(\mathbb{X}_{t-1}, \tau_t) - \beta\varphi_I E_t\left\{q(\mathbb{X}_t, \tau_{t+1})\mu_{t+1}\widetilde{\lambda}_j(\widetilde{i}_j \exp(\epsilon_{Z,t+1}) - 1)\widetilde{i}_j^2 \exp(\epsilon_{Z,t+1})\right\} \quad (67) \\ -q_j(\mathbb{X}_{t-1}, \tau_t)\mu_t\frac{\varphi_I}{2}\left(\frac{i_j(\mathbb{X}_{t-1}, \tau_t)}{i_{t-1}}\exp(\epsilon_{Z,t}) - 1\right)^2 = 0, \end{aligned}$$

$$\begin{aligned} \nu_{\pi,j}(\mathbb{X}_{t-1}, \tau_t) = V_{\pi,j}(\mathbb{X}_{t-1}, \tau_t) - \beta E_t\left\{\widetilde{\lambda}_j\left[\frac{\pi(\mathbb{X}_t, \tau_{t+1})}{\widetilde{\pi}_t} - 1\right]\frac{\pi(\mathbb{X}_t, \tau_{t+1})}{\widetilde{\pi}_t}\frac{\mathrm{y}(\mathbb{X}_t, \tau_{t+1})}{\mathrm{y}_j(\mathbb{X}_{t-1}, \tau_t)}\right\} + \quad (68) \\ \frac{\varepsilon_p}{\varphi_p}\left\{mc_j(\mathbb{X}_{t-1}, \tau_t) - \frac{\varepsilon_p - 1}{\varepsilon_p}\right\} = 0, \\ \nu_{w,j}(\mathbb{X}_{t-1}, \tau_t) = V_{w,j}(\mathbb{X}_{t-1}, \tau_t) - \beta E_t\left\{\left[\frac{\pi_w(\mathbb{X}_t, \tau_{t+1})}{\widetilde{\pi}_{w,t+1}} - 1\right]\frac{\pi_w(\mathbb{X}_t, \tau_{t+1})}{\widetilde{\pi}_{w,t+1}}\right\} + \quad (69) \\ N_j(\mathbb{X}_{t-1}, \tau_t)\lambda_j(\mathbb{X}_{t-1}, \tau_t)\varepsilon_w\varphi_w^{-1}\left\{\psi_L\frac{N_j(\mathbb{X}_{t-1}, \tau_t)^{\sigma_L}}{\lambda_j(\mathbb{X}_{t-1}, \tau_t)} - \frac{\varepsilon_w - 1}{\varepsilon_w}w_j(\mathbb{X}_{t-1}, \tau_t)\right\} = 0, \\ \nu_{u,j}(\mathbb{X}_{t-1}, \tau_t) = V_{u,j}(\mathbb{X}_{t-1}, \tau_t) - 1 - \frac{1}{\sigma_a}\log\left(\frac{r_j^k(\mathbb{X}_{t-1}, \tau_t)}{r^k}\right) = 0, \quad (70) \end{aligned}$$

where

$$\begin{aligned} \widetilde{r}^k(\mathbb{X}_t, \tau_{t+1}) &= r^k(\mathbb{X}_t, \tau_{t+1})u(\mathbb{X}_t, \tau_{t+1}) - a(u(\mathbb{X}_t, \tau_{t+1})), \\ \widetilde{\lambda}_j &\equiv \widetilde{\lambda}_j(\mathbb{X}_t, \mathbb{X}_{t-1}, \tau_t, \tau_{t+1}) = \frac{\lambda(\mathbb{X}_t, \tau_{t+1})}{\lambda_j(\mathbb{X}_{t-1}, \tau_t)}, \\ \widetilde{i}_j &\equiv \widetilde{i}_j(\mathbb{X}_t, \mathbb{X}_{t-1}, \tau_t, \tau_{t+1}) = \frac{i(\mathbb{X}_t, \tau_{t+1})}{i_j(\mathbb{X}_{t-1}, \tau_t)}, \end{aligned}$$

and $\mathbb{X}_t = g_{X,j}(\mathbb{X}_{t-1}, \tau_t)$.

Because the functions, $V_l(\mathbb{X}_{t-1}, \tau_t)$ depend directly on the nominal interest rate, we expect them to have a kink or non-differentiability. By contrast, the counterpart functions, $V_{l,j}(\mathbb{X}_{t-1}, \tau_t)$, that are indexed by the interest-rate regime do not depend on the current indicator function and thus are more likely to be smooth. The regime-specific functions still depend on a secondary effect that the kink in the nominal rate next period has on the expectations of future variables. This secondary effect enters through evaluating the decision rule in the next period (i.e., $X_{t+1} = g_X(\mathbb{X}_t, \tau_{t+1})$) but it does not affect $\mathbb{X}_t$ which depends on $g_{X,j}(\mathbb{X}_t, \tau_t)$. Following the arguments in Christiano and Fisher (2000), the secondary effects of the kink on the regime-specific functions should be small because of the presence of the expectations operator, which involves summing over the future states of τ and acts to smooth out the regime-specific functions. While our approach of using relatively smooth functions is similar to Christiano and Fisher (2000), our approach is more general as it does not require that we parameterize functions that depend only on future variables.

3.2 Approximating the Solution

We approximate the functions, $V_{l,j}(\mathbb{X}_{t-1}, \tau_t)$, as follows:

$$V_{l,j}(\mathbb{X}_{t-1}, \tau_t) \approx \sum_{k=1}^{N^\tau} T(\varphi(\mathbb{X}_{t-1}))\, a_{l,j,k} \Gamma_k(\tau_t), \tag{71}$$

where $T(\varphi(\mathbb{X}_{t-1}))$ is a 1×17 vector constructed from an anistropic Smolyak method. Specifically, $T(\varphi(\mathbb{X}_{t-1}))$ includes a constant and the first and second degree Chebyshev polynomials for each variable in $\mathbb{X}_{t-1}$. It also includes the third and fourth degree Chebyshev polynomials for investment. We augment investment with higher-order polynomials, because we found that in practice this helped reduce the size of the residual error, $\nu_{q,j}(\mathbb{X}_{t-1}, \tau_t)$. (See Judd, Maliar, Maliar, and Valero (2014) for a discussion of anistropic Smolyak methods.)

For each state variable in $\mathbb{X}_{t-1}$, we use $\varphi_f : [\underline{\mathbb{X}}_f, \overline{\mathbb{X}}_f] \to [-1, 1]$ for $f = 1, 2, .., 7$, where $\varphi_f(\mathbb{X}_{t-1,f}) = \frac{2(\mathbb{X}_{t-1,f} - \underline{\mathbb{X}}_f)}{\overline{\mathbb{X}}_f - \underline{\mathbb{X}}_f}$ and f indexes one of the state variable in $\mathbb{X}_{t-1}$. So $\varphi(\mathbb{X}_{t-1})$ is given by:

$$\varphi(\mathbb{X}_{t-1}) = [\varphi_1(\mathbb{X}_{t-1,1}), ..., \varphi_7(\mathbb{X}_{t-1,7})],$$

and $\overline{\mathbb{X}}_f$ and $\underline{\mathbb{X}}_f$) are maximum and minimum values of each state variable chosen to encompass a wide interval.

In equation (71), $a_{l,j,k}$ is a 17×1 vector of parameters for $l \in \{\lambda, c, q, i, \pi, w, u\}$, $j = 1, 2$, and $k = 1, 2, .., N^\tau$. The function $\Gamma_k(\tau_t)$ is the product of univariate piecewise linear basis functions for each variable in τ_t. These basis functions use evenly-spaced breakpoints which are also the interpolation nodes. We use 3 breakpoints for each shock except η_t for which we use 7 so that $N^\tau = 567$. We could in principle use a sparse grid to construct $\Gamma_k(\tau_t)$ as we did for $T(\varphi(\mathbb{X}_{t-1}))$.

Our solution strategy involves finding the matrix of coefficients, $a^\star$ such that:

$$\nu_{l,j}(\mathbb{X}_m, \tau_k; a^\star) = 0, \tag{72}$$

for $j = 1, 2$, and $l \in \{\pi, w, c, i, \lambda, q, u\}$. In equation (72), $\mathbb{X}_m$ denotes that the vector of state variables is evaluated at each point, $m = 1, 2, .., 17$, on an appropriately constructed Smolyak grid using Chebyshev extrema for the unidimensional grid points. Also, each exogenous state variable is evaluated at one of its equally-spaced breakpoints, $k = 1, 2, .., N^\tau$. The matrix $a^\star$ consists of:

$$a^\star = \left[a^\star_{\lambda,1,1}, ..., a^\star_{\lambda,1,N^\tau}, a^\star_{\lambda,2,1}, ..., a^\star_{\lambda,2,N^\tau}, a^\star_{c,1,1}, ..., a^\star_{u,2,N^\tau}\right],$$

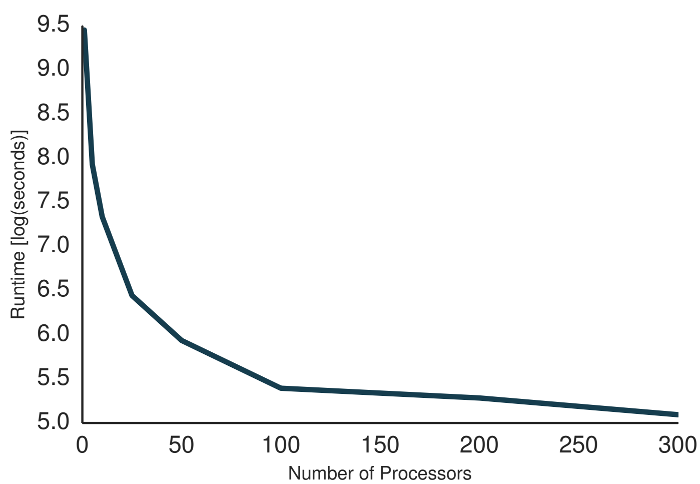
and is $17 \times 14N^\tau$. With $N^\tau = 567$, the matrix $a^\star$ has 134,946 elements. We find even though this a very large number of coefficients, we are able to solve and estimate the model because, as discussed below, there are substantial gains to parallelizing the solution algorithm.

Before discussing how we determine $a^{\star}$, it is important to note that we use Gauss-Hermite integration to approximate the conditional expectations operator in $\nu_{l,j}\left(\mathbb{X}_{t-1}, \tau_t\right)$. We use 3 nodes per shock and construct the multidimensional integral as a tensor product of the one-dimensional nodes so that there are 243 nodes in total. It may be possible to speed up the solution algorithm further using the monomial rule discussed in Judd, Maliar, and Maliar (2011) to approximate the multidimensional integrals.

3.3 Parallelization of Solution Algorithm

For an initial guess of $a^{\star}$, we can evaluate the decision rule, $g_X\left(\mathbb{X}_t, \tau_t\right)$,, and conditional expectations operators in equations (64)-(70). With these expressions in hand, we use the fixed point algorithm described in Judd, Maliar, Maliar, and Valero (2014) to update our guess for $a^{\star}$. Given that our polynomial approximation is linear in these coefficients, updating these coefficients involves only trivial calculations and avoids using a numerical routine for solving nonlinear equations. Moreover, updating these coefficients is easily parallelizable, as updating each of the $17X1$ vectors, $a_{l,j,k}$, involves a relatively small set of calculations that is independent from the calculations necessary to update the other coefficients. Using a Message Passing Interface, we can distribute this updating step for $a_{l,j,k}$ across processors and make the necessary calculation independently. Figure 1 shows the runtime for solving the model 100 times (in log seconds) against the number of processors used in the procedure. With 1 processor, the model takes about three and half hours to solve 100 times, or a little over 2 minutes on average. With 300 processors, the model takes about 3 minutes to solve 100 times, or about 1.8 seconds on average.

Figure 1: Effects of Parallelization: Model Solution



4 Particle Filter

This section describes the particle filter used to estimate the likelihood given a set of parameters. The literature on particles filters is vast: surveys and tutorials can be found, for instance, in Arulampalam, Maskell, Gordon, and Clapp (2002), Cappé, Godsill, and Moulines (2007), Doucet and Johansen (2011), and Creal (2012). The presentation of the algorithm is adapted from Herbst and Schorfheide (2015).

The starting point is the nonlinear state space model:

$$s_t = \Phi(s_{t-1}, \epsilon_t; \theta), \quad \epsilon_t \sim N(0, I), \tag{73}$$

$$y_t = \Psi(s_t, \epsilon_t; \theta) + u_t, \quad u_t \sim N(0, \Sigma_u), \tag{74}$$

where

$$\begin{aligned} s_t &= \left[\overline{k}_{t+1}, \mathrm{y}_t, c_t, i_t, \pi_t, R_t^N, w_t, \eta_t, \mu_t, g_t, \mathrm{y}_{t-1}, c_{t-1}, i_{t-1}\right], \\ \epsilon_t &= \left[\epsilon_{\eta,t}, \epsilon_{\mu,t}, \epsilon_{Z,t}, \epsilon_{g,t}, \epsilon_{R,t}\right], \text{ and} \\ \theta &= \left[\beta, \bar{\Pi}, g_z, \alpha, \rho_R, \gamma_\Pi, \gamma_g, \gamma_x, \gamma, \sigma_L, \sigma_a, \varphi_I, \varphi_p, \varphi_w, a, a_w, \rho_g, \rho_\mu, \sigma_\eta, \sigma_{\mu_I}, \sigma_Z, \sigma_g, \sigma_R\right]. \end{aligned}$$

The nonlinear transition equations, $\Phi(s_{t-1}, \epsilon_t; \theta)$, are formed using the nonlinear approximation to the decision rules, $g_X(\mathbb{X}_t, \tau_t)$, described in Section 3 and the transition equations for the shocks:

$$\tau_t = \Omega_1 \tau_{t-1} + \epsilon_t, \tag{75}$$

where Ω_1 is a 5×5 diagonal matrices whose diagonal elements are the AR(1) coefficients of the shocks.

The particle filter recursively produces discrete approximations to the distribution of the states, s_t, condition on time $t-1$ information (*forecasting distribution*) and t information (*updated distribution*). We generically refer to the set of the tuples $\{s_t^j, W_t^j\}_{j=1}^M$, which approximates $s_t|Y_{1:t}$, as particles, where M denotes the number of particles in the approximation. The object s_t^j references a point in the state space and W_t^j denotes the weight associated with that point. Note that $\sum_j^M W_t^j = M$.

The recursive formulation is useful for helping to understand the particle filter. Given a time $t-1$ particle approximation $\{s_{t-1}^j, W_{t-1}^j\}_{j=1}^M$, obtain a t time particle approximation, roughly speaking, as follows. First, simulate forward the proposed particles s_{t-1}^j to obtain particles s_t^j, and second, reweight these particles using the new data, y_t. The first step is known as forecasting and the second updating. We use the simplest particle filter—with some modifications described in Section 4.1—known as the Bootstrap particle filter (BSPF), which was first used in Gordon, Salmond, and Smith (1993). The key idea is that forecasted states come by drawing ϵ_t^j and iterating the state equation (73) forward to obtain: $s_t^j = \Phi(s_{t-1}^j, \epsilon_t^j; \theta)$. A nice by-product of the BSPF is that the likelihood function can be computed directly from the weights. The details are produced in Algorithm 1.

Algorithm 1 (Bootstrap Particle Filter)

1. **Initialization.** *Simulate the initial particles from the distribution* $s_0^j \overset{iid}{\sim} p(s_0)$ *and set* $W_0^j = 1$, $j = 1, \ldots, M$.
2. **Recursion.** *For* $t = 1, \ldots, T$:

   (a) **Forecasting s_t.** *Propagate the period* $t-1$ *particles* $\{s_{t-1}^j, W_{t-1}^j\}$ *by iterating the state-transition equation forward:*

   $$\tilde{s}_t^j = \Phi(s_{t-1}^j, \epsilon_t^j; \theta), \quad \epsilon_t^j \sim F_\epsilon(\cdot; \theta). \tag{76}$$

   An approximation of $\mathbb{E}[h(s_t)|Y_{1:t-1}, \theta]$ *is given by*

   $$\hat{h}_{t,M} = \frac{1}{M} \sum_{j=1}^M h(\tilde{s}_t^j) W_{t-1}^j. \tag{77}$$

   (b) **Forecasting y_t.** *Define the incremental weights*

   $$\tilde{w}_t^j = p(y_t|\tilde{s}_t^j, \theta). \tag{78}$$

The predictive density $p(y_t|Y_{1:t-1},\theta)$ can be approximated by

$$\hat{p}(y_t|Y_{1:t-1},\theta) = \frac{1}{M}\sum_{j=1}^{M}\tilde{w}_t^j W_{t-1}^j. \tag{79}$$

The incremental weights take the form

$$\tilde{w}_t^j \quad = \quad (2\pi)^{-n/2}|\Sigma_u|^{-1/2}\exp\left\{-\frac{1}{2}\big(y_t-\Psi(\tilde{s}_t^j,t;\theta)\big)'\Sigma_u^{-1}\big(y_t-\Psi(\tilde{s}_t^j,t;\theta)\big)\right\}, \tag{80}$$

where n here denotes the dimension of y_t.

(c) **Updating.** *Define the normalized weights*

$$\tilde{W}_t^j = \frac{\tilde{w}_t^j W_{t-1}^j}{\frac{1}{M}\sum_{j=1}^{M}\tilde{w}_t^j W_{t-1}^j}. \tag{81}$$

An approximation of $\mathbb{E}[h(s_t)|Y_{1:t},\theta]$ is given by

$$\tilde{h}_{t,M} = \frac{1}{M}\sum_{j=1}^{M}h(\tilde{s}_t^j)\tilde{W}_t^j. \tag{82}$$

(d) **Selection.** *Define the Effective Sample Size as:*

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

Let $\rho_t = \mathbf{1}_{\{ESS_t<M/2\}}$. Case (i): *If $\rho_t = 1$ resample the particles via multinomial resampling. Let $\{s_t^j\}_{j=1}^M$ denote M iid draws from a multinomial distribution characterized by support points and weights $\{\tilde{s}_t^j,\tilde{W}_t^j\}$ and set $W_t^j = 1$ for $j =, 1\ldots, M$.*
Case (ii): *If $\rho_t = 0$, let $s_t^j = \tilde{s}_t^j$ and $W_t^j = \tilde{W}_t^j$ for $j =, 1\ldots, M$.*
An approximation of $\mathbb{E}[h(s_t)|Y_{1:t},\theta]$ is given by

$$\bar{h}_{t,M} = \frac{1}{M}\sum_{j=1}^{M}h(s_t^j)W_t^j. \tag{83}$$

3. **Likelihood Approximation.** *The approximation of the log-likelihood function is given by*

$$\ln\hat{p}(Y_{1:T}|\theta) = \sum_{t=1}^{T}\ln\left(\frac{1}{M}\sum_{j=1}^{M}\tilde{w}_t^j W_{t-1}^j\right). \tag{84}$$

4.1 Adaption

While the BSPF is extremely simple to implement, it can perform poorly in practice, particularly in the presence of outliers. The forecast distribution can be highly mismatched with the updated distribution, which manifests itself as an extremely uneven distribution of the weights, and thus imprecise estimates of the likelihood function. To avoid degeneration of the particle filter in $t =$2008:Q4, we adapt the innovations of the proposal distribution, similar to Aruoba, Cuba-Borda, and Schorfheide (2016).

Details. For the BSPF, $\tilde{s}_t$ is generated by drawing $\epsilon_t^j \sim N(0,I)$ and using equation (73), and has density $p(\tilde{s}_t^j|s_{t-1}^j;\theta)$. Under a generic PF, we construct $\tilde{s}_t$ by sampling from an arbitrary distribution with density $g_t(\tilde{s}_t^j|s_{t-1}^j;\theta)$. Specifically, we simulate ϵ_t^j instead from $N(\mu,\Sigma)$, to elicit a proposal from $g_t(\tilde{s}_t^j|s_{t-1}^j;\theta)$.

When using this proposal distribution, the weights in the particle filter must be adjusted by a factor:

$$\kappa = \frac{p(\tilde{s}_t^j|s_{t-1}^j;\theta)}{g_t(\tilde{s}_t|s_{t-1}^j;\theta)}. \tag{85}$$

When applying a change of variable formula to represent $p(\tilde{s}_t^j|s_{t-1}^j;\theta)$ and $g_t(\tilde{s}_t^j|s_{t-1}^j;\theta)$, both densities contain the same Jacobian. This term drops out from the multiplicative ratio in (85), and it is easy to deduce that

$$\kappa = \frac{\exp\left\{-\frac{1}{2}\epsilon_t^{j\prime}\epsilon_t^j\right\}}{|\Sigma|^{-1/2}\exp\left\{(\epsilon_t^j-\mu)'\Sigma^{-1}(\epsilon_t^j-\mu)\right\}}, \tag{86}$$

with n the dimensionality of y_t. Unlike Aruoba, Cuba-Borda, and Schorfheide (2016), we do not use a gridsearch algorithm to generate μ and Σ. Instead, after extensive experimentation, for $t=$2008:Q4, we set

$$\mu = [3,0,0,0,0]' \text{ and } \Sigma = 1.2\times I.$$

For every other time period we set $\mu=0$ and $\Sigma=I$; i.e., we use the standard BSPF.

5 Parallelization of the Particle Filter

Parallelization of the particle filter is difficult because of the required communication during the resampling phase. We construct a particle filter adapted to a distributed computing environment. In this section, we sketch out some key aspects of the filter.

Suppose that we have K processors[1] and we are using M total particles in the particle filter. Let $M_{local}=M/K$, assume that M_{local} is an integer, and let $s_t^{i,k}$ denote the ith particle on the kth processor and similarly for $W_t^{i,k}$. It is obvious that the forecasting step can be done in parallel across the K processors. To update, we use a Message Passing Interface to aggregate the weights. The key part of the parallelized particle filter is selection and rebalancing. First, we resample every period ($\rho_t=1$, for all t), but we only resample (using systematic resampling) among the M_{local} particles on each processor. Instead of a sample with uniform weights, we are left with a sample which is evenly weighted *on a given processor*, with each particle having weight:

$$W_t^{i,k} = {M_{local}}^{-1}\sum_{j=1}^{M_{local}}\tilde{W}_t^{j,k}.$$

We account for the fact that distribution of total weight across processors can become uneven using the following procedure. Let:

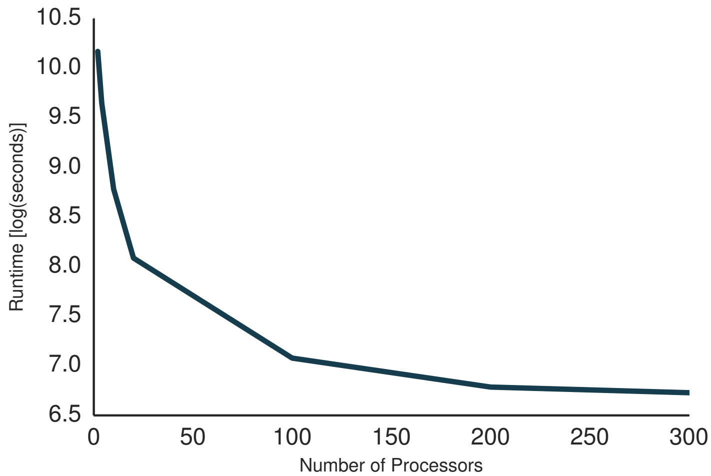
$$\alpha_k = \frac{\sum_{i=1}^{M_{local}} W_t^{i,k}}{\sum_{j=1}^{K}\sum_{i=1}^{M_{local}} W_t^{i,j}}, \tag{87}$$

where α_k is the mass of the particle distribution located on processor k. Define the effective number of processors as

$$EP_t = \frac{1}{\sum_{k=1}^{K}\alpha_k^2}. \tag{88}$$

[1]This is shorthand for whatever the lowest unit of instruction is.

Figure 2: Effects of Parallelization: Particle Filter



If $EP_t < K$, then reshuffle the particles among the processors by first ranking the processors according to α_k. Then assign each processor a partner in reverse order: the k with the largest α_k with the k with the smallest α_k, and so on. To rebalance the weights across particles, have each partner exchange $M_{exchange}$ ($< M_{local}$) particles with one another deterministically (i.e., use the first $M_{exchange}$ particles).

This procedure helps ensure that particles will not degenerate on any one processor, while still reducing the size of the resampling problem. Moreover, parallelization allows one to use extremely large number of particles without performance degradation because of memory constraints. Figure 2 demonstrates the effectiveness of parallelization. by plotting the run time for 100 evaluations of the particle filter on the vertical axis and for different values of K in Figure 2. As seen in the Figure, using two processors, it takes about 7 hours for 100 evaluations of the particle filter, or 4 minutes and 12 seconds per evaluation. When using 300 processors, it takes only about 13 minutes. While the returns to parallelization are not as large using as for the solution algorithm, they are still substantial.

6 Stability Analysis

It is crucial for the particle filter to produce *stable*—i.e., with low variance—estimates of the likelihood. When used within an MCMC algorithm, particle-filter-based estimators of the likelihood with high variance will tend to get stuck at a single value after an improbably high likelihood estimate. This means that the chain generated by the MCMC algorithm will converge very slowly (or not at all). In this section, we repeatedly apply the particle filter on a single representative parameter draw to show that our filter produces stable estimates.

In this section, we repeatedly apply the particle filter to a single, representative parameter value to highlight the stability of our particle filter. We also examine the contribution of the adaption discussed in Section 4.1. The parameter values use are given in Table 1.

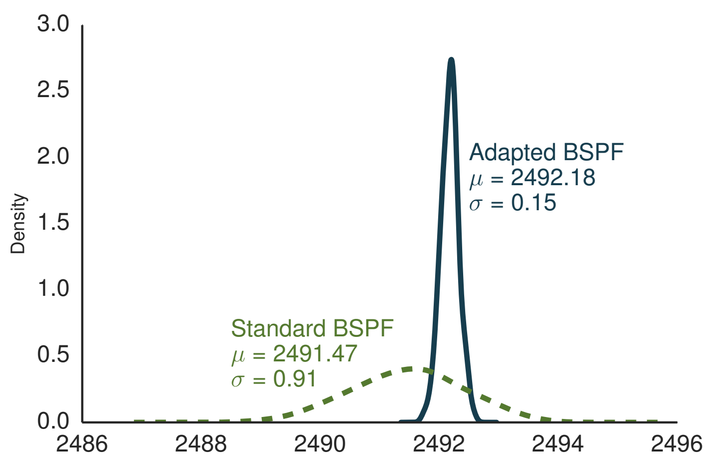
We apply the particle filter $N_{sim} = 100$ using $M = 500,000$ particles. (In the actual estimation, we use $M = 1,500,000$.) The sampling distribution of standard BSPF and the adapted BSPF are shown in Figure 3. Both estimators are fairly stable, with standard deviations less than one, the rule-of-thumb given in Pitt, Silva, Giordani, and Kohn (2012). However, the adapted BSPF has a

Table 1: Parameter Values for Stability Analysis

Parameter	Value	Parameter	Value	Parameter	Value
$100(\beta^{-1}-1)$	0.11	γ	0.69	ρ_g	0.53
$100(\bar{\Pi}-1)$	0.60	σ_L	1.27	$100\sigma_g$	0.14
$100\ln(G_z)$	0.54	σ_a	4.60	ρ_{μ_I}	0.72
α	0.21	φ_I	2.79	$100\sigma_{\mu_I}$	2.33
ρ_R	0.77	φ_p	61.32	$100\sigma_\eta$	0.45
γ_Π	1.18	$1-a$	0.59	$100\sigma_Z$	0.56
γ_g	0.91	φ_w	4615.94	$100\sigma_R$	0.15
γ_x	0.20	$1-a_w$	0.69		

standard deviation of 0.15 about six times less than the standard BSPF, indicating that there is substantial gain to adapting the particle filter during the Great Recession period.

Figure 3: Sampling Distribution of Log Likelihood Function Plus Log Prior



7 Particle Filter Metropolis-Hastings

In this section we describe the particle filter Metropolis-Hastings (PFMH) algorithm for generating a Markov Chain that converges to the posterior distribution of interest. The general algorithm to construct $\{\theta^i\}_{i=0}^N$ is given in Algorithm 2.

Algorithm 2 (PFMH Algorithm) *For $i = 1$ to N:*

1. *Draw ϑ from a density $q(\vartheta|\theta^{i-1})$.*
2. *Set $\theta^i = \vartheta$ with probability*

$$\alpha(\vartheta|\theta^{i-1}) = \min\left\{1,\ \frac{\hat{p}(Y|\vartheta)p(\vartheta)/q(\vartheta|\theta^{i-1})}{\hat{p}(Y|\theta^{i-1})p(\theta^{i-1})/q(\theta^{i-1}|\vartheta)}\right\}$$

and $\theta^i = \theta^{i-1}$ otherwise. The likelihood approximation $\hat{p}(Y|\vartheta)$ is computed using Algorithm 1.

As in Fernandez-Villaverde and Rubio-Ramirez (2007), we use the random walk variant of the PFMH algorithm, which amounts to:

$$q(\cdot|\theta^{i-1}) = N(\theta^{i-1}, c\hat{\Sigma}). \tag{89}$$

The key choices in this algorithm are the matrix $\hat{\Sigma}$ and scaling factor c. We set

$$\hat{\Sigma} = diag(\mathbb{V}[\{\theta^j\}_{j=1}^{N_{tuning}}]),$$

the estimated variance from a tuning run with $N_{tuning} = 5000$[2] and we set c to ensure a reasonable acceptance rate. Table 2 gives an overview of the hyperparameter choices we made for the PFMH algorithm. Setting $c = 0.20$ yields us an average acceptance rate of about 27 percent. To obtain 50,000 draws, the algorithm took about 10 days.

Table 2: PFMH details

Object	Description	Value
N	Length of Chain	50000
K	Number of Processors	336
M	Number of Particles in Adapted BSPF	1500000
$\hat{\Sigma}$	Proposal Variance	Tuning Run
c	Scaling Factor	0.2
Acceptance Rate		0.27
Run Time		10 days

Notes: We run 4 chains each of length 50000.

Table 3 reproduces Table 1 in the main text, while also providing the standard deviation of the posterior means across the 4 runs.

8 Computational Environment

We performed all computations at the High Performance Computing (HPC) Cluster maintained at the Federal Reserve Board. The project is coded in `Fortran`, and compiling using the **Intel Fortran Compiler** (version: 13.1.0 20130121), including the Math Kernel Library. The distributed aspects of the computation (i.e., parallelization) use `MPICH`, an implementation of the Message Passing Interfact standard–the servers are connected using `Infiniband`, a high-speed, low-latency connection.

[2]The proposal variance for the tuning run came from an estimation the linearized version of the model.

Table 3: Posterior Distribution – Nonlinear Model

Parameter	Mean	[05, 95]	Parameter	Mean	[05, 95]
		Steady State			
$100(\beta**{-1}-1)$	0.14 (0.01)	[0.06, 0.23]	$100(\bar{\Pi}-1)$	0.61 (0.01)	[0.54, 0.68]
$100\log(G_z)$	0.50 (0.00)	[0.46, 0.54]	α	0.19 (0.00)	[0.16, 0.22]
		Policy Rule			
ρ_R	0.70 (0.01)	[0.59, 0.78]	γ_Π	1.67 (0.04)	[1.21, 2.14]
γ_g	0.73 (0.02)	[0.39, 1.07]	γ_x	0.14 (0.01)	[0.07, 0.24]
		Endogenous Propagation			
γ	0.70 (0.01)	[0.63, 0.76]	σ_L	2.00 (0.05)	[1.01, 3.17]
σ_a	5.32 (0.08)	[3.78, 7.09]	φ_I	3.70 (0.15)	[2.24, 5.21]
φ_p	100.41 (1.09)	[65.10, 136.88]	$1-a$	0.56 (0.00)	[0.36, 0.76]
φ_w	4420.49 (170.02)	[1693.15, 8356.34]	$1-a_w$	0.51 (0.02)	[0.29, 0.72]
		Exogenous Processes			
ρ_G	0.67 (0.04)	[0.29, 0.96]	$100\sigma_G$	0.15 (0.00)	[0.11, 0.20]
ρ_{μ_I}	0.80 (0.02)	[0.64, 0.92]	$100\sigma_{\mu_I}$	2.39 (0.16)	[1.43, 3.70]
$100\sigma_\eta$	0.44 (0.02)	[0.34, 0.54]	$100\sigma_Z$	0.56 (0.02)	[0.38, 0.80]
$100\sigma_R$	0.18 (0.01)	[0.14, 0.24]			

Notes. Table reports the mean, fifth, and ninety-fifth percentile of the posterior distribution estimated by pooling 4 MCMC chains with 50,000 draws each (including 10,000 draw burn-in period.) The number in parentheses is the standard deviation of the mean across the 4 runs.

References

Arulampalam, S., S. Maskell, N. Gordon, and T. Clapp (2002): "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking," *IEEE Transactions on Signal Processing*, 50, 174–188.

Aruoba, B., P. Cuba-Borda, and F. Schorfheide (2016): "Macroeconomic Dynamics Near the ZLB: A Tale of Two Countries," International Finance Discussion Paper 1163.

Cappé, O., S. J. Godsill, and E. Moulines (2007): "An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo," *Proceedings of the IEEE*, 95, 899–924.

Christiano, L. and J. Fisher (2000): "Algorithms for Solving Dynamic Models with Occasionally Binding Constraints," *Journal of Economic Dynamics and Control*, 24, 1179–1232.

Creal, D. (2012): "A Survey of Sequential Monte Carlo Methods for Economics and Finance," *Econometric Reviews*, 31, 245–296.

Doucet, A. and A. M. Johansen (2011): "A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later," in *Handook of Nonlinear Filtering*, ed. by D. Crisan and B. Rozovsky, Oxford University Press.

Fernandez-Villaverde, J. and J. Rubio-Ramirez (2007): "Estimating Macroeconomic Models: A Likelihood Approach," *Review of Economic Studies*, 74, 1059–1087.

Gordon, N. J., D. J. Salmond, and A. F. Smith (1993): "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation," in *IEE Proceedings F, Radar and Signal Processing*, IET, vol. 140, 107–113.

Gust, C., D. López-Salido, and M. E. Smith (2012): "The Empirical Implications of the Interest-Rate Lower Bound," Finance and Economics Discussion Papers, Federal Reserve Board, 2012-83.

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

Judd, K. L., L. Maliar, and S. Maliar (2011): "Numerically Stable and Accurate Stochastic Simulation Approaches for Solving Dynamic Economic Models," *Quantitative Economics*, 2, 173–210.

Judd, K. L., L. Maliar, S. Maliar, and R. Valero (2014): "Smolyak Method for Solving Dynamic Economic Models: Lagrange Interpolation, Anisotropic Grid and Adaptive Domain," *Journal of Economic Dynamics and Control*, 44, 92–123.

Pitt, M. K., R. d. S. Silva, P. Giordani, and R. Kohn (2012): "On Some Properties of Markov Chain Monte Carlo Simulation Methods Based on the Particle Filter," *Journal of Econometrics*, 171, 134–151.