

# An Efficient Pseudo Marginal Method for State Space Models

David Gunawan<sup>\*</sup>, Pratiti Chatterjee<sup>‡</sup>, and Robert Kohn<sup>\*\*</sup>

September 29, 2021

## Abstract

Pseudo Marginal Metropolis-Hastings (PMMH) is a general approach to carry out Bayesian inference when the likelihood is intractable but can be estimated unbiasedly. Our article develops an efficient PMMH method for estimating the parameters of complex and high-dimensional state space models and has the following features. First, it runs multiple particle filters in parallel and uses their averaged unbiased likelihood estimate. Second, it combines block and correlated PMMH sampling. The first two features enable our sampler to scale up better to longer time series and higher dimensional state vectors than previous approaches. Third, the article develops an efficient auxiliary disturbance particle filter, which is necessary when the bootstrap filter is inefficient, but the state transition density cannot be expressed in closed form. Fourth, it uses delayed acceptance to make the make the sampler more efficient. The performance of the sampler is investigated empirically by applying it to Dynamic Stochastic General Equilibrium models with relatively high state dimensions and with intractable state transition densities. Although our focus is on applying the method to state space models, the approach will be useful in a wide range of applications such as large panel data models and stochastic differential equation models with mixed effects.

Keywords: Averaged likelihood estimate; block PMMH; Correlated PMMH; Delayed acceptance; Dynamic Stochastic General Equilibrium (DSGE) model

---

<sup>‡</sup> Level 4, West Lobby, School of Economics, University of New South Wales Business School – Building E-12, Kensington Campus, UNSW Sydney – 2052, *Email*: pratiti.chatterjee@unsw.edu.au, *Phone Number*: (+61) 293852150. *Website*: <http://www.pratitichatterjee.com>

<sup>\*</sup> 39C. 164, School of Mathematics and Applied Statistics (SMAS), University of Wollongong, Wollongong, 2522; Australian Center of Excellence for Mathematical and Statistical Frontiers (ACEMS); National Institute for Applied Statistics Research Australia (NIASRA); *Email*: dgu-

# 1 Introduction

The pseudo-marginal Metropolis-Hastings (PMMH) (Andrieu and Roberts, 2009) approach is a standard method for conducting Bayesian inference in a wide range of challenging statistical models having an intractable likelihood which can be estimated unbiasedly. Our article develops a PMMH sampling method for estimating parameters of the complex and high-dimensional state space models that is efficient and scalable (relative to current methods) in the number of observations and the latent states and can handle state transition densities that cannot be expressed in closed form; for example, many dynamic stochastic general equilibrium (DSGE) models, which are popular class of macroeconomic time series state space models, do not have closed form transition densities.

A key issue in efficiently estimating statistical models using a PMMH approach is that the variance of the log of the estimated likelihood grows with the number of observations and the dimension of the latent states (Deligiannidis et al., 2018). Pitt et al. (2012) show that to obtain a balance between computational time and the mixing of the Markov chain Monte Carlo (MCMC) chain, the number of particles used in the particle filter should be such that the variance of the log of the estimated likelihood is in the range 1 to 3, depending on the efficiency of the proposal for  $\theta$ , and that the efficiency of PMMH schemes deteriorates exponentially as that variance increases. In many complex statistical applications, it is computationally very expensive ensuring that the variance of the log of the estimated likelihood is within the required range. Deligiannidis et al. (2018) propose a more efficient PMMH scheme, called the correlated pseudo-marginal (CPM) method, which correlates the random numbers used in constructing the estimated likelihoods at the current and proposed values of the parameters. This correlation induces a correlation between the estimated likelihoods and reduces the variance of the difference in the logs of the estimated likelihoods which appears in the Metropolis-Hastings (MH) acceptance ratio. They show that the CPM scales up with the number of observations compared to the standard pseudo marginal method of Andrieu et al. (2010) if the state dimension is moderate.

Tran et al. (2016) propose an alternative approach, called the block pseudo marginal (BPM) method, which divides the random numbers into blocks and then updates the parameters jointly with one randomly chosen block of the random numbers in each MCMC iteration; this induces a positive correlation between the numerator

---

nawan@uow.edu.au. *Phone Number:* (+61) 424379015.

\*\* Level 4, West Lobby, School of Economics, University of New South Wales Business School – Building E-12, Kensington Campus, UNSW Sydney – 2052, and ACEMS *Email:* r.kohn@unsw.edu.au, *Phone Number:* (+61) 424802159.

and denominator of the MH acceptance ratio, similarly to the CPM. They showed that for large samples the correlation of the logs of the estimated likelihoods at the current and proposed values is close to  $1 - 1/G$ , where  $G$  is the number of blocks in the blocking approach.

Our paper proposes a new PMMH sampler, which we call the mixed PMMH algorithm (MPM); it builds on, and extends, the CPM method of Deligiannidis et al. (2018) and the BPM method of Tran et al. (2016). The important innovations of the MPM sampler compared to block and correlated PMMH are that: (a) the unbiased likelihood estimator is an average of the likelihood estimators obtained by multiple independent particle filters (PFs) which lowers its variance; these PFs can be run in parallel. (b) The unknown parameters, and only the random numbers used in one of the PFs, are updated jointly. Section 3.3 provides the implementation details. Sections 4.2 and 4.4 show that the MPM sampler is able to maintain the correlation between the logs of the estimated likelihoods for relatively high dimensional state space models compared to the BPM and CPM. (c) In many applications of state space models, such as DSGE models, it is difficult to estimate the likelihood efficiently because the bootstrap filter is very inefficient while it is difficult to construct an auxiliary particle filter because the state transition density does not have a closed form. We follow Murray et al. (2013) and Hall et al. (2014) and use the disturbance particle filter (DPF), which works on the disturbances rather than the states. An improved disturbance PF (IDPF) algorithm is proposed to improve the performance of the standard disturbance particle filter. Section 3.5 discusses this further. (d) The delayed-acceptance version (Christen and Fox, 2005) of the MPM algorithm is also developed to speed up the computation for non-linear state space models. The motivation for the delayed acceptance algorithm is to minimize the expensive computation of the likelihood or its estimate when it appears likely that the proposal will be rejected. See Section 3.4 for details.

We illustrate the MPM sampler empirically, and compare its performance to the BPM and CPM methods, using a standard linear Gaussian state space model and two Dynamic Stochastic General Equilibrium (DSGE) models, using simulated and real data. We also compare the performance of the improved disturbance particle filter (IDPF) to the tempered particle filter (TPF) proposed by Herbst and Schorfheide (2019); however, we note that Herbst and Schorfheide (2019) only apply the TPF to linear (first order) DSGE models. Sections 4.2 to 4.5 show that: (1) the IDPF is much more efficient than the standard bootstrap particle filter (BPF) and the tempered particle filter (TPF); (2) the MPM, which runs multiple independent particle filters and only updates the random numbers used in one particle filter, maintains the correlation between the logs of the estimated likelihoods much better than both

CPM and BPM. The BPM method of Tran et al. (2016), which runs a single particle filter and updates a block of random numbers for all time periods in the particle filter algorithm, is unsuitable for time-series state space model; (3) the delayed acceptance version of the MPM sampler is much more efficient than the standard MPM sampler.

Our work in estimating non-linear DSGE models is also related to Fernández-Villaverde and Rubio-Ramírez (2007) and Hall et al. (2014) who use standard PMMH methods. We note that the MPM sampler will also be useful for other complex statistical models, such as partially observed diffusions models, panel data models, and stochastic differential equation mixed effects models.

The rest of the article is organised as follows. Section 2 introduces the state space model and gives some examples. Section 3 discusses the MPM sampler. Section 4 presents results from both simulated and real data. Section 5 concludes with a discussion of our major results and findings. The paper has an online supplement containing some further technical and empirical results.

## 2 State Space Models

### 2.1 General State Space Models

We use the colon notation for collections of variables, i.e.,  $a_t^{r:s} := (a_t^r, \dots, a_t^s)$  and for  $t \leq u$ ,  $a_{t:u}^{r:s} := (a_t^{r:s}, \dots, a_u^{r:s})$ . Consider the stochastic process  $\{(Z_t, Y_t)\}$ , with parameter  $\theta$ , where the  $Y_t$  are the observations and the  $Z_t$  are the latent state vectors; random variables are denoted by capital letters and their realizations by lower case letters. We consider the state space model with  $p(z_0|\theta)$  the density of  $Z_0$ ,  $p(z_t|z_{t-1}, \theta)$  the density of  $Z_t = z_t$  given  $Z_{0:t-1} = z_{0:t-1}$  for  $t \geq 1$ , and  $p(y_t|z_t, \theta)$  is the density of  $Y_t = y_t$  given  $Z_{0:t} = z_{0:t}$ ,  $Y_{1:t-1} = y_{1:t-1}$ .

The stochastic volatility model is an example of a popular non-linear and non-Gaussian state space model. It is described by

$$y_t = \exp(z_t/2) \eta_t \quad (t \geq 1),$$

$$z_{t+1} = \mu + \phi(z_t - \mu) + \tau \epsilon_{t+1} \quad (t \geq 0), \quad \text{with } z_0 \sim N\left(\mu, \frac{\tau^2}{1 - \phi^2}\right);$$

$\{z_t\}$  is the latent volatility process,  $\theta := (\mu, \tau, \phi)$  is the vector of unknown parameters, with the persistence parameter  $\phi$  satisfying  $|\phi| < 1$  to ensure stationarity; the sequence  $(\eta_t, \epsilon_t)^\top \sim N(0, I)$  is independent for all  $t$ .

## 2.2 State space representations of DSGE models

Our article illustrates the proposed methods using Dynamic Stochastic General Equilibrium (DSGE) models; however, they apply more generally to nonlinear state space models.

We start by describing the state space representation of DSGE models and then highlight the specific source of nonlinearity we tackle in our applications. In the article we deal with DSGE models having a state transition equation of the form

$$z_{t+1} = F(z_t, \epsilon_{t+1}, \zeta; \theta), \quad (1)$$

such that  $\{\epsilon_t\}$  is an independent  $N(0, \Sigma_\epsilon)$  sequence;  $\zeta$  is a perturbation parameter and  $\theta$  is a vector of the unknown parameters of the DSGE model.

For most applications, the function  $F(\cdot)$  in Eq. (1) is analytically intractable and is approximated using local solution techniques. We use first and second order Taylor series approximations around the *deterministic* steady state  $z^s$  with  $\zeta = 0$  to approximate  $F$ ;  $z_s$  satisfies  $z^s = F(z^s, 0, 0; \theta)$ .

A first order-accurate approximation, around the *deterministic* steady state, to Eq. (1) is

$$z_{t+1}^d = F_1(\theta)z_t^d + F_2(\theta)\epsilon_{t+1} \quad (2)$$

where  $z_t^d = (z_t - z^s)$ . For Eq. (2) to be stable, it is necessary that all the eigenvalues of  $F_1(\theta)$  are less than 1 in absolute value. The initial value  $z_0^d = 0$  because we work with approximations around the *deterministic* steady state. For a given set of parameters  $\theta$ , the matrices  $F_1(\theta)$  and  $F_2(\theta)$  can be solved using existing software. Our applications use Dynare.

The second order accurate approximation (around the *deterministic* steady state) to Eq. (1) is

$$z_{t+1}^d = F_0(\theta)\zeta^2 + F_1(\theta)z_t^d + F_2(\theta)\epsilon_{t+1} + F_{11}(\theta)P(z_t^d) + F_{12}(\theta)(z_t^d \otimes \epsilon_{t+1}) + F_{22}(\theta)P(\epsilon_{t+1}); \quad (3)$$

where  $z_t^d = (z_t - z^s)$  and for any vector  $x := (x_1, \dots, x_m)^\top$ , we define  $P(x) := \text{vech}(xx^\top)$ , where  $\text{vech}(xx^\top)$  is the strict upper triangle and diagonal of  $xx^\top$ , such that

$\text{vech}(xx^\top) := (x_1^2, x_1x_2, x_2^2, x_1x_3, \dots, x_m^2)^\top$ . The term  $F_0(\theta)\zeta^2$  captures the level correction due to uncertainty that arises from taking a second-order approximation. For the scope of our analysis we normalize the perturbation parameter  $\zeta$  to 1. As before, for a given set of parameters  $\theta$  the matrices  $F_1(\theta)$ ,  $F_2(\theta)$ ,  $F_{11}(\theta)$ ,  $F_{12}(\theta)$ ,  $F_{22}(\theta)$  can be solved using Dynare.

The measurement (observation) equation for the DSGE model and its approximations is

$$y_t = Hz_t + \eta_t, \quad \eta_t \sim N(0, \Sigma_\eta), \quad (4)$$

where  $H$  is a known matrix;  $\{\eta_t\}$  is an independent  $N(0, \Sigma_\eta)$  and sequence and it is also independent of the  $\{\epsilon_t\}$  sequence. The matrix  $\Sigma_\eta$  is usually unknown and estimated from the data.

Although it is possible to apply the bootstrap filter for the second order solution, the resulting estimated likelihood is too variable to be useful in a PMMH scheme. Instead, we use an auxiliary disturbance particle filter because the transition density of the second order DSGE models is usually intractable.

We do not use Eq. (3) as the state space transition equation because it can generate explosive solutions for the variables in the model. Instead, this problem is solved by modifying the state transition equation using the pruning approach of Kim et al. (2008), as in Section S6.1 of the supplement. Such a pruned representation can substantially increase the size of the state vector; if  $z_t$  in Eq. (S8) is  $d$  dimensional, then the pruned state vector is  $2d + d(d+1)/2$  dimensional. The pruned state vector is then also likely to be substantially greater than the dimension of the disturbances  $\{\epsilon_t\}$ , thus giving a further motivation for using the disturbance filter.

## 3 Bayesian Inference

### 3.1 Preliminaries

The objective in Bayesian inference is to obtain the posterior distributions of the model parameters  $\theta$  and the latent states  $z_{0:T}$ , given the observations  $y_{1:T}$  and a prior distribution  $p(\theta)$ ; i.e.,

$$p(\theta, z_{0:T} | y_{1:T}) = p(y_{1:T} | \theta, z_{0:T}) p(z_{0:T} | \theta) p(\theta) / p(y_{1:T}), \quad (5)$$

where

$$p(y_{1:T}) = \int \int p(y_{1:T} | \theta, z_{0:T}) p(z_{0:T} | \theta) p(\theta) dz_{0:T} d\theta \quad (6)$$

is the marginal likelihood. The likelihood

$$p(y_{1:T} | \theta) = \int p(y_{1:T} | \theta, z_{0:T}) p(z_{0:T} | \theta) p(\theta) dz_{0:T}$$

can be calculated exactly using the Kalman filter for the linear Gaussian state space model (LGSS) and hence for the linear (first order) approximation. The posterior samples can then be obtained using a MCMC sampling scheme. However, the likelihood cannot be computed exactly if a higher-order approximation is used, such as Eq. (S10), for the evolution of the state and other non-linear and non-Gaussian state space models. In this case, the likelihood can only be estimated.

The bootstrap particle filter of Gordon et al. (1993) provides an unbiased estimator of the likelihood for a general state space model. Andrieu and Roberts (2009) and Andrieu et al. (2010) show that it is possible to use this unbiased estimator of the likelihood to carry out full Bayesian inference for the parameters of the general state space model. They call this MCMC approach pseudo marginal Metropolis-Hastings (PMMH).

The non-linear (second-order) DSGE models we consider lie within the class of general non-linear state space models whose state transition density is difficult to work with or cannot be expressed in closed form. In such cases, it is useful to express the model in terms of its latent disturbance variables as they can be expressed in closed form. The posterior in Eq. (5) becomes

$$p(\theta, \epsilon_{1:T} | y_{1:T}) \propto \prod_{t=1}^T p(y_t | \epsilon_{1:t}, \theta) p(\epsilon_t) p(\theta), \quad (7)$$

which Murray et al. (2013) call the disturbance state-space model. The standard state space model can be recovered from the disturbance state space model by using the deterministic function  $F(z_{t-1}, \epsilon_t; \theta) \rightarrow z_t$ . This gives us a state trajectory  $z_{0:T}$  from any sample  $(\epsilon_{1:T}, z_0)$ , where  $z_0 = 0$ . In the disturbance state-space model the target becomes the posterior distribution over the parameters  $\theta$  and the latent noise variables  $\epsilon_{1:T}$ , rather than  $(\theta, z_{0:T})$ .

### 3.2 Standard Pseudo Marginal Metropolis-Hastings

This section outlines the standard PMMH scheme which carries out MCMC on an expanded space using an unbiased estimate of the likelihood. Our paper focuses on estimating the posterior density of the parameters  $\theta$ , but not of the states. Let  $u$  consist of all the random variables required to compute the unbiased likelihood estimate  $\hat{p}_N(y|\theta, u)$ , with  $p(u)$  the density of  $u$ ; let  $p(\theta)$  be the prior of  $\theta$ . The joint posterior density  $\theta$  and  $u$  is

$$p(\theta, u | y_{1:T}) = \hat{p}_N(y_{1:T} | \theta, u) p(\theta) p(u) / p(y_{1:T}),$$

so that

$$p(\theta|y_{1:T}) = \int p(\theta, u|y_{1:T}) du = p(y_{1:T}|\theta) p(\theta) / p(y_{1:T})$$

is the posterior of  $\theta$  and  $\int \widehat{p}_N(y_{1:T}|\theta, u) p(u) du = p(y_{1:T}|\theta)$  because the likelihood estimate is unbiased. We can therefore sample from the posterior density  $p(\theta|y_{1:T})$  by sampling  $\theta$  and  $u$  from  $p(\theta, u|y_{1:T})$ . The subscript  $N$  indicates the number of particles used to estimate likelihood.

Let  $q(\theta'|\theta)$  be the proposal density for  $\theta'$  with the current value  $\theta$  and  $q(u'|u)$  the proposal density for  $u'$  given  $u$ . We always assume that the  $q(u'|u)$  satisfies the reversibility condition

$$q(u'|u) p(u) = q(u|u') p(u'); \quad (8)$$

it is clearly satisfied by the standard PMMH where  $q(u'|u) = p(u')$ . We generate a proposal  $\theta'$  from  $q(\theta'|\theta)$  and  $u'$  from  $q(u'|u)$ , and accept both proposals with probability

$$\begin{aligned} \alpha(\theta, u; \theta', u') &= \min \left( 1, \frac{\widehat{p}_N(y|\theta', u') p(\theta') p(u') q(\theta|\theta') q(u|u')}{\widehat{p}_N(y|\theta, u) p(\theta) p(u) q(\theta'|\theta) q(u'|u)} \right) \\ &= \min \left( 1, \frac{\widehat{p}_N(y|\theta', u') p(\theta') q(\theta|\theta')}{\widehat{p}_N(y|\theta, u) p(\theta) q(\theta'|\theta)} \right). \end{aligned} \quad (9)$$

The expression in Eq. (9) is identical to a standard Metropolis-Hastings algorithm except that estimates of the likelihood at the current and proposed parameters are used. Andrieu and Roberts (2009) show that the resulting PMMH algorithm has the correct invariant distribution regardless of the variance of the estimated likelihoods. However, the performance of the PMMH approach crucially depends on the number of particles  $N$  used to estimate the likelihood. The variance of the log of the estimated likelihood should be between 1 and 3 depending on the quality of the proposal for  $\theta$ ; see Sherlock et al. (2015) and Doucet et al. (2015) for further details. In many applications of the non-linear state space models considered in this paper, it is computationally very expensive to ensure that the variance of the log of the estimated likelihood is within the required range. Section 3.3 discusses the new PMMH sampler, which we call the mixed PMMH algorithm (MPM), that builds on and extends, the CPM of Deligiannidis et al. (2018) and BPM of Tran et al. (2016).

### 3.3 Mixed PMMH (MPM)

This section discusses the proposed mixed PMMH (MPM) method that uses multiple particle filters to obtain the unbiased estimate of the likelihood, similarly to the approach of Sherlock et al. (2017b), and uses a combination of the BPM and CPM methods to induce a high correlation between successive logs of the estimated likelihoods. Suppose that  $G$  particle filters are run in parallel. Let  $\widehat{p}_N(y|\theta, u_g)$  be the unbiased estimate of the likelihood obtained from the  $g$ th particle filter, for  $g = 1, \dots, G$ . We now define the joint target density of  $\theta$  and  $\tilde{u} = (u_1, \dots, u_G)$  as

$$p(\theta, \tilde{u}|y_{1:T}) \propto \bar{p}_N(y|\theta, \tilde{u}) p(\theta) \prod_{g=1}^G p(u_g), \quad (10)$$

where  $\bar{p}_N(y|\theta, u) = \sum_{g=1}^G \widehat{p}_N(y|\theta, u_g) / G$  is the average of the  $G$  unbiased likelihood estimates and hence it is also unbiased (Sherlock et al., 2017b). We update the parameters  $\theta$  jointly with a randomly selected block  $u_g$  in each MCMC iteration, with  $\Pr(G = g) = 1/G$  for any  $g = 1, \dots, G$ . The selected block  $u_g$  is updated using  $u'_g = \rho_u u_g + \sqrt{1 - \rho_u^2} \eta_u$ , where  $\rho_u$  is the non-negative correlation between the random numbers  $u_g$  and  $u'_g$  and  $\eta_u$  is a standard normal vector of the same length as  $u_g$ . Using this scheme, the acceptance probability is

$$\alpha(\theta, \tilde{u}; \theta', \tilde{u}') = \min\left(1, \frac{\bar{p}_N(y|\theta', \tilde{u}' = (u_1, \dots, u_{g-1}, u'_g, u_{g+1}, \dots, u_G)) p(\theta') q(\theta|\theta')}{\bar{p}_N(y|\theta, \tilde{u} = (u_1, \dots, u_{g-1}, u_g, u_{g+1}, \dots, u_G)) p(\theta) q(\theta'|\theta)}\right). \quad (11)$$

It is important that the logs of the likelihood estimates evaluated at the current and proposed values of  $\theta$  and  $\tilde{u}$  are highly correlated to reduce the variance of  $\log \bar{p}_N(y|\theta', \tilde{u}') - \log \bar{p}_N(y|\theta, \tilde{u})$  which helps the Markov chain to mix well. However, the resampling step in the particle filter introduces discontinuities even when  $\theta$  and  $\theta'$  are close, where  $\theta$  is the current value and  $\theta'$  is the proposed value of the parameters. Section 3.5 discusses this discontinuity issue and the proposed correlated disturbance particle filter algorithm, which helps to maintain the correlation. Algorithm S3 in Section S2 of the Supplement gives the MPM algorithm.

### 3.4 Delayed Acceptance PMMH

We propose the delayed acceptance version of the MPM algorithm, which we call delayed acceptance mixed PMMH (DA-MPM), to speed up the computation. The motivation for the delayed acceptance sampler (Christen and Fox, 2005) is to avoid the computation of the expensive likelihood estimate if it is likely that the proposed

draw will ultimately be rejected. A first accept-reject stage uses a cheap (or deterministic) approximation to the likelihood instead of the expensive likelihood estimate in the MH acceptance ratio. Then, the particle filter is used to estimate the likelihood only for a proposal that is accepted in the first stage; a second accept-reject stage ensures that detailed balance is satisfied with respect to the true posterior. We use the likelihood obtained from the central difference Kalman filter (CDKF) proposed by Norgaard et al. (2000) in the first accept-reject stage of the delayed acceptance scheme. We chose the CDKF for the first stage rather than the extended Kalman Filter (EKF) because it frequently outperforms the EKF for general non-linear and non-Gaussian state-space models (Andreasen, 2013). Section S4 of the Supplement gives further details.

### 3.5 The Disturbance Particle Filter

This section discusses the disturbance particle filter we use to obtain the unbiased estimates of the likelihood in the MPM sampler described in Section 3.3. Suppose that  $z_1 = \Phi(z_0, \epsilon_1; \theta)$ , where  $z_0$  is the initial state vector; for  $t \geq 2$ ,  $z_t = F(z_{t-1}, \epsilon_t; \theta)$ , where  $\epsilon_t$  is a  $n_e \times 1$  vector of normally distributed latent noise with density  $p(\epsilon_t)$ . Murray et al. (2013) express the standard state-space model in terms of the latent noise variables  $\epsilon_{1:T}$ , and call

$$\epsilon_t \sim p(\epsilon_t), y_t | \epsilon_{1:t} \sim p(y_t | \epsilon_{1:t}, x_0; \theta) = p(y_t | z_t; \theta), t = 1, \dots, T,$$

the disturbance state-space model. We note that the conditional distribution of  $y_t$  depends on all the latent error variables  $\epsilon_{1:t}$ . The disturbance particle filter provides an unbiased estimate of the likelihood.

For the MPM sampler described in Section 3.3 to work efficiently, the logs of the likelihood estimates evaluated at the current and proposed values of  $(\theta, \tilde{u})$  need to be highly correlated. However, the standard resampling step in the (disturbance) particle filter introduces discontinuities and breaks down the correlation between the logs of the likelihood estimates at the current and proposed values even when the current parameters  $\theta$  and the proposed parameters  $\theta'$  are close. Sorting the particles from smallest to largest helps to preserve the correlation between the logs of the likelihood estimates at the current and proposed values (Deligiannidis et al., 2018). However, such sorting is unavailable for multidimensional state particles. Deligiannidis et al. (2018) use the Hilbert sorting method of Skilling (2004) to order multidimensional state particles. Our article uses a simpler and faster method proposed by Choppala et al. (2016) and given in Section S3 of the Supplement. Algorithm S1 in Section S1

of the Supplement outlines the correlated disturbance particle filter algorithm used by the MPM algorithm described in Section 3.3.

## Improved Disturbance Particle Filter (IDPF)

Our article uses the mixture proposal density

$$m(\epsilon_t | u_{\epsilon,t}, \theta) = \pi p(\epsilon_t | \theta) + (1 - \pi) q(\epsilon_t | \theta, y_{1:T}); \quad (12)$$

$0 \leq \pi \leq 1$ . If  $\pi = 1$ , then  $m(\epsilon_t) = p(\epsilon_t)$  is the bootstrap disturbance particle filter. However, the empirical performance of this bootstrap filter is usually poor because the resulting likelihood estimate is too variable. In practice, we take  $0 < \pi \ll 1$  and  $q(\epsilon_t | \theta, y_{1:T}) = N(\epsilon_t | \hat{\mu}_t, \hat{\Sigma}_t)$ . The methods to obtain  $\hat{\mu}_t$  and the covariance matrix  $\hat{\Sigma}_t$  are now discussed. First, the  $G$  disturbance particle filters are run in parallel. Second, the ancestral tracing algorithm given in Section S5 of the Supplement is used to obtain  $G$  trajectories of  $\epsilon_{g,1:T}$ . Third, the mean  $\hat{\mu}_t$  and the covariance matrix  $\hat{\Sigma}_t$  are set as the mean and the covariance matrix of these  $G$  trajectories of  $\epsilon_{g,1:T}$  at each time  $t$ . Applying additional ancestral tracing step for each disturbance particle filter gives the particle trajectories from the smoothing distribution  $p(\epsilon_{1:T} | \theta, y_{1:T})$ , that is the conditional distribution of disturbances given the parameters  $\theta$  and all available data up to time  $T$ . The proposal defined in Eq. (12) is generated without much additional computational cost because the  $G$  disturbance particle filters are needed to be run at each MCMC iteration to compute the unbiased estimate of the likelihood and the ancestral tracing method is fast to compute. The proposal density in Eq. (12) is in the form of the defensive mixture approach of Hesterberg (1995) which guarantees the boundedness of the weights in the disturbance particle filter algorithm defined in Eq. (S2) of Section S1 of the Supplement, given that the observation density  $p(y_t | z_t, \theta)$  is bounded which is satisfied for all models defined in Section 2.2. We set  $\pi = 0.05$  in all our examples in Section 4.

## 4 Examples

Section 4.1 discusses the inefficiency measures we use to compare the performance of different particle filters or PMMH samplers. Section 4.2 investigates empirically the ability of the proposed mixed PMMH (MPM) sampler to maintain the correlation between successive log-likelihood estimates using a high-dimensional linear Gaussian state space model. Section 4.3 discusses the linear small-scale DSGE example used by Herbst and Schorfheide (2019). Section 4.4 discusses a non-linear Real Business

Cycle (RBC) using simulated datasets. Section 4.5 applies the MPM sampler to estimate a non-linear small scale DSGE model.

## 4.1 Definitions of Inefficiency

We define the time normalised variance (TNV) of a particle filter method

$$\text{TNV}_{PF} := \widehat{V}(\log \widehat{p}(y|\theta)) \times \text{CT},$$

as the measure of inefficiency of the method that takes computing time into account; CT is the computing time to obtain a single log of the estimated likelihood in seconds, and  $\widehat{V}(\log \widehat{p}(y|\theta))$  is the estimated variance of the log of the likelihood estimate. The relative time normalised variance (RTNV) of a particle filter method is defined as  $\text{RTNV}_{PF} := \text{TNV}_{PF} / \text{TNV}_{IDPF}$ .

We use the inefficiency factor (IF)

$$\text{IF}_{\psi} := 1 + 2 \sum_{t=1}^{\infty} \rho_{\psi}(t),$$

to measure the inefficiency of a PMMH sampler at estimating the posterior expectation of a univariate function  $\psi(\theta)$  of  $\theta$ ; here,  $\rho_{\psi}(t)$  is the  $j$ th autocorrelation of the iterates  $\psi(\theta)$  in the MCMC chain after it has converged to its stationary distribution. We estimate the IACT using the CODA R package of Plummer et al. (2006). A low value of the IACT estimate suggests that the Markov chain mixes well. Our measure of the inefficiency of a PMMH sampler that takes computing time into account for a given parameter  $\theta$  based on  $\text{IF}_{\psi}$  is the time normalised inefficiency factor (TNIF) is defined as  $\text{TNIF}_{\psi} := \text{IF}_{\psi} \times \text{CT}$ . The estimate of TNIF is the estimate of the IF times the computing time (CT). For a given sampler, let  $\text{IACT}_{\text{MAX}}$  and  $\text{IACT}_{\text{MEAN}}$  be the maximum and mean of the IACT values over all the parameters in the model.

## 4.2 Performance for the Linear Gaussian State Space Model

This section examines empirically the ability of the following methods to maintain the correlation between successive values of the log of the estimated likelihood for a linear Gaussian state space model: (1) the block PMMH (BPM) of Tran et al. (2016), (2) the correlated PMMH (CPM) of Deligiannidis et al. (2018), and (3) the mixed PMMH (MPM).

We consider the model discussed in Guarniero et al. (2017) and Deligiannidis

et al. (2018), where  $\{X_t; t \geq 1\}$  and  $\{Y_t; t \geq 1\}$  are  $\mathcal{R}^d$  valued with

$$\begin{aligned} Y_t &= X_t + W_t, \\ X_{t+1} &= A_\theta X_t + V_{t+1}, \end{aligned}$$

with  $X_1 \sim \mathcal{N}(0_d, I_d)$ ,  $V_t \sim N(0_d, I_d)$ ,  $W_t \sim N(0_d, I_d)$ , and  $A_\theta^{i,j} = \theta^{|i-j|+1}$ ; the true value of  $\theta$  is 0.4. We use the state based bootstrap filter so that the proposal density is the state transition density of the particle filter for all methods. The experimental setup is now discussed. The simulated data is generated from the model above with  $T = 100$  and  $T = 1000$  time periods and  $d = 5, 10, 20, 50, 100$  dimensions. The correlation parameter in the CPM is set to  $\rho_u = (0.5, 0.75, 0.8, 0.9, 0.99)$  and the number of blocks in the BPM is set to  $G = (2, 4, 5, 10, 100)$ . We use the same number of blocks for the MPM as in the BPM and set the correlation coefficients  $\rho_u = 0$  and  $\rho_u = 0.99$ .

There are important implementation differences between the BPM of Tran et al. (2016), the CPM of Deligiannidis et al. (2018) and the proposed MPM method. The BPM updates a block of random numbers for all time periods in the particle filter algorithm and runs a single particle filter; e.g. suppose  $N = 100$  particles, one block of the particles is updated for all time periods, with the other 99 block of particles unchanged for all time periods. The CPM method correlates all the random numbers used in the particle filter using  $u' = \rho_u u + \sqrt{1 - \rho_u^2} \eta_u$ , where  $\rho_u$  is the non-negative correlation between the random numbers  $u$  and  $u'$  and  $\eta_u$  is a standard normal vector of the same length as  $u$ , and runs a single particle filter. The MPM method updates the random numbers in one particle filter, keeping the random numbers in the other  $G - 1$  particles filters unchanged. The selected block  $u_g$  is updated using the correlated approach  $u'_g = \rho_u u_g + \sqrt{1 - \rho_u^2} \eta_u$ . Given those random numbers, the MPM then runs  $G$  particle filters in parallel.

We ran the CPM, BPM, and MPM approaches for 100 iterations holding the parameter  $\theta$  fixed at the true value. At each iteration we generated the  $u^{(s)}$  and  $u^{(s)'}$  and obtained  $\log \hat{p}_N(y|\theta, u^{(s)})$  and  $\log \hat{p}_N(y|\theta, u^{(s)'})$ , for  $s = 1, \dots, 100$ , for the BPM and CPM; we then computed their sample correlation. Similarly, we generated  $\tilde{u}$ , where  $\tilde{u} = (u_1, \dots, u_G)$  and  $\tilde{u}'$  and obtained  $\log \bar{p}_N(y|\theta, \tilde{u}^{(s)})$  and  $\log \bar{p}_N(y|\theta', \tilde{u}^{(s)'})$  for the MPM approach and computed their sample correlation.

Figures 1 and 2 report the correlation estimates of  $\log \hat{p}_N(y|\theta, u)$  and  $\log \hat{p}_N(y|\theta, u')$  for the BPM and CPM approaches and  $\log \bar{p}_N(y|\theta, \tilde{u}^{(s)})$  and  $\log \bar{p}_N(y|\theta', \tilde{u}^{(s)'})$  for the MPM approach for  $T = 100$  and 1000, respectively. The figures show that: (1) the correlation estimates for the BPM approach are close to zero for all the reported number of blocks and dimensions, supporting the observation made by Tran et al.

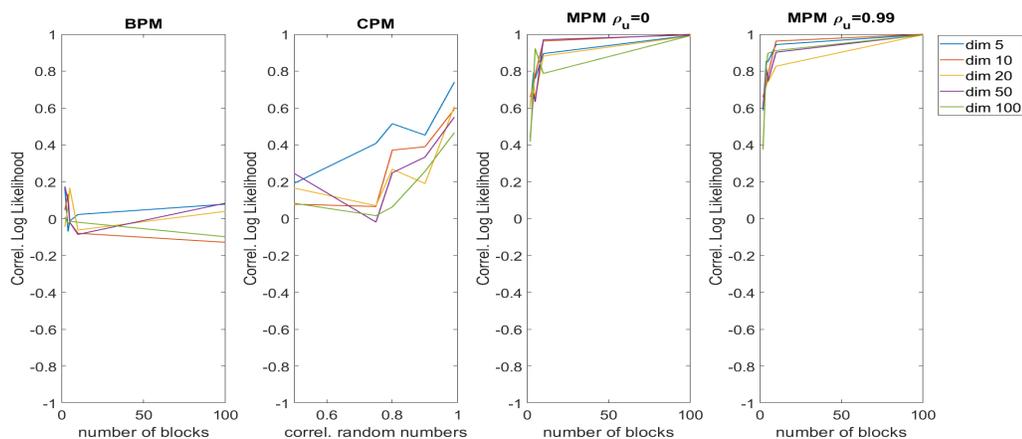
(2016) that the BPM approach is unsuitable for time series state space models; (2) when the correlation between random numbers  $u$  and  $u'$  is set to 0.99 in the CPM approach, the estimated correlations between the log of the estimated likelihoods are 0.73, 0.59, 0.59, 0.55, and 0.47 for  $T = 100$  and 0.66, 0.51, 0.52, 0.51, and 0.53 for  $T = 1000$  for  $d = 5, 10, 20, 50, 100$ , respectively. The estimated correlations seem to be slightly worse when  $T = 1000$ . The higher the dimension of the latent state variables, the lower is the estimated correlation between the log of estimated likelihoods. Section 4.4 shows that the estimated correlation between the log of the estimated likelihoods gets worse for the non-linear model; (3) the MPM method maintains a high correlation between the logs of the estimated likelihoods even for  $d = 100$  dimensions and  $T = 1000$  time periods. The estimated correlations between logs of the estimated likelihoods when the number of blocks  $G$  is 100 and  $d = 100$  dimensions are 0.99 for both  $T = 100$  and 1000. This supports the result in Tran et al. (2016) that shows that the correlation of the logs of the estimated likelihood at the current and proposed values is close to  $1 - 1/G = 1 - 1/100 = 0.99$ .

In summary, this example suggests that: (1) the BPM of Tran et al. (2016) is ineffective for time series state space models because it is unable to maintain the correlation between the logs of the estimated likelihoods in successive iterates; (2) the correlation estimates between the logs of the estimated likelihoods obtained from the CPM deteriorate as the dimension of the state variables and the length of time series increase; (3) the MPM approach maintains a high correlation even when the dimension of the state variables is 100. There is no noticeable difference between setting  $\rho_u = 0$  or 0.99 for the MPM approach in this example. It is therefore sufficient to set the correlation  $\rho_u = 0$ .

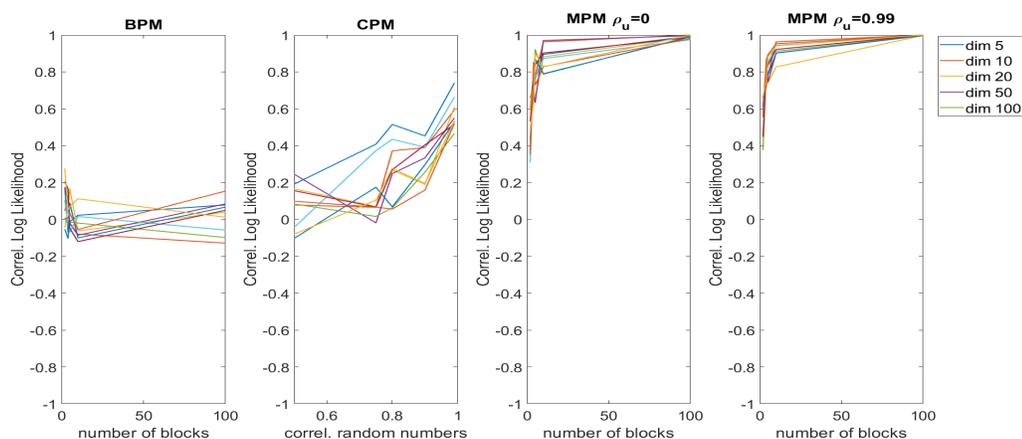
### 4.3 Linear Small Scale DSGE Model

This section investigates empirically the performance of the proposed improved disturbance particle filter (IDPF), the bootstrap particle filter (BPF), and the tempered particle filter (TPF) of Herbst and Schorfheide (2019) based on the variance of the log of the estimated likelihood for each estimator. Thus, an estimator is better than a second estimator if the variance of the log of the first estimator is smaller than that of the second. We compare the variances using the linear small scale DSGE model considered by Herbst and Schorfheide (2019), and described in Section S7 of the Supplement. A linearised DSGE model (with normally distributed innovations) leads to a linear Gaussian state-space representation, making it feasible to compute the likelihood using the Kalman filter instead of estimating the likelihood by the particle filter. The likelihood estimated using Kalman filter is treated as the ground

**Figure 1:** The estimated correlation of log of the estimated likelihoods obtained using the four approaches: (1) Block PMMH (BPM), (2) Correlated PMMH (CPM), (3) Mixed PMMH (MPM) with  $\rho_u = 0$ , and (4) Mixed PMMH (MPM) with  $\rho_u = 0.99$  for  $T = 100$  and  $d = 5, 10, 20, 50, 100$  dimensions



**Figure 2:** The estimated correlation of log of the estimated likelihoods obtained using the four approaches: (1) Block PMMH (BPM), (2) Correlated PMMH (CPM), (3) Mixed PMMH (MPM) with  $\rho_u = 0$ , and (4) Mixed PMMH (MPM) with  $\rho_u = 0.99$  for  $T = 1000$  and  $d = 5, 10, 20, 50, 100$  dimensions



truth for comparing the accuracy of different particle filter methods.

The small scale DSGE model has three observables (output growth, inflation, and the federal funds rate). Herbst and Schorfheide (2019) augment the observation equations by independent measurement errors  $\eta_t$ , whose standard deviations are fixed at 20% of the standard deviations of the observables  $y_t$ . We augment similarly, but set the standard deviations at 10% of the standard deviation of the observables, making the estimation more challenging as the signal to noise is higher. The real dataset used in this section is obtained from Herbst and Schorfheide (2019), using data from 1983Q1 to 2002Q4, a total of 80 observations for each series<sup>1</sup>.

The variance of the log of the likelihood estimate is obtained by computing the sample variance of the log of the estimated likelihood for each particle filter method using  $N_{run} = 100$  independent runs holding the parameters fixed at the values given in Table S1 in Section S8 of the Supplement. The fixed parameter values are those used by Herbst and Schorfheide (2019). The proposal for the IDPF is estimated from the collection of particles generated by the  $G$  particle filters running in parallel. These particle filters are run once and the same proposal is kept for different numbers of particles  $N$ . The tuning parameters of the TPF are set to the default values given in Herbst and Schorfheide (2019). All the particle filters are implemented in Matlab. The TPF Matlab code is obtained from the website<sup>2</sup> of Frank Schorfheide.

Table 1 reports the variance of the log of the estimated likelihoods using the BPF, TPF, and IDPF for the linear small scale DSGE model with  $T = 80$  time periods and  $N = 10, 20, 50, 100, 250, 500, 1000,$  and  $2000$  particles evaluated at the parameter values given in Table S1 of Section S8 of the Supplement. Section 3.5 shows that there is no additional computational cost in constructing the IDPF proposal because it is obtained from the collection of particles generated by the  $G$  particle filters running in parallel at each MCMC iteration, but the IDPF computing time in Table 1 is set to twice of the computing time to obtain a single estimate of the likelihood to account for the time spent constructing the proposal.

In this example, the exact log-likelihood obtained from the Kalman filter is  $-296.45$ . In general, the IDPF is much more efficient than BPF and TPF for all  $N$ . For example, the IDPF is 350 times more efficient than BPF and 2855 times more efficient than TPF for  $N = 500$ . The log of the likelihood estimate obtained from the IDPF is very close to the exact value even with  $N = 20$ . The BPF fails to provide reasonable values of the of the likelihood estimate even with  $N = 2000$  and the TPF requires  $N = 500$  particles to obtain the value of the log of the likelihood estimate

---

<sup>1</sup>The standard deviations of the measurement errors are 0.057 for output growth, 0.147 for inflation, and 0.223 for the interest rates. The DSGE model in this section is solved using the GENSYS software.

<sup>2</sup><https://web.sas.upenn.edu/schorf/publications/>

that is close to the exact value. In addition, in Table 1, we note that for all three estimators  $\log \widehat{p}_N(y|\theta, u)$  increases as  $N$  increases and seems to converge to the log of the likelihood obtained by the Kalman filter. This can be explained heuristically as follows. Suppose that  $E(X) = a$ ,  $\sigma_X^2 = \text{var}(X)$  and  $|X - a|/|a|$  is small. Then

$$E(\log X) \approx \log a - \frac{\sigma_X^2}{a^2} \leq \log a$$

and  $\log X$  monotonically increases to  $\log a$  as  $\sigma_X^2$  tends to 0.

## 4.4 Non-Linear Real Business Cycle Models

This section reports the results of three studies using data generated from the non-linear Real Business Cycle (RBC) model described in Section S10 of the Supplement. The true values of the parameters are set to  $\beta = 0.95$ ,  $\delta = 0.025$ ,  $\alpha = 0.90$ ,  $\sigma = 2$ ,  $\rho_a = \rho_e = 0.8$ ,  $\sigma_e = \sigma_a = 0.5$ . The measurement error covariance matrix is set to

$$\Sigma_\eta = \begin{bmatrix} 0.0135 & 0 \\ 0 & 0.0137 \end{bmatrix},$$

which corresponds to a large signal to noise ratio (SNR). We simulated three data series of  $T = 50, 100$ , and  $200$  observations. These numbers of observations are chosen because they are roughly the same size as standard ( $T = 50$  and  $100$ ) and large ( $T = 200$ ) macroeconomic time series. The particle filters and the parameter samplers are implemented in Matlab. All the computations are done on a single desktop computer with 6-CPU cores.

The **first** study compares variance of the log of the estimated likelihoods for the improved disturbance particle filter (IDPF), the block particle filter (BPF) and the tempered particle filter (TPF) using  $N_{run} = 100$  independent runs holding the parameters fixed at the true values. The proposal for the IDPF is estimated from the collection of particles generated by the  $G$  particle filters running in parallel. These particle filters are run once and the same proposal is kept for different numbers of particles  $N$ . The tuning parameters of the TPF are set to the default values given in Herbst and Schorfheide (2019).

Table 2 reports the variance of the log of the estimated likelihood for the BPF, TPF, and IDPF methods for the non-linear RBC model with  $T = 50, 100, 200$  time periods and  $N = 100, 250, 500, 1000$ , and  $2000$  particles. In general, the IDPF is much more efficient than BPF and TPF for all  $T$  and  $N$ , for the non-linear RBC model. For example, the IDPF is 5373 times more efficient than BPF and 294020 times more efficient than TPF when the number of particles  $N = 500$  and  $T = 200$

**Table 1:** The variance of the log of the estimated likelihoods obtained using various particle filter methods: (1) the bootstrap particle filter (BPF), (2) the tempered particle filter (TPF), and (3) the improved disturbance particle filter (IDPF) for the small scale linear DSGE model with  $T = 80$  periods and  $N$  is the of particles. The computation time (CT) in seconds to obtain a single log of the estimated likelihood is in seconds. The  $\log \tilde{p}_N(y|\theta, u)$  is the average of the log of the estimated likelihoods over 100 replications. The TNV and RTNV are the time normalised variance and the relative time normalised variance of a particle filter method relative to the IDPF method, respectively, and are defined in Section 4.1.

N	BPF				TPF				IDPF						
	Var	CT	TNV	RTNV	$\log \tilde{p}_N(y \theta, u)$	Var	CT	TNV	RTNV	$\log \tilde{p}_N(y \theta, u)$	Var	CT	TNV	RTNV	$\log \tilde{p}_N(y \theta, u)$
10	373665.65	0.01	3736.60	54.04	-2855.24	14335.67	9.60	137622.43	1990.20	-594.7452	1728.84	0.04	69.15	1	-301.58
20	131366.07	0.01	1313.66	57.24	-2016.73	1346.30	13.31	17919.25	780.80	-388.1918	573.66	0.04	22.95	1	-298.93
50	30537.59	0.01	305.37	136.94	-1113.85	213.04	18.74	3992.37	1790.30	-318.1254	55.77	0.04	2.23	1	-297.82
100	20879.66	0.02	417.59	707.78	-729.90	97.64	21.51	2100.24	3559.73	-302.4716	14.68	0.04	0.59	1	-296.64
250	7246.88	0.02	144.94	630.17	-444.62	29.99	29.04	870.91	3786.57	-302.4189	5.66	0.04	0.23	1	-296.28
500	3682.07	0.02	73.64	350.67	-420.32	15.41	38.91	599.60	2855.24	-296.7779	2.57	0.08	0.21	1	-296.79
1000	2610.05	0.03	78.30	652.50	-368.68	6.98	56.14	391.86	3265.50	-296.5825	1.22	0.10	0.12	1	-296.48
2000	1558.63	0.05	77.93	519.53	-328.80	3.78	96.09	363.22	2421.47	-297.0468	0.75	0.20	0.15	1	-296.32

time periods.

The **second** study compares the ability of the CPM of Deligiannidis et al. (2018), and the MPM methods to maintain the correlation between successive log-likelihood estimates for the non-linear RBC model; the TPF is not used in the CPM method because it is not straightforward to correlate all the random numbers, especially with the random walk Metropolis-Hastings mutation steps and the adaptive tempering iterations. The correlation parameter in the CPM approach is set to  $\rho_u = (0.5, 0.75, 0.8, 0.9, 0.99, 0.9999, 0.999999)$ . The number of blocks in the MPM approach is set to  $G = (2, 4, 5, 10, 100)$  and the correlation coefficient  $\rho_u$  is set to 0.99. Two different sorting methods are considered. The first sorts the state particles and the second sorts the disturbance particles. We ran the CPM and the MPM approaches for  $N_{run} = 100$  iterations holding the parameters fixed at the true values.

Figure 3 reports the correlation estimates between successive values of the log of the estimated likelihood for the CPM method with disturbance sorting and state sorting. It shows that: (a) the correlation estimates are quite low even when the correlation between the random numbers are set to be very close to 1; (b) the correlation estimates with the states sorted are 0.49, 0.28, 0.29 when the correlation between random numbers is set to 0.999999, for  $T = 50, 100, 200$ , respectively; (c) sorting the state particles gives similar results to sorting the disturbance particles.

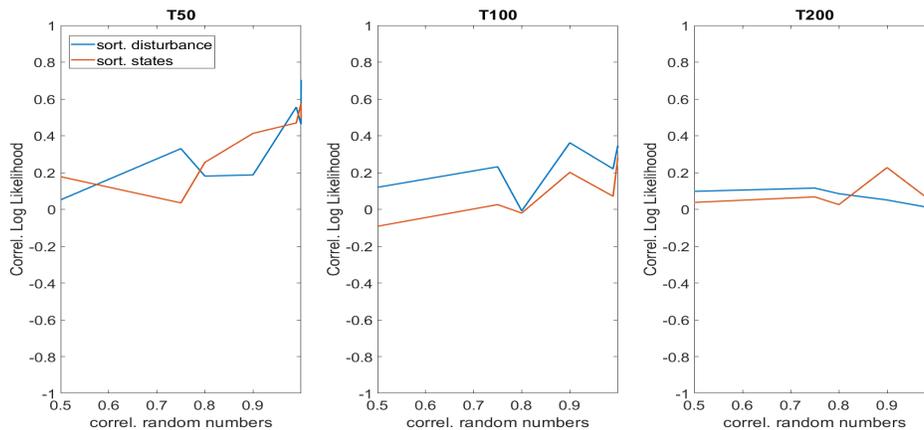
Figure 4 reports the correlation estimates between the log of the estimated likelihoods for four different methods: (i) the MPM with the BPF and disturbance sorting, (ii) the MPM with the IDPF and disturbance sorting, (iii) the MPM with the BPF and state sorting, and (iv) the MPM with the IDPF and state sorting. The figure shows that: (a) the correlation estimates are 0.98, 0.99, 0.98 when the number of blocks is set to 100, for  $T = 50, 100, 200$ , respectively, when sorting the state particles and the IDPF are used. This suggests that the proposed MPM sampler is able to maintain high correlations between the logs of the estimated likelihoods in scenarios where the other methods fail. (b) Sorting the disturbances gives bigger correlation estimates than sorting the states, especially when  $T = 200$ . This is because the dimension of the disturbances is smaller than that of the states.

The **third** study compares the performance of the following PMMH samplers in estimating the parameters of the non-linear RBC model for  $T = 200$ : (i) The delayed acceptance mixed PMMH (DA-MPM) with the improved disturbance particle filter (IDPF), disturbance sorting, and the correlation between random numbers used in estimating the log of estimated likelihoods set to  $\rho_u = 0.99$ ; (ii) the DA-MPM with the IDPF, state sorting, and the  $\rho_u = 0.99$ ; (iii) the DA-MPM with the bootstrap particle filter (BPF), disturbance sorting, and the  $\rho_u = 0.99$ ; (iv) the DA-MPM with the IDPF, disturbance sorting, and  $\rho_u = 0$ , (v) the MPM with the IDPF, disturbance

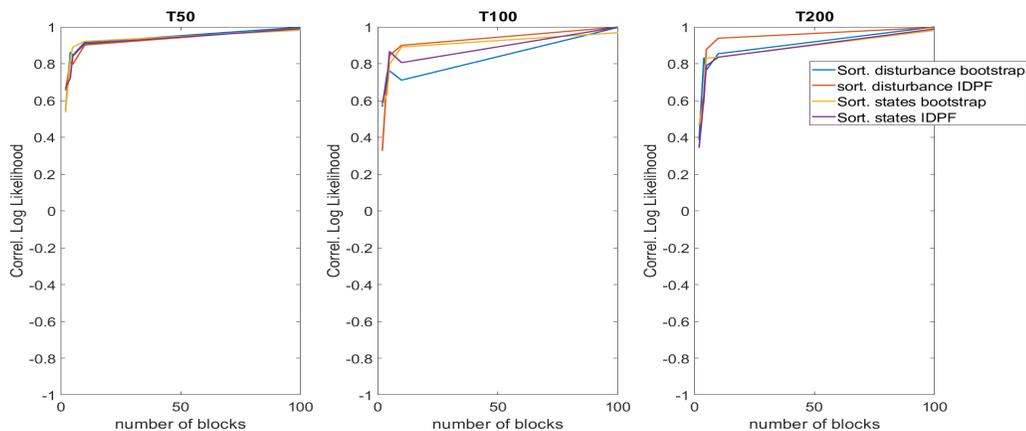
**Table 2:** The variance of the log of the estimated likelihood obtained using three particle filter methods: (1) the bootstrap particle filter (BPF), (2) the tempered particle filter (TPF), and (3) the improved disturbance particle filter (IDPF) for the non-linear RBC model with  $T = 50, 100, 100,$  and  $200$  time periods and  $N = 100, 250, 500, 1000, 2000$  particles. We also report the time normalised variance (TNV), the relative time normalised variance (RTNV), and the computation time (CT) per iteration in seconds as defined in Section 4.1.

$T$	$N$	BF					TPF					IDPF					
		Var	CT	TNV	RTNV	Var	CT	TNV	RTNV	Var	CT	TNV	RTNV	Var	CT	TNV	RTNV
50	100	36.81	0.015	0.552	17.25	9.23	3.171	29.268	914.625	0.85	0.038	0.032	1				
	250	9.60	0.017	0.163	11.64	3.64	8.082	29.418	2101.290	0.33	0.044	0.014	1				
	500	4.69	0.029	0.136	11.33	1.60	15.843	25.349	2112.41	0.16	0.078	0.012	1				
	1000	2.22	0.050	0.111	11.10	0.84	31.532	26.487	2648.70	0.08	0.136	0.010	1				
	2000	1.03	0.080	0.082	10.25	0.42	63.942	26.856	3357.00	0.04	0.196	0.008	1				
100	100	1424.20	0.029	41.302	0.83	421.48	8.701	3667.297	73.48	779.84	0.064	49.910	1				
	250	760.33	0.030	22.810	3.29	375.70	20.346	7643.992	1101.44	86.75	0.080	6.940	1				
	500	596.48	0.052	31.017	22.22	214.52	37.74	8095.984	5799.41	10.27	0.128	1.396	1				
	1000	265.98	0.087	23.140	71.42	89.02	76.947	6849.822	21141.42	1.36	0.238	0.324	1				
	2000	50.54	0.149	7.531	36.21	16.28	146.671	2387.804	11489.83	0.52	0.398	0.208	1				
200	100	89615.69	0.035	3136.549	2.96	18178.00	19.732	358688.300	338.56	9459.50	0.112	1059.460	1				
	250	41655.73	0.053	2207.754	1760.57	9704.34	37.996	368726.100	294039.95	8.25	0.152	1.254	1				
	500	26864.78	0.098	2632.748	5372.95	2088.47	73.720	153962.010	314208.18	1.91	0.256	0.490	1				
	1000	6590.42	0.175	1153.324	2270.32	29.47	147.365	4342.850	8548.92	1.08	0.470	0.508	1				
	2000	919.93	0.296	272.299	677.36	9.49	294.720	2796.890	6957.44	0.52	0.774	0.402	1				

**Figure 3:** The estimated correlation of log of the estimated likelihoods  $(\log \hat{p}_N(y|\theta, u), \log \hat{p}_N(y|\theta, u'))$  obtained using the two approaches: (1) CPM with disturbance sorting, (2) CPM with state sorting for  $T = 50, 100, 200$



**Figure 4:** The estimated correlation of the log of the estimated likelihoods,  $\log \tilde{p}_N^{(s)}(y|\theta, \tilde{u})$  and  $\log \tilde{p}_N^{(s)}(y|\theta', \tilde{u}')$  obtained using 4 approaches: (1) MPM with the BPF and disturbance sorting, (2) MPM with the IDPF and disturbance sorting, (3) MPM with the BPF and state sorting, and (4) MPM with the IDPF and state sorting, for  $T = 50, 100, 200$ .



sorting, and the  $\rho_u = 0.99$ ; and (vi) the delayed acceptance correlated PMMH (DA-CPM) with  $\rho_u = 0.99$  and the BPF. The samplers ran for 25000 iterations, with the initial 5000 iterations discarded as burn-in. All the samplers are initialized at the posterior means from a second order approximation of the model obtained from the central difference Kalman Filter (CDKF) proposed by Norgaard et al. (2000). We chose this initialisation method because the variance of the log of the estimated likelihood increased significantly in some areas of the support of  $\theta$  away from the true values, making it difficult for the MCMC sampler to converge. We use the adaptive random walk proposal of Roberts and Rosenthal (2009) for  $q(\theta'|\theta)$  and use the adaptive scaling approach of Garthwaite et al. (2015) to tune the Metropolis-Hastings acceptance probability to 20%. The covariance matrix of the proposals is fixed to the estimates from the preliminary run of the MCMC. We fix  $\beta$ ,  $\delta$ ,  $\alpha$ , and  $\sigma$  and estimate the four parameters  $\rho_a, \rho_e, \sigma_e, \sigma_a$ . This gives a smaller number of parameters so that we can focus on the comparisons of different PMMH samplers. All the samplers ran on a single desktop computer with 6-CPU cores.

Table 3 reports the inefficiency factors for each parameter and the relative time normalised inefficiency factor (RTNIV) of a PMMH sampler relative to the DA-MPM with the IDPF, disturbance sorting, and  $\rho_u = 0.99$  for the non-linear RBC models with  $T = 200$  time periods. The computing time reported in the table is the time to run a single particle filter for the CPM and  $G$  particle filters for the MPM approach using a single desktop computer with 6 cores. The table shows that: (a) the DA-MPM sampler with the IDPF, disturbance sorting,  $\rho_u = 0.99$ , and  $N = 100$  particles is much more efficient than the CPM with 2000 particles in terms of  $\widehat{\text{RTNIF}}_{\text{MAX}}$  and  $\widehat{\text{RTNIF}}_{\text{MEAN}}$ ; (b) the time taken for running the MPM method with  $G = 100$  particles filters each with  $N = 100$  particles is similar to the CPM with  $N = 1000$  particles. The MPM method can be much faster than the CPM method if it was run using high-performance computing with a large number of cores. This also shows that the multidimensional sorting algorithm is quite expensive for a large number of particles and long time periods; (c) the MPM sampler allows us to use of a much smaller number of particles for each independent PF and these multiple PFs can be run in parallel; (d) the performance of DA-MPM with IDPF, disturbance sorting, and  $\rho_u = 0.99$  is 4.49 and 2.66 times more efficient than the DA-MPM method with state sorting method when  $N = 250$  and it has similar performance to the DA-MPM approach with IDPF, disturbance sorting, and  $\rho_u = 0$ ; (e) the delayed acceptance MPM is much more efficient than the standard MPM approach. The delayed acceptance algorithm is 5 times faster on average because the target Metropolis-Hastings acceptance probability is set to 20% using the Garthwaite et al. (2015) approach.

**Table 3:** Real Business Cycle (RBC) models using simulated dataset with  $T = 200$ . Sampler I: DA-MPM ( $\rho_u = 0.99$ , disturbance sorting, and IDPF), Sampler II: DA-MPM ( $\rho_u = 0.99$ , state sorting, and IDPF), Sampler III: DA-MPM ( $\rho_u = 0.99$ , disturbance sorting, and BPF), Sampler IV: DA-MPM ( $\rho_u = 0$ , disturbance sorting, and IDPF), Sampler V: MPM ( $\rho_u = 0.99$ , disturbance sorting, and IDPF), and Sampler VI: DA-CPM ( $\rho_u = 0.99$  and BPF). The table reports the time normalised inefficiency factor (TNIV), the relative time normalised inefficiency factor (RTNIV), and the computation time (CT) per iteration (in seconds) as defined in Section 4.1. The symbol NA means the Markov chain gets stuck and does not converge.

Param	I		II		III		IV		V			VI		
	100	250	100	250	250	100	250	100	250	100	250	500	1000	2000
$N$	53.81	25.02	150.31	108.08	NA	60.81	27.69	112.66	15.34	2495.09	439.28	NA	2495.09	439.28
$\rho_a$	53.40	26.40	122.10	39.62	NA	43.70	24.50	39.12	15.88	5040.98	874.69	NA	5040.98	874.69
$\rho_e$	72.82	24.07	116.61	34.45	NA	65.98	27.64	38.68	15.44	4790.05	544.69	NA	4790.05	544.69
$\sigma_e$	71.87	31.02	107.30	37.85	NA	40.32	31.51	34.91	17.65	4901.89	943.21	NA	4901.89	943.21
$\widehat{\text{IACT}}_{\text{MAX}}$	72.82	31.03	150.31	108.08	NA	65.99	31.51	112.66	17.65	5040.98	943.22	NA	5040.98	943.22
$\widehat{\text{TNIF}}_{\text{MAX}}$	327.69	347.85	858.27	1560.68	NA	296.95	353.23	2534.85	989.28	22029.08	13535.21	NA	22029.08	13535.21
$\widehat{\text{RTNIF}}_{\text{MAX}}$	0.94	1	2.47	4.49	NA	0.85	1.02	7.29	2.84	63.33	38.91	NA	63.33	38.91
$\widehat{\text{IACT}}_{\text{MEAN}}$	62.98	26.63	124.08	55.00	NA	52.71	27.84	56.34	16.08	4307.00	700.47	NA	4307.00	700.47
$\widehat{\text{TNIF}}_{\text{MEAN}}$	283.41	298.52	1791.72	794.20	NA	237.19	312.09	1267.65	901.28	18821.59	10051.74	NA	18821.59	10051.74
$\widehat{\text{RTNIF}}_{\text{MEAN}}$	0.95	1	6.00	2.66	NA	0.79	1.05	4.25	3.02	63.05	33.67	NA	63.05	33.67
CT	4.50	11.21	5.71	14.44	13.04	4.50	11.21	22.5	56.05	1.54	4.37	14.35	4.37	14.35

## 4.5 Nonlinear Small Scale DSGE Model

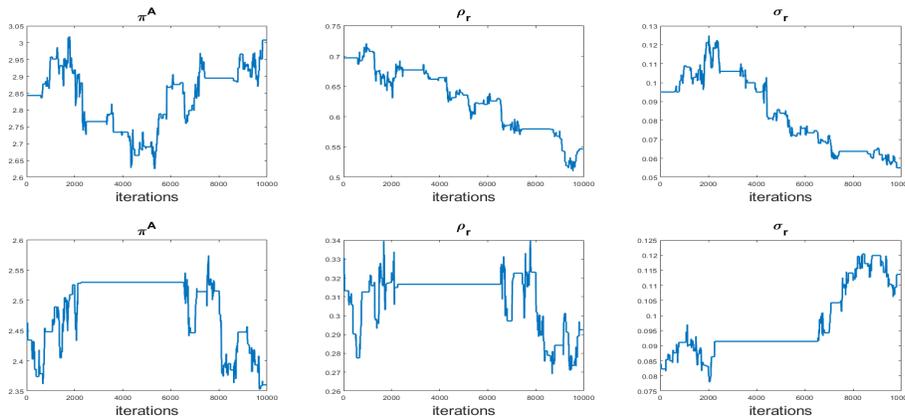
This section reports on how well a number of PMMH samplers of interest estimate the parameters of the nonlinear small scale DSGE model described in Section S7 of the Supplement, using the same data as in Section 4.3. The PMMH samplers considered are: (a) the MPM sampler with the IDPF, disturbance sorting, and  $\rho_u = 0.99$ ; (b) the MPM sampler with the IDPF, disturbance sorting, and  $\rho_u = 0$ ; and (c) the CPM sampler with the BPF, disturbance sorting, and  $\rho_u = 0.99$ . Each sampler ran for 55000 iterations, with the initial 5000 iterations discarded as burn-in. All the samplers are initialized at the posterior means from a second order approximation of the model obtained from the central difference Kalman Filter (CDKF) proposed by Norgaard et al. (2000). We use the adaptive random walk proposal of Roberts and Rosenthal (2009) for  $q(\theta'|\theta)$  and the adaptive scaling approach of Garthwaite et al. (2015) to tune the Metropolis-Hastings acceptance probability to 20%. The covariance matrix of the proposals is fixed to the estimates from the preliminary run of the MCMC. Section S7 of the supplement gives details on model specifications and model parameters.

Table 4 reports the inefficiency factors for each parameter and the relative time normalised inefficiency factor (RTNIV) of a sampler relative to the DA-MPM with the IDPF, disturbance sorting, and  $\rho_u = 0.99$  with  $T = 80$  time periods. The computing time reported in the table is the time to run a single particle filter for the CPM and  $G$  particle filters for the MPM approach using a single desktop computer with 6 cores. The table shows that: (a) the DA-MPM sampler with the IDPF, disturbance sorting,  $\rho_u = 0.99$ , and  $N = 250$  particles has similar performance to the DA-MPM approach with IDPF, disturbance sorting, and  $\rho_u = 0$ ; (b) the time taken to run MPM with  $G = 100$  particles filters each with  $N = 250$  particles is similar to the CPM with  $N = 2000$  particles. The MPM method can be much faster than the CPM method if it runs using high-performance computing with a large number of cores. In addition, the multidimensional sorting algorithm is quite expensive for a large number of particles; (c) Figure 5 shows the trace plot of some of the parameters for the last 10000 iterations estimated using the CPM method with  $N = 2000, 3000$  particles. The CPM still gets stuck and does not converge even with  $N = 3000$  particles. Section S9 of the supplement provides further analysis comparing the linear and non-linear small scale DSGE models.

**Table 4:** Small scale DSGE model using a real dataset with  $T = 80$ . Sampler I: DA-MPM ( $\rho_u = 0.99$ , disturbance sorting, and IDPF), Sampler II: DA-MPM ( $\rho_u = 0$ , disturbance sorting, and IDPF), Sampler III: DA-CPM ( $\rho_u = 0.99$ , disturbance sorting, and BPF). We also report the time normalised inefficiency factor (TNIV), the relative time normalised inefficiency factor (RTNIV), and the computation time (CT) per iteration in second, defined in Section 4.1. The symbol NA means the Markov chain got stuck and did not converge.  $\sigma_r^m$ ,  $\sigma_g^m$ , and  $\sigma_m^m$  are the measurement error variances.

Param	I	II	III	
$N$	250	250	2000	3000
$\pi^{(A)}$	139.24	157.92	NA	NA
$\tau$	206.79	244.24	NA	NA
$\psi_1$	169.58	169.89	NA	NA
$\psi_2$	269.15	214.37	NA	NA
$\gamma^{(Q)}$	158.96	148.30	NA	NA
$r^{(A)}$	171.73	156.14	NA	NA
$\rho_r$	385.22	275.34	NA	NA
$\rho_g$	203.93	239.14	NA	NA
$\rho_m$	334.20	516.61	NA	NA
$\sigma_r$	292.06	206.95	NA	NA
$\sigma_g$	318.61	433.34	NA	NA
$\sigma_m$	177.47	172.45	NA	NA
$\sigma_r^m$	193.66	202.54	NA	NA
$\sigma_g^m$	190.19	197.82	NA	NA
$\sigma_m^m$	187.61	222.94	NA	NA
$\widehat{\text{IACT}}_{\text{MAX}}$	385.22	516.61	NA	NA
$\widehat{\text{TNIF}}_{\text{MAX}}$	2037.81	2732.87	NA	NA
$\widehat{\text{RTNIF}}_{\text{MAX}}$	1	1.34	NA	NA
$\widehat{\text{IACT}}_{\text{MEAN}}$	226.56	237.20	NA	NA
$\widehat{\text{TNIF}}_{\text{MEAN}}$	1198.50	1254.79	NA	NA
$\widehat{\text{RTNIF}}_{\text{MEAN}}$	1	1.05	NA	NA
Time	5.29	5.29	6.25	14.04

**Figure 5:** The trace plots of some of the parameters of the non-linear (second order) small scale DSGE models estimated using DA-CPM with disturbance sorting,  $\rho_u = 0.99$ ,  $N = 2000$  particles (top) and  $N = 3000$  particles (bottom).



## 5 Summary and conclusions

Our article proposes a general particle MCMC approach (MPM) for estimating the posterior density of the parameters of complex and high-dimensional state-space models. It is especially useful when the bootstrap filter is inefficient, while the auxiliary particle filter cannot be used because the state transition density is computationally intractable. The MPM method is a general extension of the PMMH method and consists of four parts; (a) it is based on an average of unbiased likelihood estimators; (b) it combines block and correlated PMMH methods; (c) a delayed acceptance proposal is used to speed up the computation, when estimating the likelihood is expensive; (d) an auxiliary disturbance particle filter sampler is proposed to estimate the likelihood. The MPM methodology is then applied to complex DSGE models with many latent state variables.

Our empirical results suggest that: (i) the improved disturbance particle filter (IDPF) is much more efficient than the standard bootstrap particle filter (BPF) and the tempered particle filter (TPF); (ii) the MPM maintains the correlation between logs of the estimated likelihoods in successive iterates much better than the CPM and BPM; (iii) the delayed acceptance version of the MPM sampler is much more efficient than the standard MPM sampler; (iv) the MPM with disturbance sorting is more efficient than the MPM with state sorting; (v) the performance of the MPM just using block sampling is as efficient as that using both block sampling and correlated sampling.

Finally, we believe that the methods in the paper will be very useful for many other models where particle alternative sophisticated methods, such as the particle

Gibbs, are either inefficient or impossible to use; e.g., partially observed diffusions and large panel data models. Future research will also consider developing better proposals for the parameters  $\theta$ .

## References

- Andreasen, M. M. (2013). Non-linear DSGE models and the central difference Kalman filter. *Journal of Applied Econometrics*, (28):929–955.
- Andreasen, M. M., Fernández-Villaverde, J., and Rubio-Ramírez, J. F. (2018). The pruned state-space system for non-linear DSGE models: Theory and empirical applications. *Review of Economic Studies*, 85(1):1–49.
- Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B*, 72:1–33.
- Andrieu, C. and Roberts, G. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, 37:697–725.
- Choppala, P., Gunawan, D., Chen, J., Tran, M. N., and Kohn, R. (2016). Bayesian inference for state-space models using block and correlated pseudo marginal methods. *arXiv:1612.07072v1*.
- Christen, J. A. and Fox, C. (2005). Markov chain Monte Carlo using an approximation. *Journal of Computational and Graphical Statistics*, 14(4):795–810.
- Deligiannidis, G., Doucet, A., and Pitt, M. (2018). The correlated pseudo-marginal method. *Journal of Royal Statistical Society Series B*, 80(5):839–870.
- Doucet, A., Pitt, M., Deligiannidis, G., and Kohn, R. (2015). Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika*, 102(2):295–313.
- Fernández-Villaverde, J. and Rubio-Ramírez, J. F. (2007). Estimating macroeconomic models: A likelihood approach. *Review of Economic Studies*, 74:1059–1087.
- Garthwaite, P. H., Fan, Y., and Sisson, S. A. (2015). Adaptive optimal scaling of Metropolis-Hastings algorithms using the Robbins-Monro process. *Communications in Statistics - Theory and Methods*, 45(17):5098–5111.
- Gordon, N., Salmond, D., and Smith, A. (1993). A novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEE-Proceedings F*, 140:107–113.

- Guarniero, P., Johansen, A. M., and Lee, A. (2017). The iterated auxiliary particle filter. *Journal of American Statistical Association*, 112:1636–1647.
- Hall, J., Pitt, M. K., and Kohn, R. (2014). Bayesian inference for nonlinear structural time series models. *Journal of Econometrics*, 179(2):99–111.
- Herbst, E. and Schorfheide, F. (2015). *Bayesian Estimation of DSGE Models*.
- Herbst, E. and Schorfheide, F. (2019). Tempered particle filtering. *Journal of Econometrics*, 210:26–44.
- Hesterberg, T. (1995). Weighted average importance sampling and defensive mixture distributions. *Technometrics*, 37:185–194.
- Kim, J., Kim, S., Schaumburg, E., and Sims, C. A. (2008). Calculating and using second-order accurate solutions of discrete time dynamic equilibrium models. *Journal of Economic Dynamics and Control*, 32:3397–3414.
- Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25.
- Kollmann, R. (2015). Tractable latent state filtering for non-linear DSGE models using a second-order approximation and pruning. *Computational Economics*, 45:239–260.
- Murray, L. M., Jones, E. M., and Parslow, J. (2013). On disturbance state-space models and the particle marginal Metropolis-Hastings sampler. *SIAM/ASA Journal on Uncertainty Quantification*, 1(1):494–521.
- Norgaard, M., Poulsen, N. K., and Ravn, O. (2000). Advances in derivative-free state estimation for non-linear systems. *Automatica*, 36(11):1627–1638.
- Pitt, M. K., Silva, R. S., Giordani, P., and Kohn, R. (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics*, 171(2):134–151.
- Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis of MCMC. *R News*, 6(1):7–11.
- Roberts, G. O. and Rosenthal, J. S. (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367.
- Rotemberg, J. J. (1982). Sticky prices in the united states. *Journal of Political Economy*, 90(6):1187–1211.

- Sherlock, C., Golightly, A., and Henderson, D. A. (2017a). Adaptive, delayed-acceptance MCMC for targets with expensive likelihoods. *Journal of Computational and Graphical Statistics*, 26(2):434–444.
- Sherlock, C., Thiery, A., Robert, G., and Rosenthal, J. (2015). On the efficiency of pseudo-marginal random walk Metropolis algorithms. *The Annals of Statistics*, 43(1):238–275.
- Sherlock, C., Thiery, A. H., and Lee, A. (2017b). Pseudo-marginal Metropolis-Hastings sampling using averages of unbiased estimators. *Biometrika*, 104(3):727–734.
- Skilling, J. (2004). Programming the Hilbert curve. *AIP Conference Proceedings*, 707:381–387.
- Tran, M. N., Kohn, R., Quiroz, M., and Villani, M. (2016). Block-wise pseudo-marginal Metropolis-Hastings. *arXiv:1603.02485v2*.

# Online Supplement for “Efficient Pseudo Marginal Method for State Space Models”

## S1 The Disturbance Particle Filter Algorithm

This section discusses the disturbance particle filter algorithm. Let  $u$  be the random vector used to obtain the unbiased estimate of the likelihood. It has the two components  $u_{\epsilon,1:T}^{1:N}$  and  $u_{A,1:T-1}^{1:N}$ ;  $u_{\epsilon,t}^i$  is the vector random variable used to generate the particles  $\epsilon_t^i$  given  $\theta$ . We can write

$$\epsilon_1^i \sim m(\epsilon_1^i | u_{\epsilon,1}^i, \theta), z_1^i = F(z_0, \epsilon_1^i; \theta) \text{ and } \epsilon_t^i \sim m(\epsilon_t^i | u_{\epsilon,t}^i, \theta), z_t^i = F(z_{t-1}^{a_{t-1}^i}, \epsilon_t^i; \theta), t \geq 2, \quad (\text{S1})$$

where  $m(\epsilon_t^i | u_{\epsilon,t}^i, \theta)$  is the proposal density to generate  $\epsilon_t^i$ , and  $z_0$  is a vector of zeros. Denote the distribution of  $u_{\epsilon,t}^i$  as  $\psi_{\epsilon t}(\cdot)$ . For  $t \geq 2$ , let  $u_{A,t-1}$  be the vector of random variables used to generate the ancestor indices  $a_{t-1}^{1:N}$  using the resampling scheme  $M(a_{t-1}^{1:N} | \bar{w}_{t-1}^{1:N}, z_{t-1}^{1:N})$  and define  $\psi_{A,t-1}(\cdot)$  as the distribution of  $u_{A,t-1}$ . Common choices for  $\psi_{\epsilon t}(\cdot)$  and  $\psi_{A,t-1}(\cdot)$  are iid  $N(0, 1)$  and iid  $U(0, 1)$  random variables, respectively.

The algorithm takes the number of particles  $N$ , the parameters  $\theta$ , the random variables used to generate the disturbance particles  $u_{\epsilon,1:T}^{1:N}$ , and the random variables used in the resampling steps  $u_{A,t-1}^{1:N}$  as the inputs; it outputs the set of state particles  $z_{1:T}^{1:N}$ , disturbance particles  $\epsilon_{1:T}^{1:N}$ , ancestor indices  $A_{1:T-1}^{1:N}$ , and the weights  $\bar{w}_{1:T}^{1:N}$ . At  $t = 1$ , the disturbance particles  $\epsilon_1^{1:N}$  are obtained as a function of the random numbers  $u_{\epsilon,1}^{1:N}$  using Eq. (S1) in step (1a) and the state particles are obtained from  $z_1^i = F(z_0, \epsilon_1^i; \theta)$ , for  $i = 1, \dots, N$ ; the weights for all particles are then computed in steps (1b) and (1c).

Step (2a) sorts the state or disturbance particles from smallest to largest using the simple Euclidean sorting procedure of Choppala et al. (2016) to obtain the sorted disturbance particles, sorted state particles and weights. Algorithm S2 resamples the particles using multinomial resampling to obtain the ancestor index  $a_{t-1}^{1:N}$  in the original order of particles in steps (2b) and (2c). Step (2d) generates the disturbance particles  $\epsilon_t^{1:N}$  as a function of the random numbers  $u_{\epsilon,t}^{1:N}$  using Eq. (S1) and the state particles are obtained from  $z_t^i = F(z_{t-1}^{a_{t-1}^i}, \epsilon_t^i; \theta)$ , for  $i = 1, \dots, N$ ; we then compute the weights for all particles in step (2e) and (2f).

The disturbance particle filter provides an unbiased estimate

$$\hat{p}_N(y|\theta, u) := \prod_{t=1}^T \left( N^{-1} \sum_{i=1}^N w_t^i \right),$$

of the likelihood, where

$$w_t^i = \frac{p(y_t|z_t^i, \theta) p(\epsilon_t^i)}{m(\epsilon_t^i|u_{\epsilon,t}, \theta)} \text{ for } t = 1, \dots, T, \text{ and } \bar{w}_t^i = \frac{w_t^i}{\sum_{j=1}^N w_t^j}. \quad (\text{S2})$$

---

**Algorithm S1** The Correlated Disturbance Particle Filter

---

Input:  $u_{\epsilon,1:T}^{1:N}$ ,  $u_{A,t-1}^{1:N}$ ,  $\theta$  and  $N$

Output:  $\epsilon_{1:T}^{1:N}$ ,  $z_{1:T}^{1:N}$ ,  $A_{1:T-1}^{1:N}$ , and  $\bar{w}_{1:T}^{1:N}$

For  $t = 1$

- (1a) Generate  $\epsilon_1^i$  from  $m(\epsilon_1^i|u_{\epsilon,1}^i, \theta)$  and set  $z_1^i = F(z_0, \epsilon_1^i; \theta)$  for  $i = 1, \dots, N$
- (1b) Compute the unnormalised weights  $w_1^i$ , for  $i = 1, \dots, N$
- (1c) Compute the normalised weights  $\bar{w}_1^i$  for  $i = 1, \dots, N$ .

For  $t \geq 2$

- (2a) Sort the state particles  $z_{t-1}^i$  or disturbance particles  $\epsilon_{t-1}^i$  using the Euclidean sorting method of Choppala et al. (2016) and obtain the sorted index  $\zeta_i$  for  $i = 1, \dots, N$ , and the sorted state particles, disturbance particles, and weights  $\tilde{z}_{t-1}^i = z_{t-1}^{\zeta_i}$ ,  $\tilde{\epsilon}_{t-1}^i = \epsilon_{t-1}^{\zeta_i}$  and  $\tilde{\bar{w}}_t^i = \bar{w}_{t-1}^i$ , for  $i = 1, \dots, N$ .
  - (2b) Obtain the ancestor indices based on the sorted state particles  $\tilde{a}_{t-1}^{1:N}$  using the correlated multinomial resampling in Algorithm S2.
  - (2c) Obtain the ancestor indices based on original order of the particles  $a_{t-1}^i$  for  $i = 1, \dots, N$ .
  - (2d) Generate  $\epsilon_t^i$  from  $m(\epsilon_t^i|u_{\epsilon,t}^i, \theta)$  and set  $z_t^i = F(z_{t-1}^{a_{t-1}^i}, \epsilon_t^i; \theta)$ , for  $i = 1, \dots, N$
  - (2e) Compute the unnormalised weights  $w_t^i$ , for  $i = 1, \dots, N$
  - (2f) Compute the normalised weights  $\bar{w}_t^i$  for  $i = 1, \dots, N$ .
-

---

**Algorithm S2** Multinomial Resampling Algorithm

---

Input:  $u_{A,t-1}$ , sorted states  $\tilde{z}_{t-1}^{1:N}$ , sorted disturbances  $\tilde{\epsilon}_{t-1}^{1:N}$ , and sorted weights  $\tilde{w}_{t-1}^{1:N}$   
Output:  $\tilde{a}_{t-1}^{1:N}$

1. Compute the cumulative weights

$$\widehat{F}_{t-1}^N(j) = \sum_{i=1}^j \tilde{w}_{t-1}^i$$

based on the sorted state particles  $\{\tilde{z}_{t-1}^{1:N}, \tilde{w}_{t-1}^{1:N}\}$  or the sorted disturbance particles  $\{\tilde{\epsilon}_{t-1}^{1:N}, \tilde{w}_{t-1}^{1:N}\}$

2. set  $\tilde{a}_{t-1}^i = \min_j \widehat{F}_{t-1}^N(j) \geq u_{A,t-1}^i$  for  $i = 1, \dots, N$ . Note that  $\tilde{a}_{t-1}^i$  for  $i = 1, \dots, N$  is the ancestor index based on the sorted states or the sorted disturbances.
-

## S2 The Mixed PMMH (MPM) Algorithm

---

**Algorithm S3** The Mixed PMMH (MPM) algorithm

---

- Set the initial values of  $\theta^{(0)}$  arbitrarily.
  - Sample  $u_g \sim N(0, I)$  for  $g = 1, \dots, G$ , and run  $G$  particle filters to compute an unbiased estimate of likelihood  $\bar{p}_N(y|\theta, \tilde{u}) = \frac{1}{G} \sum_{g=1}^G \hat{p}_N(y|\theta, u_g)$ , and run ancestral tracing algorithm in Supplement S5 after each particle filter algorithm to obtain the initial  $G$  trajectories of  $\epsilon_{g,1:T}$ . The mean  $\hat{\mu}_t$  and the covariance matrix  $\hat{\Sigma}_t$  of the proposal defined in Section 3.5 are set as the mean and the covariance matrix of these  $G$  trajectories of  $\epsilon_{g,1:T}$  at each time  $t$ .
  - For MCMC iteration  $i$ ,  $i = 1, \dots, M$ ,
    - Sample  $\theta'$  from the proposal density  $q(\theta'|\theta)$ .
    - Choose index  $g$  with probability  $1/G$ , sample  $\eta_u \sim N(0, I)$ , and set  $u'_g = \rho_u u_g + \sqrt{1 - \rho_u^2} \eta_u$ .
    - Run  $G$  particle filters to compute the unbiased estimate of likelihood  $\bar{p}_N(y|\theta', \tilde{u}')$
    - Run ancestral tracing algorithm in Supplement S5 after each particle filter algorithm to obtain  $G$  trajectories of  $\epsilon'_{g,1:T}$ .
    - With the probability in Equation (11), set  $\bar{p}(y_{1:T}|\theta, \tilde{u})^{(i)} = \bar{p}(y_{1:T}|\theta', \tilde{u}')$ ,  $\tilde{u}^{(i)} = \tilde{u}'$ ,  $\theta^{(i)} = \theta'$ , and  $\epsilon_{g,1:T}^{(i)} = \epsilon'_{g,1:T}$ ; otherwise, set  $\bar{p}(y_{1:T}|\theta, \tilde{u})^{(i)} = \bar{p}(y_{1:T}|\theta, \tilde{u})^{(i-1)}$ ,  $\tilde{u}^{(i)} = \tilde{u}^{(i-1)}$ ,  $\theta^{(i)} = \theta^{(i-1)}$ , and  $\epsilon_{g,1:T}^{(i)} = \epsilon_{g,1:T}^{(i-1)}$ .
    - The mean  $\hat{\mu}_t$  and the covariance matrix  $\hat{\Sigma}_t$  are set as the mean and the covariance matrix of these  $G$  trajectories of  $\epsilon_{g,1:T}^{(i)}$  at each time  $t$ .
- 

## S3 Multidimensional Euclidean Sorting Algorithm

This section discusses the multidimensional Euclidean sorting algorithm used to sort the particles in Algorithm S1. Let  $x_t^i$  be  $n_x$ -dimensional particles at a time  $t$ ,  $x_t^i = (x_{t,1}^i, \dots, x_{t,n_x}^i)^\top$ . Let  $d(x_t^j, x_t^i)$  be the Euclidean distance between two multidimensional particles. Algorithm S4 gives the multidimensional Euclidean sorting algorithm to generate sorted particles and weights with associated sorted indices. The first sorting index is the index of the particle having the smallest value along its first dimension. The other particles are chosen in a way that minimises the Euclidean distance between the last selected particle and the set of all remaining particles.

---

**Algorithm S4** Multidimensional Euclidean Sorting Algorithm

---

Input:  $x_t^{1:N}, \bar{w}_t^{1:N}$ Output: sorted particles  $\tilde{x}_t^{1:N}$ , sorted weights  $\tilde{w}_t^{1:N}$ , sorted indices  $\zeta_{1:N}$ Let  $\chi^j = \{1, \dots, N\}$  be the index set.When  $j = 1$ ,

- Obtain the index  $\zeta_1 = \min_i x_{t,1}^i, \forall i \in \chi^j$

For  $j = 2, \dots, N$ 

- Set  $x_t^* = x_t^{j-1}$
- Update the index  $\chi^j$  by removing  $\zeta_{j-1}$  from the index set.
- Obtain the index  $\zeta_j = \min_i d(x_t^*, x_t^i), \forall i \in \chi^j$ .

Sort the particles and weights according to the indices  $\zeta_{1:N}$  to obtain the sorted particles  $\tilde{x}_t^{1:N}$ , and sorted weights  $\tilde{w}_t^{1:N}$ .

---

## S4 Delayed Acceptance Mixed PMMH (DA-MPM)

Delayed acceptance MCMC can be used to speed up the computation for models with expensive likelihoods (Christen and Fox, 2005). The motivation in delayed acceptance is to avoid computing of the expensive likelihood if it is likely that the proposed draw will ultimately be rejected. A first accept-reject stage is applied with the cheap (or deterministic) approximation substituted for the expensive likelihood in the Metropolis Hastings acceptance ratio. Then, only for a proposal that is accepted in the first stage, the computationally expensive likelihood is calculated with a second accept-reject stage to ensure detailed balance is satisfied with respect to the true posterior. This section discusses the delayed-acceptance mixed PMMH (DA-MPM) algorithm.

Given the current parameter value  $\theta$ , the delayed acceptance Metropolis-Hastings (MH) algorithm proposes a new value  $\theta'$  from the proposal density  $q(\theta'|\theta)$  and uses a cheap approximation  $\hat{p}^c(y|\theta')$  to the likelihood in the MH acceptance probability

$$\alpha_1(\theta, \tilde{u}; \theta', \tilde{u}') = \min\left(1, \frac{\hat{p}^c(y|\theta') p(\theta') q(\theta|\theta')}{\hat{p}^c(y|\theta) p(\theta) q(\theta'|\theta)}\right). \quad (\text{S3})$$

As an alternative to particle filtering, Norgaard et al. (2000) develop the central difference Kalman filter (CDKF) for estimating the state in general non-linear and non-Gaussian state-space models. The CDKF frequently outperforms the extended

Kalman Filter (EKF) for general non-linear and non-Gaussian state-space models Andreasen (2013). We use the likelihood approximation,  $\widehat{p}^c(y|\theta)$ , obtained from the CDKF in the first stage accept-reject in Equation (S3).

A second accept-reject stage is applied to a proposal that is accepted in the first stage, with the second acceptance probability

$$\alpha_2(\theta, \tilde{u}; \theta', \tilde{u}') = \min\left(1, \frac{\bar{\widehat{p}}_N(y|\theta', \tilde{u}') \widehat{p}^c(y|\theta)}{\bar{\widehat{p}}_N(y|\theta, \tilde{u}) \widehat{p}^c(y|\theta')}\right). \quad (\text{S4})$$

The overall acceptance probability  $\alpha_1(\theta, \tilde{u}; \theta', \tilde{u}') \alpha_2(\theta, \tilde{u}; \theta', \tilde{u}')$  ensures that detailed balance is satisfied with respect to the true posterior. If a rejection occurs at stage one, then the expensive evaluation of the likelihood at the second stage is unnecessary (Sherlock et al., 2017a). Algorithm S5 describes the Delayed Acceptance MPM algorithm.

- Set the initial values of  $\theta^{(0)}$  arbitrarily
  - Sample  $u_g \sim N(0, I)$  for  $g = 1, \dots, G$ , and run  $G$  particle filters to compute an unbiased estimate of likelihood  $\widehat{p}_N(y|\theta, \tilde{u}) = \frac{1}{G} \sum_{g=1}^G \widehat{p}_N(y|\theta, u_g)$ , and run ancestral tracing algorithm in Supplement S5 after each particle filter algorithm to obtain the initial  $G$  trajectories of  $\epsilon_{g,1:T}$ . The mean  $\widehat{\mu}_t$  and the covariance matrix  $\widehat{\Sigma}_t$  of the proposal defined in Section 3.5 are set as the mean and the covariance matrix of these  $G$  trajectories of  $\epsilon_{g,1:T}$  at each time  $t$ .
  - For MCMC iteration  $i, i = 1, \dots, M$ ,
    - (1) Sample  $\theta'$  from the proposal density  $q(\theta'|\theta)$ .
    - (2) Compute the likelihood approximation  $\widehat{p}^c(y|\theta')$  using the central difference Kalman filter (CDKF).
    - (3) Accept the first stage proposal with the acceptance probability in Equation (S3). If the proposal is accepted, then go to step (4), otherwise go to step (8)
    - (4) Choose index  $g$  with probability  $1/G$ , sample  $\eta_u \sim N(0, I)$ , and set  $u'_g = \rho_u u_g + \sqrt{1 - \rho_u^2} \eta_u$ .
    - (5) Run  $G$  particle filters to compute an unbiased estimate of likelihood  $\widehat{p}_N(y|\theta', \tilde{u}')$ .
    - (6) Run ancestral tracing algorithm after each particle filter algorithm to obtain  $G$  trajectories of  $\epsilon'_{g,1:T}$ .
    - (7) Accept the second stage Metropolis-Hastings with acceptance probability in Equation (S4). If the proposal is accepted, then set  $\widehat{p}_N(y|\theta, \tilde{u})^{(i)} = \widehat{p}_N(y|\theta', \tilde{u}')$ ,  $\widehat{p}^c(y|\theta)^{(i)} = \widehat{p}^c(y|\theta')$ ,  $\tilde{u}^{(i)} = \tilde{u}'$ ,  $\epsilon_{g,1:T}^{(i)} = \epsilon'_{g,1:T}$  and  $\theta^{(i)} = \theta'$ . Otherwise, go to step (8).
    - (8) Otherwise,  $\widehat{p}_N(y|\theta, \tilde{u})^{(i)} = \widehat{p}_N(y|\theta, \tilde{u})^{(i-1)}$ ,  $\widehat{p}^c(y|\theta)^{(i)} = \widehat{p}^c(y|\theta)^{(i-1)}$ ,  $\tilde{u}^{(i)} = \tilde{u}^{(i-1)}$ ,  $\epsilon_{g,1:T}^{(i)} = \epsilon_{g,1:T}^{(i-1)}$  and  $\theta^{(i)} = \theta^{(i-1)}$
    - (9) The mean  $\widehat{\mu}_t$  and the covariance matrix  $\widehat{\Sigma}_t$  are set as the mean and the covariance matrix of these  $G$  trajectories of  $\epsilon_{g,1:T}^{(i)}$  at each time  $t$ .
- 

## S5 Ancestral Tracing

The simplest way of sampling from the particle approximation of  $p(\epsilon_{1:T}|y_{1:T}, \theta)$  is called ancestral tracing (Kitagawa, 1996). This consists of sampling one particle trajectory from the particle filter. The method is equivalent to sampling an index  $J = j$  with probability  $\bar{w}_t^j$ , tracing back its ancestral lineage  $b_{1:T}^j$  ( $b_T^j = j$  and  $b_{t-1}^j = a_{t-1}^{b_t^j}$ )

and choosing the particle trajectory  $\epsilon_{1:T}^j = (\epsilon_1^j, \dots, \epsilon_T^j)$ . The selected particle trajectory is used to generate the IDPF proposal in Section 3.5.

## S6 Overview of DSGE models

This section presents a brief overview of DSGE models. To make it self-contained, we repeat the solution and the state space representation explained in Section 2.2.

The equilibrium conditions for a wide variety of DSGE models can be summarized by

$$E_t G(z_{t+1}, z_t, \epsilon_{t+1}; \theta) = 0; \quad (\text{S5})$$

$E_t$  is the mathematical expectation conditional on date  $t$  information and  $G : \mathbb{R}^{2n+m} \mapsto \mathbb{R}^n$ ;  $z_t$  is an  $n \times 1$  vector containing all variables known at time  $t$ ;  $\epsilon_{t+1}$  is an  $m \times 1$  vector of serially independent innovations. The solution to Eq. (S5) for  $z_{t+1}$  can be written as

$$z_{t+1} = F(z_t, \epsilon_{t+1}, \zeta; \theta) \quad (\text{S6})$$

such that

$$E_t G(F(z_t, \epsilon_{t+1}, \zeta; \theta), z_t, \epsilon_{t+1}; \theta) = 0 \text{ for any } t.$$

For our applications, we assume that  $\epsilon_t \sim N(0, \zeta^2 \Sigma_\epsilon)$ , where  $\zeta$  is a scalar perturbation parameter. Under suitable differentiability assumptions, and using the notation in Kollmann (2015), we now describe the first and second order solutions.

For most applications, the function  $F(\cdot)$  in Eq. (S6) is analytically intractable and is approximated using local solution techniques. We use first and second order Taylor series approximations around the *deterministic* steady state with  $\zeta = 0$  to approximate  $F$ . The *deterministic* steady state  $z^s$  satisfies  $z^s = F(z^s, 0, 0; \theta)$ .

A first order-accurate approximation, around the *deterministic* steady state is

$$z_{t+1}^d = F_1(\theta)z_t^d + F_2(\theta)\epsilon_{t+1}, \quad z_0^d = 0. \quad (\text{S7})$$

For Eq. (S7) to be stable, it is necessary for all the eigenvalues of  $F_1(\theta)$  to be less than 1 in absolute value. The initial value  $z_0^d = 0$  because we work with approximations around the *deterministic* steady state. For a given set of parameters  $\theta$ , the matrices  $F_1(\theta)$  and  $F_2(\theta)$  can be solved using existing software; our applications use Dynare.

The second order accurate approximation (around the *deterministic* steady state) is

$$z_{t+1}^d = F_0(\theta)\zeta^2 + F_1(\theta)z_t^d + F_2(\theta)\epsilon_{t+1} + F_{11}(\theta)P(z_t^d) + F_{12}(\theta)(z_t^d \otimes \epsilon_{t+1}) + F_{22}(\theta)P(\epsilon_{t+1}); \quad (\text{S8})$$

where  $z_t^d = (z_t - z^s)$  and for any vector  $x := (x_1, \dots, x_m)^\top$ , we define  $P(x) := \text{vech}(xx^\top)$ , where  $\text{vech}(xx^\top)$  is the strict upper triangle and diagonal of  $xx^\top$ , such that

$\text{vech}(xx^\top) := (x_1^2, x_1x_2, x_2^2, x_1x_3, \dots, x_m^2)^\top$ . The term  $F_0(\theta)\zeta^2$  captures the level correction due to uncertainty that arises from taking a second-order approximation. For our analysis we normalize the perturbation parameter  $\zeta$  to 1. As before, for a given set of parameters  $\theta$  the matrices  $F_1(\theta), F_2(\theta), F_{11}(\theta), F_{12}(\theta), F_{22}(\theta)$  can be solved using Dynare.

The measurement (observation) equation for the DSGE model and its approximations is

$$y_t = Hz_t + \eta_t, \quad \eta_t \sim N(0, \Sigma_\eta), \quad (\text{S9})$$

where  $H$  is a known matrix;  $\{\eta_t\}$  is an independent  $N(0, \Sigma_\eta)$  and sequence and it is also independent of the  $\{\epsilon_t\}$  sequence. The matrix  $\Sigma_\eta$  is usually unknown and estimated from the data.

## S6.1 Pruned State Space Representation of DSGE Models

To obtain a stable solution we use the pruning approach recommended in Kim et al. (2008); for details of how pruning is implemented in DSGE models, see Kollmann (2015) and Andreasen et al. (2018). The pruned second order solution preserves only second order accurate terms by using the first order accurate solution to calculate  $P(z_t^{(2)})$  and  $(z_t^{(2)} \otimes \epsilon_{t+1})$ ; where  $P(\cdot)$  is defined below. That is, we obtain the solution preserving *only* the second order effects by using  $P(z_t^{(1)})$  instead of  $P(z_t^{(2)})$  and  $(z_t^{(1)} \otimes \epsilon_{t+1})$  instead  $(z_t^{(2)} \otimes \epsilon_{t+1})$  in Eq. (S8). We note that the variables are still expressed as deviation from steady state; however, for notational simplicity, we drop the superscript  $d$ . For a vector  $x = (x_1, \dots, x_q)^\top$

$$P(x) := \text{vech}(xx^\top) = ((x_1)^2, (x_1(x_2)), \dots, (x_1(x_q)), (x_2)_2, (x_2(x_3)), \dots, (x_2(x_q)), \dots, (x_{q-1})^2, (x_{q-1}(x_q)), \dots, (x_q)^2).$$

The evolution of  $P(z_t^{(1)})$  is

$$P(z_{t+1}^{(1)}) = K_{11}P(z_t^{(1)}) + K_{12}(z_t^{(1)} \otimes \epsilon_{t+1}) + K_{22}P(\epsilon_{t+1})$$

with  $K_{11}, K_{12}, K_{22}$  being functions of  $F_1$  and  $F_2$  respectively. A consequence of using pruning in preserving only second order effects is that it increases the state space.

The pruning-augmented solution of the DSGE model is given by

$$\begin{bmatrix} z_{t+1}^{(2)} \\ P(z_{t+1}^{(1)}) \\ z_{t+1}^{(1)} \end{bmatrix} = \begin{bmatrix} F_0 \zeta^2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} F_1 & F_{11} & 0 \\ 0 & K_{11} & 0 \\ 0 & 0 & F_1 \end{bmatrix} \begin{bmatrix} z_t^{(2)} \\ P(z_t^{(1)}) \\ z_t^{(1)} \end{bmatrix} + \begin{bmatrix} F_2 \\ 0 \\ F_2 \end{bmatrix} \epsilon_{t+1} + \begin{bmatrix} F_{12} \\ K_{12} \\ 0 \end{bmatrix} (z_t^{(1)} \otimes \epsilon_{t+1}) + \begin{bmatrix} F_{22} \\ K_{22} \\ 0 \end{bmatrix} P(\epsilon_{t+1}).$$

The augmented state representation of the pruned second order accurate system is therefore

$$\tilde{z}_{t+1} = g_0 + G_1 \tilde{z}_t + G_2 \epsilon_{t+1} + G_{12} (z_t^{(1)} \otimes \epsilon_{t+1}) + G_{22} P(\epsilon_{t+1}),$$

where  $\tilde{z}_t = [z_t^{(2)}, P(z_t^{(1)}), z_t^{(1)}]^\top$ . We can now summarize the state transition and the measurement equations for the pruned second order accurate system as

$$\begin{aligned} y_{t+1} &= H \tilde{z}_{t+1} + \eta_{t+1}, \quad \eta_{t+1} \sim N(0, \Sigma_\eta) \\ \tilde{z}_{t+1} &= g_0 + G_1 \tilde{z}_t + G_2 \epsilon_{t+1} + G_{12} (z_t^{(1)} \otimes \epsilon_{t+1}) + G_{22} P(\epsilon_{t+1}), \quad \epsilon_{t+1} \sim N(0, \zeta^2 \Sigma_\epsilon) \end{aligned} \quad (\text{S10})$$

The matrix  $H$  selects the observables from the pruning augmented state space representation of the system. If  $\dim(z_t^{(2)}) = n \times 1$  and  $\dim(\epsilon_t) = m \times 1$  then  $\dim(P(z_t^{(1)})) = n(n+1)/2 \times 1$  and  $\dim(\tilde{z}_t) = [n + n(n+1)/2 + n] \times 1$ . If  $n_y \leq n$  denotes the number of observables then the selection matrix

$$H = [H_{ny}; \mathbf{0}]$$

where  $\dim(H_{ny}) = n_y \times [n + n(n+1)/2 + n]$  and  $\dim(\mathbf{0}) = (n - n_y) \times [n + n(n+1)/2 + n]$ . The selection matrix  $H_{ny}$  consists only of zeros and ones.

## S7 Description: Small scale DSGE model

The specification of the small scale DSGE model follows Herbst and Schorfheide (2015), and is also used by Herbst and Schorfheide (2019).

**Households** Time is discrete, households live forever, the representative household derives utility from consumption  $C_t$  relative to a habit stock (which is approximated by the level of technology  $A_t$ )<sup>3</sup> and real money balances  $\frac{M_t}{P_t}$  and disutility from hours

---

<sup>3</sup>This assumption ensures that the economy evolves along a balanced growth path even if the utility function is additively separable in consumption, real money balances and hours.

worked  $N_t$ . The household maximizes

$$E_0 \sum_{t=0}^{\infty} \beta^t \left[ \frac{\left(\frac{C_t}{A_t}\right)^{1-\tau} - 1}{1-\tau} + \chi_M \log \frac{M_t}{P_t} - \chi_N N_t \right]$$

subject to the budget constraint

$$P_t C_t + B_t + M_t + T_t = P_t W_t N_t + R_{t-1} B_{t-1} + M_{t-1} + P_t D_t + P_t SC_t;$$

$T_t$ ,  $D_t$  and  $SC_t$  denote lump-sum taxes, aggregate residual profits and net cash inflow from trading a full set of state contingent securities;  $P_t$  is the aggregate price index, and  $W_t$  is the real wage;  $\beta$  is the discount factor,  $\tau$  is the coefficient of relative risk aversion,  $\chi_M$  and  $\chi_N$  are scale factors determining the steady state money balance holdings and hours. We set  $\chi_N = 1$ .  $A_t$  is the level of aggregate productivity.

**Firms** Final output is produced by a perfectly competitive representative firm which uses a continuum of intermediate goods  $Y_t(i)$  and the production function

$$Y_t = \left( \int_0^1 Y_t(i)^{1-\nu} di \right)^{\frac{1}{1-\nu}},$$

with  $\nu < 1$ . The demand for intermediate good  $i$  is

$$Y_t(i) = \left( \frac{P_t(i)}{P_t} \right)^{-\frac{1}{\nu}} Y_t,$$

and the aggregate price index is

$$P_t = \left( \int_0^1 P_t(i)^{\frac{\nu-1}{\nu}} di \right)^{\frac{\nu}{\nu-1}}.$$

Intermediate good  $i$  is produced using the linear production technology

$$Y_t(i) = A_t N_t(i);$$

$A_t$  is an exogenous productivity process common to all firms, and  $N_t(i)$  is the labor input of firm  $i$ . Intermediate firms set prices ( $P_t(i)$ ) and labor input ( $N_t(i)$ ) by maximizing the net present value of future profit. Nominal rigidities are introduced

through price adjustment costs following Rotemberg (1982).

$$E_t \sum_{s=0}^{\infty} \beta^s Q_{t,t+s} \left[ \frac{P_{t+s}(i)}{P_{t+s}} Y_{t+s}(i) - W_{t+s} N_{t+s}(i) - AC_{t+s}(i) \right]$$

with

$$AC_t(i) = \frac{\phi}{2} \left( \frac{P_t(i)}{P_{t-1}(i)} - \pi \right)^2 Y_t(i)$$

The parameter  $\phi$  governs the extent of price rigidity in the economy and  $\pi$  is the steady state inflation rate associated with the final good. In equilibrium, households and firms use the same stochastic discount factor  $Q_{t,t+s}$  where

$$Q_{t,t+s} = \left( \frac{C_{t+s}}{C_t} \right)^{-\tau} \left( \frac{A_t}{A_{t+s}} \right)^{1-\tau} \quad (\text{S11})$$

In a symmetric equilibrium all firms choose the same price.

**Monetary and Fiscal Policy** The central bank conducts monetary policy following an interest rate feedback rule given by

$$R_t = R_t^{*1-\rho_R} R_{t-1}^{\rho_R} \epsilon_t^R$$

where  $\epsilon_t^R \sim IID(0, \sigma_r)$  is an iid shock to the nominal interest rate,  $R_t^*$  is the nominal target and  $(1 - \rho_R)$  captures interest rate smoothing in the conduct of policy,

$$R_t^* = r \pi^* \left( \frac{\pi_t}{\pi^*} \right)^{\psi_1} \left( \frac{Y_t}{Y_t^*} \right)^{\psi_2};$$

$\pi_t := \frac{P_t}{P_{t-1}}$  is the gross inflation rate and  $\pi^*$  is the target inflation rate.  $Y_t^*$  is the output in the absence of nominal rigidities. The parameters  $\psi_1$  and  $\psi_2$  capture the intensity with which the central bank responds to inflation and output gap in the model. Government expenditure accounts for a fraction  $v_t \in [0, 1]$  of final output such that  $G_t = v_t Y_t$ . The government budget constraint is

$$P_t G_t + R_{t-1} B_{t-1} + M_{t-1} = T_t + B_t + M_t$$

**Aggregation** Combining the household budget constraint with the government budget constraint gives

$$C_t + G_t + AC_t = Y_t$$

where in equilibrium,  $AC_t = \frac{\phi}{2} (\pi_t - \pi)^2 Y_t$ .

**Exogenous Processes** Aggregate technology grows at the rate  $\gamma$  and  $m_t$  is the shock to aggregate demand such that

$$\log A_t = \log \gamma + \log A_{t-1} + \log m_t$$

$$\log m_t = (1 - \rho_m) \log m + \rho_m \log m_{t-1} + \epsilon_t^m$$

with  $\epsilon_t^m \sim IID(0, \sigma_m)$ . We define  $g_t := 1/(1 - v_t)$  and  $g_t$  evolves as

$$\log g_t = (1 - \rho_g) \log g + \rho_g \log g_{t-1} + \epsilon_t^g$$

with  $\epsilon_t^g \sim IID(0, \sigma_g)$ . We summarize the nonlinear equilibrium conditions after detrending  $C_t, Y_t, G_t$  by deterministic technology i.e define  $\tilde{X}_t := X_t/A_t$ .

$$1 = \beta E_t \left[ \left( \frac{\tilde{C}_{t+1}}{\tilde{C}_t} \right)^{-\tau} z_t \frac{R_t}{\pi_{t+1}} \right] \quad (\text{S12})$$

$$1 = \phi(\pi_t - \pi) \left[ \left( 1 - \frac{1}{2\nu} \right) \pi_t + \frac{\pi}{2\nu} \right] - \phi E_t \left[ \left( \frac{\tilde{C}_{t+1}}{\tilde{C}_t} \right)^{-\tau} \frac{\tilde{Y}_{t+1}}{\tilde{Y}_t} (\pi_{t+1} - \pi) \pi_{t+1} \right] + \frac{1}{\nu} \left[ 1 - \left( \tilde{C}_t \right)^\tau \right] \quad (\text{S13})$$

$$R_t = R_t^{*1-\rho_R} R_{t-1}^{\rho_R} \exp \epsilon_t^R \quad (\text{S14})$$

$$R_t^* = r \pi^* \left( \frac{\pi_t}{\pi^*} \right)^{\psi_1} \left( \frac{\tilde{Y}_t}{\tilde{Y}_t^*} \right)^{\psi_2} \quad (\text{S15})$$

$$\tilde{C}_t + \tilde{G}_t + \frac{\phi}{2} (\pi_t - \pi)^2 \tilde{Y}_t = \tilde{Y}_t \quad (\text{S16})$$

$$\log A_t = \log \gamma + \log A_{t-1} + \log m_t \quad (\text{S17})$$

$$\log m_t = (1 - \rho_m) \log m + \rho_m \log m_{t-1} + \epsilon_t^m \quad (\text{S18})$$

$$\log g_t = (1 - \rho_g) \log g + \rho_g \log g_{t-1} + \epsilon_t^g \quad (\text{S19})$$

After detrending, the steady state solution of the model is  $\pi = \pi^*$ ,  $\tilde{R} = \frac{\gamma\pi}{\beta}$ ,  $\tilde{C} = (1 - \nu)^{\frac{1}{\tau}}$  and  $\tilde{Y} = g\tilde{C} = \tilde{y}^*$ .

Equations (S12)-(S19) can now be rewritten as

$$E_t G(z_{t+1}, z_t, \epsilon_{t+1}) = 0,$$

where

$$z_t = [\tilde{C}_t, \tilde{Y}_t, \tilde{G}_t, R_t, R_t^*, A_t, m_t, g_t] \quad \text{and} \quad \epsilon_t = [\epsilon_t^R, \epsilon_t^m, \epsilon_t^g],$$

and solved using the methods in Section S6 of the Supplement.

We solve the model in log deviations from steady state, i.e.  $\hat{x}_t := \log\left(\frac{\tilde{x}_t}{\bar{x}}\right)$  using Dynare. The small scale DSGE model therefore consists of a consumption Euler equation, a new Keynesian Philip curve, a monetary policy rule, fiscal policy rule, three exogenous shock processes, and eight endogenous latent variables. We estimate the model using both first and second order approximations to assess the performance of our proposed method relative to existing methods. As explained, the source of nonlinearity in our case stems from taking a second order approximation of the equilibrium conditions.

The observables used in estimating the model consist of per capita GDP growth rate ( $YGR_t$ ), annualized quarter on quarter inflation rate ( $Infl_t$ ) and annualized nominal rates ( $Int_t$ ). The observed data is measured in percentages, and, after applying a log transformation to the endogenous variables, the measurement equations for our system using the transformed endogenous variables given by (a hatted variable denotes the log transformation).

$$YGR_t = \gamma^{(Q)} + 100(\hat{y}_t - \hat{y}_{t-1} + \hat{m}_t), \quad (\text{S20})$$

$$Infl_t = \pi^{(A)} + 400\hat{\pi}_t, \quad (\text{S21})$$

$$Int_t = \pi^{(A)} + r^{(A)} + 4\gamma^{(Q)} + 400\hat{R}_t. \quad (\text{S22})$$

Here,  $\hat{y}_t = \log\left(\frac{\tilde{Y}_t}{\bar{Y}}\right)$ ,  $\hat{\pi}_t = \log\left(\frac{\tilde{\pi}_t}{\bar{\pi}}\right)$ ,  $\hat{R}_t = \log\left(\frac{\tilde{R}_t}{\bar{R}}\right)$  and  $\hat{m}_t = \log\left(\frac{\tilde{m}_t}{\bar{m}}\right) = \log m_t$  since  $\log(\bar{m}) = 0$  in steady state as evident from Equation (S18).

The model has 15 parameters:

$$\left\{ \nu, \nu, \phi, \tau, \pi^{(A)}, r^{(A)}, \gamma^{(Q)}, \psi_1, \psi_2, \rho_r, \rho_g, \rho_m, \sigma_r, \sigma_g, \sigma_m \right\}.$$

We calibrate the parameters characterizing the share of fiscal expenditure in GDP  $\nu = 0.2$ , the elasticity of substitution across varieties  $1/\nu = 11$  and the parameter guiding the extent of nominal rigidities,  $\phi = 100$ . The remaining parameters along with the measurement errors in the observation equation are estimated. The steady-state inflation rate,  $\pi$ , in the model relates to the estimated parameter for annualized inflation  $\pi^{(A)}$  such that  $\pi = \pi^{(A)}/400$  and the discount factor  $\beta$  to the estimated parameter for annualized interest rate  $r^{(A)}$  such that  $\beta = \left((1 + r^{(A)})/400\right)^{-1}$ . The sample used for estimation is 1983Q1-2002Q4. The data set and the variables are

identical to those used in Herbst and Schorfheide (2019).<sup>4</sup>

## S8 Additional Tables

**Table S1:** The parameter values used in Section 4.3

Param.	$\theta$	Param.	$\theta$
$\tau$	2.09	$\kappa$	0.98
$\psi_1$	2.25	$\psi_2$	0.65
$\rho_r$	0.81	$\rho_g$	0.98
$\rho_m$	0.93	$r^{(A)}$	0.34
$\pi^{(A)}$	3.16	$\gamma^{(Q)}$	0.51
$\sigma_r$	0.19	$\sigma_g$	0.65
$\sigma_m$	0.24		

**Table S2:** Marginal prior distribution for each model parameter for the non-linear small scale DSGE model in Section 4.5. Param (1) and Param (2) list the means and the standard deviation for truncated normal (TN) and Normal distribution;  $v$  and  $s$  for the inverse gamma (IG) distribution. The domain  $R^+$  is  $(0, \infty)$ .  $\text{TN}_{(0,\infty)}$  is the normal distribution lies within the interval  $(0, \infty)$ .  $\text{TN}_{(0,1)}$  is the normal distribution lies within the interval  $(0, 1)$ .  $\sigma_r^m$ ,  $\sigma_g^m$ , and  $\sigma_m^m$  are the measurement error variances.

Param.	Domain	Density	Param (1)	Param (2)
$r^{(A)}$	$R^+$	$\text{TN}_{(0,\infty)}$	0.7	0.5
$\pi^{(A)}$	$R^+$	$\text{TN}_{(0,\infty)}$	3.12	0.5
$\gamma^{(Q)}$	$R$	N	0.60	0.5
$\tau$	$R^+$	$\text{TN}_{(0,\infty)}$	1.50	0.5
$\psi_1$	$R^+$	$\text{TN}_{(0,\infty)}$	2.41	0.5
$\psi_2$	$R^+$	$\text{TN}_{(0,\infty)}$	0.39	0.5
$\rho_r$	$[0, 1]$	$\text{TN}_{(0,1)}$	0.5	0.2
$\rho_g$	$[0, 1]$	$\text{TN}_{(0,1)}$	0.5	0.2
$\rho_m$	$[0, 1]$	$\text{TN}_{(0,1)}$	0.5	0.2
$\sigma_r$	$R^+$	IG	5	0.5
$\sigma_g$	$R^+$	IG	5	0.5
$\sigma_m$	$R^+$	IG	5	0.5
$\sigma_r^m$	$R^+$	$\text{TN}_{(0,\infty)}$	2	2
$\sigma_g^m$	$R^+$	$\text{TN}_{(0,\infty)}$	2	2
$\sigma_m^m$	$R^+$	$\text{TN}_{(0,\infty)}$	2	2

<sup>4</sup>Data on all three observables are sourced from FRED. Per capita GDP growth rate is calculated using data on real gross domestic product (FRED mnemonic ‘GDPC1’) and Civilian Non-institutional Population (FRED mnemonic ‘CNP16OV’ / BLS series ‘LNS1000000’), Annualized Inflation is calculated using CPI price level (FRED mnemonic ‘CPIAUCSL’), the Federal Funds Rate (FRED mnemonic ‘FEDFUNDS’).

**Table S3:** Marginal prior distribution for the RBC model parameters discussed in Section 4.4. Param (1) and Param (2) list the means and the standard deviation for truncated normal (TN) and Normal distribution; the shape parameter  $v$  and the scale parameter  $s$  for the inverse gamma (IG) distribution. The domain  $R^+$  is  $(0, \infty)$ .  $TN_{(0,1)}$  is the normal distribution lies within the interval  $(0, 1)$ .

Param.	Domain	Density	Param (1)	Param (2)
$\rho_a$	$[0, 1]$	$TN_{(0,1)}$	0.8	0.05
$\rho_e$	$[0, 1]$	$TN_{(0,1)}$	0.8	0.05
$\sigma_a$	$R^+$	IG	5	0.5
$\sigma_e$	$R^+$	IG	5	0.5

## S9 Further Results for the Non-Linear Small Scale DSGE Models

Table S4 reports the posterior estimates when the measurement error is fixed to 50% of the variance of the corresponding observable. Table S5 reports the estimates when the measurement errors are also estimated.

**Estimation results – posterior estimates:** We compare the parameter estimates obtained from the linear model (first order approximation) estimated using the Kalman filter and the nonlinear model (second order approximation) when the measurement error variances are fixed to 50% of the variance of the observables and when the measurement error variances are estimated. Tables S4 and S5 show that the model solved using a second order approximation produces different estimates from that taking a first order approximation.

**Table S4:** Mean, 5%, and 95% quantiles of the posterior distributions of each of the parameters in the small scale DSGE model estimated using Kalman Filter (linear model) and DA-MPM (second order approximation). The measurement error variances are fixed to 50% of the variance of the observables. Param (1) and Param (2) list the means and the standard deviation for truncated normal (TN) and Normal distribution;  $v$  and  $s$  for the inverse gamma (IG) distribution. The domain  $R^+$  is  $(0, \infty)$ .  $TN_{(0,\infty)}$  is the normal distribution lies within the interval  $(0, \infty)$ .  $TN_{(0,1)}$  is the normal distribution lies within the interval  $(0, 1)$

Parameters	Prior			Posterior					
	Distribution	Param(1)	Param(2)	Mean (1 <sup>st</sup> Order)	5 Percent	95 Percent	Mean (2 <sup>nd</sup> Order)	5 Percent	95 Percent
$\tau$	$TN_{(0,\infty)}$	1.5	0.5	1.638	1.005	2.346	1.049	0.807	1.436
$\psi_1$	$TN_{(0,\infty)}$	2.41	0.5	2.309	1.677	2.956	2.371	1.790	3.046
$\psi_2$	$TN_{(0,\infty)}$	0.39	0.5	0.541	0.062	1.164	0.254	0.006	0.793
$\pi^{(A)}$	$TN_{(0,\infty)}$	3.12	0.5	3.575	3.06	4.089	2.922	2.514	3.317
$r^{(A)}$	$TN_{(0,\infty)}$	0.7	0.5	0.621	0.234	1.031	0.711	0.225	1.242
$\gamma^{(Q)}$	N	0.6	0.5	0.769	0.545	1.001	0.433	0.293	0.574
$\rho_r$	$TN_{(0,1)}$	0.5	0.2	0.592	0.374	0.766	0.193	0.037	0.392
$\rho_g$	$TN_{(0,1)}$	0.5	0.2	0.557	0.216	0.923	0.576	0.241	0.899
$\rho_m$	$TN_{(0,1)}$	0.5	0.2	0.975	0.941	0.997	0.934	0.889	0.971
$\sigma_r$	IG	5	0.5	0.081	0.046	0.131	0.072	0.043	0.112
$\sigma_g$	IG	5	0.5	0.1	0.049	0.187	0.096	0.050	0.171
$\sigma_m$	IG	5	0.5	0.135	0.104	0.17	0.076	0.050	0.110

**Table S5:** Mean, 5%, and 95% quantiles of the posterior distributions of each of the parameters in the small scale DSGE model estimated using Kalman Filter (linear model) and DA-MPM (second order approximation). The measurement error variances are estimated. Param (1) and Param (2) list the means and the standard deviation for truncated normal (TN) and Normal distribution;  $v$  and  $s$  for the inverse gamma (IG) distribution. The domain  $R^+$  is  $(0, \infty)$ .  $TN_{(0,\infty)}$  is the normal distribution lies within the interval  $(0, \infty)$ .  $TN_{(0,1)}$  is the normal distribution lies within the interval  $(0, 1)$

Parameters	Mean (1 <sup>st</sup> Order)	5 Percent	95 Percent	Mean (2 <sup>nd</sup> Order)	5 Percent	95 Percent
$\tau$	1.627	0.99	2.342	0.891	0.868	0.915
$\psi_1$	2.576	2.094	3.107	3.417	2.890	3.972
$\psi_2$	0.566	0.084	1.219	1.199	0.651	1.822
$\pi^{(A)}$	3.96	3.601	4.32	4.000	3.574	4.371
$r^{(A)}$	0.666	0.267	1.072	1.600	1.060	2.101
$\gamma^{(Q)}$	0.961	0.812	1.114	0.909	0.707	1.075
$\rho_r$	0.695	0.574	0.801	0.534	0.404	0.670
$\rho_g$	0.524	0.218	0.837	0.563	0.224	0.842
$\rho_m$	0.989	0.971	0.999	0.998	0.995	0.999
$\sigma_r$	0.066	0.044	0.09	0.046	0.032	0.069
$\sigma_g$	0.091	0.05	0.151	0.108	0.046	0.182
$\sigma_m$	0.132	0.106	0.16	0.118	0.100	0.140

## S10 Description of the Real Business Cycle (RBC) model

This section describes a simple RBC model without labor and estimates it using the simulated data in Section 4.4. The model assumes that households live infinitely and choose consumption and capital stock by maximizing the CRRA utility function, subject to the budget constraint, production function, law of motion of capital and exogenous processes guiding the evolution of technology and preferences. The utility function is

$$U_t = E_t \left[ \sum_{s=0}^{\infty} \beta^s e_t \frac{c_{t+s}^{1-\sigma} - 1}{1-\sigma} \right],$$

where  $e_t$  is a preference shock, modeled as

$$\log e_{t+1} = \rho_e \log e_t + \epsilon_{e,t+1} \quad (\text{S23})$$

with  $\epsilon_{e,t+1} \sim \mathcal{NID}(0, \sigma_e^2)$ . The budget constraint is

$$c_t + i_t = y_t, \quad (\text{S24})$$

where  $i_t$  is investment and  $y_t$  is total income (= total production) at time  $t$ . Production is carried out using

$$y_t = a_t k_t^\alpha, \quad (\text{S25})$$

where  $k_t$  is the capital stock in the economy and  $a_t$  is the exogenously given level of technology. The law of motion for capital is

$$k_{t+1} = (1 - \delta) k_t + i_t. \quad (\text{S26})$$

Finally, we let

$$\log a_{t+1} = \rho_a \log a_t + \epsilon_{a,t+1}, \quad (\text{S27})$$

where  $\epsilon_{a,t+1} \sim \mathcal{NID}(0, \sigma_a^2)$ . The optimality conditions of the model are given by

$$c_t^{-\sigma} e_t = E_t [\beta c_{t+1}^{-\sigma} e_{t+1} (\alpha a_{t+1} k_{t+1}^{\alpha-1} + 1 - \delta)] \quad (\text{S28})$$

$$c_t + k_{t+1} = a_t k_t^\alpha + (1 - \delta) k_t \quad (\text{S29})$$

Equations (S23), (S27)–(S29) summarize the equilibrium conditions of the model and can be expressed as

$$E_t G(z_{t+1}, z_t, \epsilon_{t+1}) = 0,$$

where

$$z_t = [c_t, k_t, i_t, e_t, a_t] \quad \text{and} \quad \epsilon_t = [\epsilon_{a,t}, \epsilon_{e,t}];$$

they are solved using the methods described in Section S6 of the Supplement. We solve the model in log deviation from steady state i.e.  $\hat{x}_t := \log(\frac{X_t}{\bar{X}})$ . The parameters of the model are calibrated as follows:  $\beta = 0.95$ ,  $\delta = 0.025$ ,  $\alpha = 0.90$ ,  $\sigma = 2$ ,  $\rho_a = \rho_e = 0.8$ ,  $\sigma_e = \sigma_a = 0.5$ .