Dynare Working Papers Series
http://www.dynare.org/wp/

An Improved Auxiliary Particle Filter for Nonlinear Dynamic Equilibrium Models

Yuan Yang
Lu Wang

Working Paper no. 47

November 2015

CEPREMAP
CENTRE POUR LA RECHERCHE ECONOMIQUE ET SES APPLICATIONS
142, rue du Chevaleret — 75013 Paris — France
http://www.cepremap.fr

An Improved Auxiliary Particle Filter for Nonlinear Dynamic Equilibrium Models

Yuan Yang* Lu Wang†

October 21, 2015

Abstract

We develop a procedure that efficiently computes likelihood function in nonlinear dynamic stochastic general equilibrium (DSGE) models. The procedure employs linearization to the measurement equation and delivers competitive results as the fully-adapted particle filter. The resulting likelihood approximation has much lower Monte Carlo variance than currently available particle filters, which greatly enhances the likelihood-based inference of DSGE models. We illustrate our procedure in applications to Bayesian estimation of a new Keynesian macroeconomic model.
Keywords: DSGE model, auxiliary particle filter, Bayesian estimation
JEL classification: C11, C15, C32, C63

1 Introduction

The particle filter (PF) has become an important tool for likelihood-based inference in dynamic stochastic general equilibrium (DSGE) models with non-linear policy functions and non-normal exogenous shocks. PF approximates the intractable likelihood function in the spirit of sequential importance sampling and resampling. The approximated likelihood function is then plugged into the Metropolis-Hastings (MH) algorithm to obtain draws from posterior distribution, or into a derivative-free optimization algorithm to get a simulated maximum likelihood estimator (Fernández-Villaverde and Rubio-Ramírez, 2007).

It is well-known that PF can be quite inefficient when the variance of measurement error is small or in the presence of outliers in data. This explains why a huge number of particles (usually between 10^4 and 10^5) are required to achieve sensible estimation results when taking DSGE models to real data (An and Schorfheide, 2007).

In this article, we propose an efficient auxiliary particle filter (APF) algorithm that reduces variance of likelihood approximation significantly. This algorithm adds a nonlinear

*The Wang Yanan Institute for Studies in Economics (WISE), Xiamen University, P.R.China. E-mail: yuanyang200@aliyun.com.

†The Department of Statistical Science, Duke University, USA. E-mail: wangronglu22@gmail.com.

filtering step in each period of particle filter, leading to as accurate an approximation as the fully-adapted particle filter. This algorithm is robust to data generating mechanisms with different parameter values. It leads to a fast mixing rate in MH algorithm and provides accurate posterior estimators.

2 Particle Filter for Likelihood Approximation

Most dynamic equilibrium models have the following state space representation:

$$\begin{aligned} x_t &= h_\theta(x_{t-1}) + R\epsilon_t, \\ y_t &= g_\theta(x_t) + e_t. \end{aligned} \tag{1}$$

Here x_t is $n_x \times 1$ latent pre-determined state variable with $x_1 \sim \mu_\theta(\cdot)$, y_t is $n_y \times 1$ imperfectly observed economic data, $\epsilon_t \sim N(0, \Sigma_\epsilon)$ is $n_\epsilon \times 1$ exogenous shock to the economic system, $e_t \sim N(0, \Sigma_e)$ is $n_y \times 1$ measurement error, R is a $n_x \times n_\epsilon$ matrix, $\theta \in \Theta$ is vector of structural parameters and $h_\theta : \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ and $g_\theta : \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$ are non-linear functions. In many applications, these policy functions (1) are solved by perturbation method(Schmitt-Grohe and Uribe, 2004). For example, the second-order policy functions have the following form (σ is parameter scales the variance of shocks): [1]

$$\begin{aligned} h_\theta(x) &= \frac{1}{2}h_{\sigma\sigma}(\theta) + h_x(\theta)x + \frac{1}{2}(I_{n_x} \otimes x)^T h_{xx}(\theta)x \\ g_\theta(x) &= \frac{1}{2}g_{\sigma\sigma}(\theta) + g_x(\theta)x + \frac{1}{2}(I_{n_z} \otimes x)^T g_{xx}(\theta)x. \end{aligned}$$

The likelihood function $L(\theta) \equiv p_\theta(y_{1:T})$ of the model is analytically intractable due to non-linear terms in policy functions. Note that $p_\theta(y_{1:T})$ can be decomposed as

$$p_\theta(y_{1:T}) = p_\theta(y_1)\prod_{t=2}^{T} p_\theta(y_t|y_{1:t-1}) \tag{2}$$

and each term in (2) can be represented as

$$p_\theta(y_t|y_{1:t-1}) = \int p_\theta(y_t|x_t)p_\theta(x_t|x_{t-1})p_\theta(x_{t-1}|y_{1:t-1})dx_{t-1:t}. \tag{3}$$

Particle filter approximates $p_\theta(y_{1:T})$ recursively: At period t, PF replaces $p_\theta(x_{t-1}|y_{1:t-1})$ in (3) with the empirical filtering density obtained at period $t-1$

$$\tilde{p}_\theta^N(x_{t-1}|y_{1:t-1}) \triangleq \sum_{j=1}^{N} \bar{w}_{t-1}^{(j)}\delta(x_{t-1} - x_{t-1}^{(j)})$$

and then approximated (3) by importance sampling. A generic auxiliary particle filter follows

[1] The function $g_\theta(\cdot)$ in this article is not the same as that defined in Schmitt-Grohe and Uribe (2004).

1. At $t = 1$, for $i = 1, \ldots, N$, draw $x_1^{(i)} \sim \mu_\theta(x_1)$, let $w_1^{(i)} = p_\theta(y_1|x_1^{(i)})$ and $\bar{w}_1^{(i)} = w_1^{(i)} / \left(\sum_{j=1}^N w_1^{(j)}\right)$. Approximate $p_\theta(y_1)$ by $\hat{p}_\theta^N(y_1) \triangleq N^{-1} \sum_{j=1}^N w_1^{(j)}$.

2. For $t = 2, \ldots, T$,

   (a) For $i = 1, \ldots, N$, let $u_t^{(i)} = \bar{w}_{t-1}^{(i)} \nu_t^{(i)}$, $\bar{u}_t^{(i)} = u_t^{(i)} / \left(\sum_{j=1}^N u_t^{(j)}\right)$ and draw a_t^i from $\{1, \ldots, N\}$ with probabilities $\{\bar{u}_t^{(1)}, \ldots, \bar{u}_t^{(N)}\}$ and $x_t^{(i)} \sim q_\theta(\cdot|x_{t-1}^{(a_t^i)}, y_t)$.

   (b) For $i = 1, \ldots, N$, evaluate the importance weight

   $$w_t^{(i)} = \frac{p_\theta(y_t|x_t^{(i)}) p_\theta(x_t^{(i)}|x_{t-1}^{(a_t^i)})}{q_\theta(x_t^{(i)}|x_{t-1}^{(a_t^i)}, y_t) \nu_t^{(a_t^i)}}, \quad \bar{w}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{j=1}^N w_t^{(j)}}.$$

   (c) Approximate $p_\theta(y_t|y_{1:t-1})$ by $\hat{p}_\theta^N(y_t|y_{1:t-1}) \triangleq (N^{-1} \sum_{j=1}^N w_t^{(j)})(\sum_{j=1}^N u_t^{(j)})$.

3. Approximate the likelihood by $\hat{p}_\theta^N(y_{1:T}) \triangleq \hat{p}_\theta^N(y_1) \prod_{t=2}^T \hat{p}_\theta^N(y_t|y_{1:t-1})$.

In the above PF algorithm, $\nu_t = \nu_\theta(x_{t-1}, y_t)$ is the weight-adjustment multiplier (Pitt and Shephard, 1999) and a_t^i is the index for the ancestor of $x_t^{(i)}$ at period $t-1$. Under mild conditions, $\mathbb{E}[\hat{p}_\theta^N(y_{1:T})] = p_\theta(y_{1:T})$ for any $N \geq 1$ and $\theta \in \Theta$ (Pitt et al., 2012). The unbiasedness property enables us to simulate from the *true* posterior distribution by MH algorithm and achieve *exact* Bayesian inference for θ even though the true likelihood function is unknown (Flury and Shephard, 2011). Since the mixing rate of the resulting particle MH chain highly depends on the Monte Carlo variance of $\hat{p}_\theta^N(y_{1:T})$, large $\mathbb{V}[\hat{p}_\theta^N(y_{1:T})]$ will result in high autocorrelation in the generated posterior samples and hence inaccurate posterior estimates. It is possible that the particle MH scheme is no longer geometrically ergodic in this case, retaining the chain in the same value for a long period due to many rejections of the proposals.

The variance of $\hat{p}_\theta^N(y_{1:T})$ is closely related to the variance of importance weights. According to Doucet et al. (2000), the optimal proposal distribution $q_\theta^*(x_t, |x_{t-1}, y_t) \triangleq p_\theta(x_t, |x_{t-1}, y_t)$ minimizes the variance of importance weights given x_{t-1}. However, even when q_θ^* is employed, important weights $w_t^{(i)} = p_\theta(y_t|x_{t-1}^{(a_t^i)})$ may still be tremendously unevenly distributed if y_t is an outlier, resulting in large variance of w_t. This is of particular concern if the measurement error is small because $p_\theta(y_t|x_t)$ is extremely sensitive to x_t in this case.

To alleviate this problem, auxiliary particle filter (APF) assigns more weights in the resampling step to particles with higher conditional likelihood $p_\theta(y_t|x_{t-1})$ such that those "promising" particles have higher probabilities to survive(Pitt and Shephard, 1999). This is done by multiplying the standard importance weight w_t by an multiplier ν_t which conveys information in observation y_t.[2] When $\nu_\theta^*(x_{t-1}, y_t) \triangleq p_\theta(y_t|x_{t-1})$ and q_θ^* are employed, the weights reduce to all being one and the procedure is called "fully-adapted" particle filter (FAPF). The FAPF is generally the optimal filter and most efficient in estimating the likelihood $p_\theta(y_{1:T})$ (Pitt et al., 2012).

[2] Note that $\nu_t = 1$ in standard particle filter.

In general non-linear DSGE models, neither q_θ^* nor ν_θ^* is analytically available. Amisano and Tristani (2010) suggest a normal approximation to q_θ^* by linearizing $g_\theta(\cdot)$ around forecasted mean of latent state $\mathbb{E}[x_t|y_{1:t-1}]$. As we will see, this approximation is inaccurate as it ignores the rich information in y_t and neglects the linearization error. Andreasen (2011) proposes to draw $x_t \sim q_\theta(x_t)$ independently of x_{t-1}, where $q_\theta(x_t)$ is an approximation to $p_\theta(x_t|y_{1:t})$. This proposal distribution loses efficiency by ignoring the dependence of latent state process. Hall et al. (2014) proposes an APF for state-space models with intractable transition density. This APF is computationally demanding since numerical optimization steps are implemented in each period of filtering. DeJong et al. (2013) develops an iterated least squares procedure to obtain a global approximation to the "unconditional" optimal proposal distribution $p_\theta(x_t, x_{t-1}|y_{1:t})$. This approach introduces bias in likelihood approximation and hence not desirable in Bayesian computation.

As Johansen and Doucet (2008) points out, a preferable approach for APF is to select an approximation $\hat{p}_\theta(y_t, x_t|x_{t-1})$ to the distribution $p_\theta(y_t, x_t|x_{t-1})$ such that the importance weights have an upper bound. When the approximation is accurate, the resulted APF will achieve similar performance as with FAPF. In DSGE models, the moderate non-linearity in measurement equation enables us to linearize $g_\theta(\cdot)$ to obtain a good approximation to FAPF. The small variance of measurement error leads to a highly spiky distribution $p_\theta(y_t|x_t)$ and hence the mode of $p_\theta(y_t|x_t)$ may be far away from $\mathbb{E}[x_t|y_{1:t-1}]$. Therefore a local approximation to $p_\theta(y_t|x_t)$ around $\mathbb{E}[x_t|y_{1:t-1}]$ would be rather inaccurate resulting in large errors in regions where $p_\theta(y_t|x_t)$ is high. To reduce the local approximation error, we suggest linearizing $g_\theta(\cdot)$ around $\mathbb{E}(x_t|y_{1:t})$, which though is unknown before applying particle filter yet can be easily computed by running one extra step of non-linear Kalman filter. We propose the following procedure at period t $(t = 1, \ldots, T)$ of APF:

1. Approximate $\mathbb{E}[x_{t-1}|y_{1:t-1}]$ and $\mathbb{V}[x_{t-1}|y_{1:t-1}]$ with particles $\{x_{t-1}^{(i)}, w_{t-1}^{(i)}\}_{i=1}^N$. Run one step non-linear Kalman filter to get $\mathbb{E}[x_t|y_{1:t}]$.

2. Linearize $g_\theta(x)$ around $\mathbb{E}[x_t|y_{1:t}]$ to get normal approximations to $p_\theta(y_t|x_{t-1})$ and $p_\theta(x_t|x_{t-1}, y_t)$.

3. Implement APF with $\nu_\theta(x_{t-1}, y_t) = \hat{p}_\theta(y_t|x_{t-1})$ and $q_\theta(x_t|x_{t-1}, y_t) = \hat{p}_\theta(x_t|x_{t-1}, y_t)$.

3 Numerical Experiment

This section illustrates the performance of the proposed APF by estimating the new Keynesian model built in An and Schorfheide (2007). The model is solved up to second-order accuracy around the non-stochastic steady state (Schmitt-Grohe and Uribe, 2004). The prior distributions and calibrated parameter values are listed in Table 1 (the parameter $1/g$ is fixed to be 0.85 in the simulation). The competing algorithms to our APF include bootstrap filter (BF) in Fernández-Villaverde and Rubio-Ramírez (2007), conditional particle filter (CPF) in Amisano and Tristani (2010) and central difference particle filter (CDPF) in Andreasen (2011). The proposal distribution in CDPF is multivariate student-t with degree of freedom 15. In APF, the preliminary filtered mean $E[x_t|y_{1:t}]$ is computed by

Table 1: Priors Distribution and Calibrated Values

Name	Distribution	Mean	S.D.	Value
τ	Gamma	3.0	0.50	2.00
ν	Beta	0.1	0.05	0.10
κ	Gamma	0.3	0.01	0.33
ψ_1	Gamma	1.5	0.25	1.50
ψ_2	Gamma	0.5	0.25	0.13
$r^{(A)}$	Gamma	0.8	0.50	1.00
$\pi^{(A)}$	Gamma	4.0	0.50	3.20
$\gamma^{(Q)}$	Gaussian	0.5	0.20	0.55
ρ_R	Beta	0.8	0.15	0.75
ρ_π	Beta	0.8	0.15	0.95
ρ_u	Beta	0.8	0.15	0.90
$100\sigma_R$	Inv-Gamma	0.4	4.00	0.20
$100\sigma_\pi$	Inv-Gamma	0.4	4.00	0.60
$100\sigma_u$	Inv-Gamma	0.4	4.00	0.30

central difference Kalman filter (CDKF), which provides accurate approximation to the filtered moments in an computationally efficient way (Andreasen, 2013).

We first compare the Monte Carlo variance of log-likelihood approximation for different PFs given the calibrated parameters. We generate 100 observations from the model and set the variance of measurement errors to be r times the sample variance of data. Each PF was run 100 times to estimate $\mathbb{V}[\log \hat{p}_\theta^N(y_{1:T}|\theta)]$ given N and r and this procedure is repeatedly implemented 10 times. Table 2 reports the average $\mathbb{V}[\log \hat{p}_\theta^N(y_{1:T}|\theta)]$ across 10 implementations for each method. Besides model with normally distributed shocks, we also consider the case where structural shocks follow student-t distribution to generate outliers in the data set.

Table 2 shows that the APF clearly dominates other PFs in all cases. Even when $r = 0.01$, APF obtains a remarkably low variance of log-likelihood approximation with just 500 particles. BF performs the worst and needs more than 40000 particles to achieve the similar Monte Carlo variance as APF with $N = 500$ when $r = 0.1$. CPF performs better than CDPF if the measurement error is not extremely small. When $r = 0.01$, the performance of CPF improves rather slowly as N increases due to omitting the highly informative y_t in this case. When outriders are introduced in the data set, performances of all PFs deteriorate compared to the normal distribution case. However, CDPF and APF are more robust to outliers due to heavy-tailed proposal distribution and auxiliary variable strategy respectively.

Finally, we illustrate the benefit of the proposed APF for Bayesian parameter estimation. The observations in measurement equation include GDP growth, inflation rate and federal fund rate in the U.S. from 1984Q1 to 2007Q4. Posterior draws are generated by random-walk Metropolis algorithm, and the variance-covariance matrix in proposal distribution is c times the inverse of negative Hessian matrix evaluated at posterior

Table 2: Variance of Log-likelihood Approximation

	Normal, $r = 0.1$				Normal, $r = 0.05$			
$N(\times 10^3)$	BF	CPF	CDPF	APF	BF	CPF	CDPF	APF
0.5	7.944	0.185	1.581	0.085	26.045	0.368	2.319	0.115
1	3.446	0.090	0.820	0.040	12.138	0.192	1.247	0.056
2	1.480	0.044	0.419	0.018	5.661	0.126	0.621	0.028
4	0.812	0.021	0.218	0.010	3.168	0.078	0.300	0.015
10	0.335	0.011	0.092	0.004	1.158	0.031	0.136	0.006
20	0.152	0.005	0.042	0.002	0.741	0.018	0.064	0.003
40	0.087	0.003	0.021	0.001	0.387	0.015	0.034	0.001
	Normal, $r = 0.01$				Student-t, $r = 0.1$			
$N(\times 10^3)$	BF	CPF	CDPF	APF	BF	CPF	CDPF	APF
0.5	560.021	4.931	68.418	0.210	17.464	0.609	2.501	0.280
1	296.200	3.751	21.588	0.130	10.528	0.444	1.198	0.133
2	125.407	2.815	3.127	0.064	6.468	0.265	0.764	0.060
4	68.222	2.410	1.544	0.035	3.287	0.171	0.394	0.028
10	28.985	2.036	0.524	0.018	1.717	0.194	0.217	0.014
20	15.995	1.203	0.295	0.009	0.778	0.099	0.129	0.006
40	10.830	1.275	0.144	0.004	0.582	0.086	0.074	0.003

mode[3]. The optimal scaling of random-walk Metropolis is $c^* = 2.38/\sqrt{dim(\theta)} \approx 0.636$ and the resulting acceptance rate will be around 0.234. For each generated Metropolis chain, the scale parameter c is tuned in order to make the acceptance rate between 0.23 and 0.28. Table 3 reports some convergence diagnostic statistics of Markov chain, calculated from $M = 20000$ posterior draws with the first 2000 as burn-in. Here ASJD $= (M-1)^{-1}\sum_{m=2}^{M} \|\theta^{(m)} - \theta^{(m-1)}\|^2$ measures the average squared jump distance of the generated chain in the parameter space. $\text{IAT}_j = 1 + 2\sum_{l=1}^{1000} \hat{\rho}_j(l)$ measures the autocorrelation of posterior samples for j-th parameter.

In Table 3, APF dominates CPF in convergence of the resulting Markov chain. For CPF, when N is not large enough, the maximum of IATs is high because the rather noisy estimate of likelihood makes it hard for MH algorithm to explore the parameter space thoroughly. Thus it will lead to large Monte Carlo variance in estimating posterior means. In particular, it needs 10000 particles to achieve competitive performance. When $N = 10000$, the approximation to likelihood function is rather accurate and the resulting particle MH algorithm is very close to the idealized algorithm. In fact, 1000 particles is adequate for APF, since the gain by adding number of particles is only marginal. From the comparison, it is clear that APF greatly reduces the computational burden in posterior computation.

[3]The mode is computed by CDKF and Sim's routine.

Table 3: Convergence of Particle MH Algorithm

Filter	$N(\times 10^3)$	c	Acceptance Rate	ASJD	Median of IATs	Max. of IATs
CPF	0.5	0.50	0.24	0.030	120.6	406.4
	1	0.52	0.26	0.035	109.2	426.6
	5	0.55	0.25	0.036	104.7	339.9
	10	0.60	0.23	0.040	91.0	194.7
APF	0.5	0.57	0.27	0.042	71.0	247.7
	1	0.58	0.27	0.042	81.7	142.5
	5	0.60	0.27	0.045	59.8	116.6
	10	0.60	0.27	0.048	61.2	158.3

Note: After posterior sampling, we calculate IAT for each parameter and "Median of IATs" is the median of 14 IATs while "Max. of IATs" is the maximum.

4 Conclusion

This articles proposes an efficient particle filter algorithm for likelihood evaluation in nonlinear dynamic equilibrium models. The proposed algorithm delivers competitive results as the fully-adapted particle filter, which greatly enhances the likelihood-based inference of DSGE models. We illustrate our procedure in applications to Bayesian estimation of Keynesian macroeconomic model.

References

Gianni Amisano and Oreste Tristani. Euro area inflation persistence in an estimated nonlinear DSGE model. *Journal of Economic Dynamics and Control*, 34(10):1837–1858, 2010.

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

Martin M. Andreasen. Non-linear DSGE models and the optimized central difference particle filter. *Journal of Economic Dynamics and Control*, 35(10):1671–1695, 2011.

Martin M. Andreasen. Non-linear DSGE models and the central difference Kalman filter. *Journal of Applied Econometrics*, 28(6):929–955, 2013.

David N. DeJong, Roman Liesenfeld, Guilherme V Moura, Jean-François Richard, and Hariharan Dharmarajan. Efficient likelihood evaluation of state-space representations. *The Review of Economic Studies*, 80(2):538–567, 2013.

Arnaud Doucet, Simon Godsill, and Christophe Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and computing*, 10(3):197–208, 2000.

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

Thomas Flury and Neil Shephard. Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models. *Econometric Theory*, 27(5):933, 2011.

Jamie Hall, Michael K. Pitt, and Robert Kohn. Bayesian inference for nonlinear structural time series models. *Journal of Econometrics*, 179(2):99–111, 2014.

Adam M. Johansen and Arnaud Doucet. A note on auxiliary particle filters. *Statistics & Probability Letters*, 78(12):1498–1504, 2008.

Michael K. Pitt and Neil Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association*, 94(446):590–599, 1999.

Michael K. Pitt, Ralph dos Santos Silva, Paolo Giordani, and Robert Kohn. On some properties of markov chain monte carlo simulation methods based on the particle filter. *Journal of Econometrics*, 171(2):134–151, 2012.

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