# Non-linear DSGE Models and The Optimized Particle Filter*

Martin M. Andreasen†

Bank of England and CREATES

January 27, 2010

**Abstract**

This paper improves the accuracy and speed of particle filtering for non-linear DSGE models with potentially non-normal shocks. This is done by introducing a new proposal distribution which i) incorporates information from new observables and ii) has a small optimization step that minimizes the distance to the optimal proposal distribution. A particle filter with this proposal distribution is shown to deliver a high level of accuracy even with relatively few particles, and this filter is therefore much more efficient than the standard particle filter.

*Keywords:* Likelihood inference, Non-linear DSGE models, Non-normal shocks, Particle filtering.

JEL: C13, C15, E10, E32

*This paper is an improved version of some sections in an earlier paper entitled: "Non-linear DSGE Models, the Central Difference Kalman Filter, and the Mean Shifted Particle Filter".

†I would like to thank Juan Rubio-Ramírez, Oreste Tristani, Bent Jesper Christensen, Tom Engsted, two annonumous referees, and participants at the workshop "Modeling and Forecasting Economic and Financial Time Series with State Space models" at the Swedish Riksbank Oct. 17-18 2008 for comments and discussions. Email: martin.andreasen@bankofengland.co.uk. Telephone number: +44 207 601 3431. I greatly acknowledge financial support from the Danish Center for Scientific Computation (DCSC). I appreciate financial support to Center for Research in Econometric Analysis of Time Series, CREATES, funded by the Danish National Research Foundation. I am also grateful to Juan Rubio-Ramírez for giving me his code of the standard particle filter. Finally, the views expressed herein are solely those of the author and not necessarily those of Bank of England.

# 1 Introduction

Likelihood based inference has long been a standard method for taking linearized DSGE models to the data (see the references in Fernández-Villaverde & Rubio-Ramírez (2005)). The fascinating work by Fernández-Villaverde & Rubio-Ramírez (2007) shows how to do likelihood based inference for non-linear DSGE modes with potentially non-normal shocks. They use a particle filter to approximate the log-likelihood function and the conditional state distributions through repeated use of importance sampling and resampling. We will refer to the particle filter used by Fernándes-Villaverde and Rubio-Ramirez (2005b, 2007a) as the standard Particle Filter (PF). This filter has subsequentially been successfully applied by An (2005), Strid (2006), Doh (2007), and An & Schorfheide (2007) to non-linear DSGE models.

It is well-known that the standard PF suffers from the so-called "sample depletion problem", which means that few particles get a positive weight in the importance sampling step of the filter and this leads to inaccuracy. The sample depletion problem arises because the state transition distribution is used as the proposal distribution in the importance sampling step of the standard PF. This choice of proposal distribution is clearly sub-optimal because information from new observables is omitted. One way to reduce this problem is to increase the number of particles, but a great deal of inaccuracy may still remain. Moreover, this attempted solution comes at the cost of increasing the computational requirements which are already severe when using the standard PF to estimate non-linear DSGE models.

The purpose of this paper is to improve the accuracy and speed of particle filtering for non-linear DSGE models. We suggest introducing a new and more efficient proposal distribution which is constructed from two elements. The first being output from a non-linear deterministic filter which provides a preliminary estimate of the first and second moments in the posterior state distribution, given the previous state distribution and new observables. These estimates are then used to form a Gaussian proposal distribution which therefore contains information from new observables.

The second element is a free parameter that scales the first estimate of the covariance matrix

in the proposal distribution. This parameter allows us to increase the variance of the proposal distribution to account for fat tails and other deviations from normality in the posterior state distribution. We introduce the novel idea to determine the value of this parameter by minimizing the distance between our proposal distribution and the optimal proposal distribution. We show how to implement this optimization step in a very fast manner, and we provide evidence that the optimization step greatly improves the performance of the proposal distribution. The particle filter with the resulting new proposal distribution is therefore referred to as the Optimized Particle Filter (OPF).

A Monte Carlo study tests the performance of the OPF on a standard New Keynesian DSGE model approximated up to second order. We highlight the following results. Firstly, our new proposal distribution is shown to be a good approximation of the unknown state distribution, and this leads to much lower Monte Carlo variation in the estimated log-likelihood function than in the standard PF. Secondly, the OPF is 3 to 105 times more efficient than the standard PF when we account for differences in accuracy and computing time of the two filters. We also show that the OPF outperforms the Sigma Point Particle Filter by Merwe, Doucet, de Freitas & Wan (2000) with even higher efficiency gains. Finally, the OPF has a low Monte Carlo variation in the log-likelihood function with relative few particles. This result is robust to the presence of small measurement errors and non-normal shocks to the economy. As a result, the OPF greatly facilitates likelihood inference when taking non-linear DSGE models to the data.

The rest of the paper is organized as follows. We present the state space representation of DSGE models in section 2. Particle filters are discussed in section 3, where we describe the standard PF and derive the new OPF. Section 4 sets up a DSGE model which is calibrated to account for higher order moments in the post-war US economy. The performance of the various filters are then examined in a Monte Carlo study in section 5. Concluding comments are provided in section 6.

## 2 The state space representation of DSGE models

We consider the class of economic models which can be represented in a dynamic state space system (see for instance Thomas F. Cooley (1995) and Schmitt-Grohé & Uribe (2004) for a number of illustrations). The set of observables in period $t$ is denoted by the vector $\mathbf{y}_t$ which has dimension $n_y \times 1$. These observables are a function of the state vector $\mathbf{x}_t$ and a random vector $\mathbf{v}_t \sim \mathcal{IID}(\mathbf{0}, \mathbf{R}_{\mathbf{v}}(t))$. We let $\mathbf{x}_t$ have dimension $n_x \times 1$ and $\mathbf{v}_t$ have dimension $n_v \times 1$. More formally,

$$\mathbf{y}_t = \mathbf{g}(\mathbf{x}_t, \mathbf{v}_t; \boldsymbol{\theta}). \tag{1}$$

The function $\mathbf{g}(\cdot)$ is determined by the parameters $\boldsymbol{\theta} \in \Theta$ in the economic model and the equilibrium conditions describing the economy. The equations in (1) are known as the set of measurement equations. Lagged values of the state vector can be included in these equations by extending the state vector appropriately as shown in section 5.1.

The law of motion for the state vector is given by

$$\mathbf{x}_t = \mathbf{h}(\mathbf{x}_{t-1}, \mathbf{w}_t; \boldsymbol{\theta}), \tag{2}$$

where $\mathbf{w}_t \sim \mathcal{IID}(\mathbf{0}, \mathbf{R}_{\mathbf{w}}(t))$ is a random vector of structural shocks with dimension $n_w \times 1$. The equations in (2) are typically referred to as the set of transition equations. Our notation with one lag in these equations is without loss of generality, because additional lags can be added to the transition equations by extending the state vector. The state vector is assumed to be unobserved, but observed state variables can be introduced by letting one or more elements in $\mathbf{g}(\cdot)$ be identity mappings. All the vectors $\mathbf{y}_t$, $\mathbf{x}_t$, $\mathbf{v}_t$, and $\mathbf{w}_t$ are assumed to have continuous support.

The state vector in many economic models can often be decomposed as

$$\mathbf{x}_t \equiv \begin{bmatrix} \mathbf{x}_{1,t} \\ \mathbf{x}_{2,t} \end{bmatrix}, \tag{3}$$

where $\mathbf{x}_{1,t}$ denotes the endogenous state variables and $\mathbf{x}_{2,t}$ are the exogenous state variables, i.e. the shocks hitting the economy. The dimension of these vectors are $n_{x_1} \times 1$ and $n_{x_2} \times 1$, respectively, and $n_{x_1} + n_{x_2} = n_x$. This implies that the set of transition equations can be decomposed as

$$\mathbf{x}_{1,t} = \mathbf{h}_1 (\mathbf{x}_{t-1}; \boldsymbol{\theta}) \tag{4}$$

$$\mathbf{x}_{2,t} = \mathbf{h}_2 (\mathbf{x}_{2,t-1}, \mathbf{w}_t; \boldsymbol{\theta}) . \tag{5}$$

We do not impose any specific probability distributions on $\mathbf{v}_t$ or $\mathbf{w}_t$. However, independence between $\mathbf{v}_t$ and $\mathbf{w}_t$ is assumed at all leads and lags.

# 3 Particle filters

The objective in particle filters is to recursively estimate the probability distribution of the unknown state vector $\mathbf{x}_{0:t} \equiv \{\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_t\}$ given all information at a certain point in time, $\mathbf{y}_{1:t} \equiv \{\mathbf{y}_1, ..., \mathbf{y}_t\}$. This posterior state distribution is denoted by $p(\mathbf{x}_{0:t} | \mathbf{y}_{1:t}; \boldsymbol{\theta})$, and the unknown state vector is typically estimated by the mean of the marginal distribution $p(\mathbf{x}_t | \mathbf{y}_{1:t}; \boldsymbol{\theta})$. The probability $p(\mathbf{y}_{t+1} | \mathbf{y}_{1:t}; \boldsymbol{\theta})$ and hence the likelihood function can also be computed from the posterior state distribution. The recursive estimation of $p(\mathbf{x}_{0:t} | \mathbf{y}_{1:t}; \boldsymbol{\theta})$ is based on the probability structure of the state space system in (1) and (2) and sequential use of importance sampling and resampling.

The outline for the remaining part of this section is as follows. We proceed in section 3.1 by describing the basic algorithm for particle filters as presented in Doucet, Godsill & Andrieu (2000). Section 3.2 presents the standard PF as an example of this basic algorithm and we discuss the advantages and disadvantages of this filter. A few extensions of the standard PF are then briefly discussed in section 3.3 to motivate the construction of the Optimized Particle Filter (OPF) in section 3.4.

We adopt the standard notation in terms of the state vector $\mathbf{x}_t$ throughout the presentation. Hence, to avoid stochastic singularity it is assumed that the dimension of $\mathbf{v}_t$ is equal to the

number of observables. At the expense of a more evolved notation, some of the structural shocks in $\mathbf{w}_t$ can be used to avoid stochastic singularity, as shown by Fernández-Villaverde & Rubio-Ramírez (2007).

### 3.1 The basic algorithm for particle filters

We require that the two conditional probabilities $p(\mathbf{y}_t|\mathbf{x}_t;\boldsymbol{\theta})$ and $p(\mathbf{x}_t|\mathbf{x}_{t-1};\boldsymbol{\theta})$ can be evaluated for all values of $\mathbf{x}_t$ and $\mathbf{y}_t$ for $t = 1, ..., T$.[1] Using the well-known result we have

$$p(\mathbf{y}_t|\mathbf{x}_t;\boldsymbol{\theta}) = p(\mathbf{v}_t;\boldsymbol{\theta}) \left|\det\left(\frac{\partial \mathbf{g}(\mathbf{x}_t,\mathbf{v}_t;\boldsymbol{\theta})}{\partial \mathbf{v}_t}\right)\right|^{-1} \tag{6}$$

$$p(\mathbf{x}_t|\mathbf{x}_{t-1};\boldsymbol{\theta}) = p(\mathbf{w}_t;\boldsymbol{\theta}) \left|\det\left(\frac{\partial \mathbf{h}(\mathbf{x}_{t-1},\mathbf{w}_t;\boldsymbol{\theta})}{\partial \mathbf{w}_t}\right)\right|^{-1}, \tag{7}$$

provided that the Jacobian of $\mathbf{g}(\cdot)$ and $\mathbf{h}(\cdot)$ exist and their determinants are non-zero. We henceforth assume that the distributions in (6) and (7) are well-defined.

It is straightforward to show that the posterior state distribution has a recursive form given by

$$p(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1};\boldsymbol{\theta}) = p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t};\boldsymbol{\theta}) \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1};\boldsymbol{\theta})\, p(\mathbf{x}_{t+1}|\mathbf{x}_t;\boldsymbol{\theta})}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t};\boldsymbol{\theta})}. \tag{8}$$

It is in general not possible to calculate $p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t};\boldsymbol{\theta})$ or sample from $p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t};\boldsymbol{\theta})$. These problems can be solved by using importance sampling and the approximation

$$p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t};\boldsymbol{\theta}) \approx \sum_{i=1}^{N} w_t^{(i)} \delta\left(\mathbf{x}_{0:t} - \mathbf{x}_{0:t}^{(i)}\right), \tag{9}$$

where $\delta(\cdot)$ denotes the Dirac delta function and $\left\{\mathbf{x}_{0:t}^{(i)}\right\}_{i=1}^{N}$ are random draws from $p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t};\boldsymbol{\theta})$. Each element $\mathbf{x}_{0:t}^{(i)}$ is referred to as a particle and is assigned the weight $w_t^{(i)}$. The proposal distribution for the importance sampling is denoted $\pi(\cdot)$ and assumed to be given by

$$\pi(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1}) = \pi(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})\, \pi(\mathbf{x}_{t+1}|\mathbf{x}_{0:t}, \mathbf{y}_{1:t+1}). \tag{10}$$

[1] In the standard PF, it is only required that we can sample from $p(\mathbf{x}_t|\mathbf{x}_{t-1};\boldsymbol{\theta})$. However, DSGE models are usually specified with a known parameteric distribution for the structural shocks, and this ensures that no additional assumptions are needed to evaluate $p(\mathbf{x}_t|\mathbf{x}_{t-1};\boldsymbol{\theta})$.

The structure of $\pi(\cdot)$ implies that $p(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1};\boldsymbol{\theta})$ is approximated without modifying past estimated state values, $\mathbf{x}_{0:t}$. The specific functional form of $\pi(\cdot)$ should be chosen such that its support includes the posterior state distribution. This proposal distribution gives rise to the following recursive formula for the importance sampling weights $w_{t+1}^{(i)}$;

$$w_{t+1}^{(i)} = w_t^{(i)} \frac{p\left(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^{(i)};\boldsymbol{\theta}\right) p\left(\mathbf{x}_{t+1}^{(i)}\middle|\mathbf{x}_t^{(i)};\boldsymbol{\theta}\right)}{\pi\left(\mathbf{x}_{t+1}^{(i)}\middle|\mathbf{x}_{0:t}^{(i)},\mathbf{y}_{1:t+1}\right)} \qquad \text{for } i=1,...,N. \tag{11}$$

The normalized importance samplings weights are given by

$$\tilde{w}_{t+1}^{(i)} = \frac{w_{t+1}^{(i)}}{\sum_{i=1}^N w_{t+1}^{(i)}} \qquad \text{for } i=1,...,N. \tag{12}$$

A random sample from $p(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1};\boldsymbol{\theta})$ is then generated by sampling with replacement from $\left\{\mathbf{x}_{0:t+1}^{(i)}\right\}_{i=1}^N$ with probabilities $\left\{\tilde{w}_t^{(i)}\right\}_{i=1}^N$. This new sample is denoted by $\left\{\hat{\mathbf{x}}_{0:t+1}^{(i)}\right\}_{i=1}^N$, and the particles in this sample have uniform weights, i.e. $w_{t+1}^{(i)} = 1/N$. If this resampling step is omitted, then the unconditional variance of $w_t$ increases over time, and after a few iterations only one particle will have a non-zero weight. This means that a large number of particles essentially are removed from the approximation of the posterior state distribution which therefore deteriorates in precision. The purpose of the resampling step is to mitigate this problem by eliminating particles that are too far from the true state vector (i.e. particles with low values of $\tilde{w}_{t+1}^{(i)}$) and multiply particles which are close to the true state vector (i.e. particles with high values of $\tilde{w}_{t+1}^{(i)}$).

For any function $f(\mathbf{x}_{0:t+1})$ which is integrable with respect to $p(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1};\boldsymbol{\theta})$, it holds that

$$E[f(\mathbf{x}_{0:t+1})] = \frac{1}{N}\sum_{i=1}^N f\left(\hat{\mathbf{x}}_{0:t+1}^{(i)}\right). \tag{13}$$

A central limit theorem for this estimator is provided by Berzuini, Best, Gilks & Larizza (1997).[2]

[2]Liu & Chen (1998) recommend to do state estimation before the resampling step, i.e. by $E[f(\mathbf{x}_{0:t+1})] = \sum_{i=1}^N \tilde{\omega}_{t+1}^{(i)} f\left(\mathbf{x}_{0:t+1}^{(i)}\right)$, because the resampling step introduces additional random variation in the sample of particles. However, the estimator in (13) is computationally faster than the one recommended by Liu & Chen (1998), because $\frac{1}{N}\sum_{i=1}^N f\left(\hat{\mathbf{x}}_{0:t+1}^{(i)}\right) = \frac{1}{N}\sum_{i=1}^n N_i f\left(\tilde{\mathbf{x}}_{0:t+1}^{(i)}\right)$ where $N_i$ is the number of repetitions of the $i'$th particle and

Finally, the contribution to the likelihood function can be estimated by

$$p\left(\mathbf{y}_{t+1}|\,\mathbf{y}_{1:t};\boldsymbol{\theta}\right)=\sum_{i=1}^{N}w_{t+1}^{(i)}. \tag{14}$$

The choice of proposal distribution is clearly crucial for the performance of particle filters. Doucet et al. (2000) show that the following proposal distribution

$$\pi\left(\mathbf{x}_{t+1}|\,\mathbf{x}_{0:t},\mathbf{y}_{1:t+1}\right)=p\left(\mathbf{x}_{t+1}|\,\mathbf{x}_{t},\mathbf{y}_{t+1};\boldsymbol{\theta}\right) \tag{15}$$

is optimal in the sense that it minimizes the variance of the importance sampling weights given $\mathbf{x}_{0:t}$ and $\mathbf{y}_{1:t+1}$. This proposal distribution is in general intractable and approximations are therefore needed.

### 3.2 The standard particle filter

The standard PF approximates the optimal proposal distribution by the transition distribution for the states. That is

$$\pi\left(\mathbf{x}_{0:t+1}|\,\mathbf{y}_{1:t+1}\right)=p\left(\mathbf{x}_{t+1}|\,\mathbf{x}_{t};\boldsymbol{\theta}\right). \tag{16}$$

This is an obvious choice for the three reasons: i) it is easy to sample from $p\left(\mathbf{x}_{t+1}|\,\mathbf{x}_{t};\boldsymbol{\theta}\right)$, ii) the probability of $\mathbf{x}_{t+1}$ is conditioned on $\mathbf{x}_{t}$, and iii) the importance sampling weights are given by $w_{t+1}^{(i)}=w_{t}^{(i)}p\left(\mathbf{y}_{t+1}|\,\mathbf{x}_{t+1}^{(i)};\boldsymbol{\theta}\right)$. However, the state transition distribution does not use information from new observables (i.e. $\mathbf{y}_{t+1}$), and this proposal distribution is therefore said to be "blind" (Pitt & Shephard (1999)). It is in this sense that the proposal distribution in the standard PF is sub-optimal.

Omitting information from new observables in the proposal distribution is unfortunate for two reasons. Firstly, if new observations are very informative about the state vector, valuable information is not present in the proposal distribution. This situation can occur if small changes in the state vector generate large changes in the observables, and/or the observables are measured

$\left\{\tilde{\mathbf{x}}_{0:t+1}^{(i)}\right\}_{i=1}^{n}$ is the set of different resampled particles. This latter estimator is faster to compute because $n$ is must smaller than $N$.

with a high signal to noise ratio. Secondly, using a blind proposal distribution makes the standard PF very sensitive to state outliers, because the posterior state distribution in this case is centered in the tail of the state transition distribution. Hence, state outliers generate a poor support overlap between the proposal distribution and the posterior state distribution. The importance sampling weights in the standard PF may therefore be very unevenly distributed in such a situation, and as a result many particles are needed to get a satisfying approximation of $p\left(\mathbf{x}_{0:t+1}|\,\mathbf{y}_{1:t+1};\boldsymbol{\theta}\right)$.

### 3.3 Two alternatives to the standard particle filter

There exist numerous alternatives to the standard PF (see Doucet, de Freitas & Gordon (2001)). Many of these alternatives try to improve the performance of the standard PF by bringing information from new observables into the proposal distribution. We focus on two alternatives which do so by using deterministic filters.

The first alternative we consider is that suggested by Doucet et al. (2000) who use the Extended Kalman Filter (EKF) (Jazwinski (1970)) to generate a Gaussian approximation of the optimal proposal distribution in (15). This is done by sending *each* particle through one iteration in the EKF to generate a new probability distribution, where the mean and the covariance matrix thus contain information from new observables. The idea is that sampling from these distributions moves particles to areas of high probability. More formally,

$$\pi\left(\mathbf{x}_{t+1}|\,\mathbf{x}_t^{(i)},\mathbf{y}_{t+1}\right)=\mathcal{N}\left(\hat{\mathbf{x}}_{t+1}^{EKF,(i)},\hat{\mathbf{P}}_{\mathbf{xx}}^{EKF,(i)}\left(t+1\right)\right)\qquad\text{for } i=1,...,N. \tag{17}$$

We use the notation $\hat{\mathbf{x}}_{t+1}^{EKF,(i)}$ to denote the posterior mean in the EKF for particle $i$, and $\hat{\mathbf{P}}_{\mathbf{xx}}^{EKF,(i)}\left(t+1\right)$ to denote the posterior covariance matrix for this state estimate. Using this proposal distribution in the basic algorithm for particle filters leads to the Extended Kalman Particle Filter (EKPF). Doucet et al. (2000) and Merwe et al. (2000) show that the proposal distribution in (17) gives more precise state estimates compared to state estimates from the standard PF. The EKPF has recently been used by Amisano & Tristani (2007) in the context

of non-linear DSGE models.

Two drawbacks are related to (17). Firstly, the EKPF is very time consuming to implement, because the mean and the covariance matrix in (17) must be calculated for a large number of particles each time period.[3] Secondly, the approximations of the first and second moments in the EKF are only accurate up to first and second order, respectively, and the approximations do not take the probability distribution of the state vector into account.

Merwe et al. (2000) and Merwe & Wan (2003) suggest to solve this second problem by using the Central Difference Kalman filter (CDKF) by Norgaard, Poulsen & Ravn (2000) to calculate more accurate expressions for the mean and the covariance matrix in (17). The CDKF is an extension of the standard Kalman filter to non-linear and non-normal state space systems where the non-linear moments in the filtering equations are approximated at least up to second order accuracy by multivariate Stirling interpolations. These interpolations are computed using a deterministic sampling approach and hence no derivatives are needed in the CDKF.[4] As a result, (17) is replaced by

$$\pi\left(\mathbf{x}_{t+1}|\,\mathbf{x}_t^{(i)},\mathbf{y}_{t+1}\right)=\mathcal{N}\left(\hat{\mathbf{x}}_{t+1}^{CDKF,(i)},\hat{\mathbf{P}}_{\mathbf{xx}}^{CDKF,(i)}\left(t+1\right)\right)\qquad\text{for } i=1,...,N \tag{18}$$

Here, $\hat{\mathbf{x}}_{t+1}^{CDKF,(i)}$ denotes the posterior mean in the CDKF for particle $i$ and $\hat{\mathbf{P}}_{\mathbf{xx}}^{CDKF,(i)}\left(t+1\right)$ denotes the posterior covariance matrix for this state estimate. Using this proposal distribution in the basic algorithm for particle filters leads to the Sigma Point Particle Filter (SPPF). In an application Merwe et al. (2000) show that the SPPF clearly outperforms the EKPF. However, like the EKPF, the SPPF is very time consuming to compute.

[3] The resampling implies that some of the particles in $\left\{\hat{\mathbf{x}}_t^{(i)}\right\}_{i=1}^{N}$ are identical. Hence, only particles with different values need to be send through one iteration in the EKF, and this number of particles is often lower than $N$.

[4] The CDKF uses the same number of function evaluations as in the EKF with numerically computed derivatives.

### 3.4 The Optimized Particle Filter

This section suggests a computational less demanding method than the one used in the EKPF and the SPPF for bringing information from new observables into the proposal distribution. In doing so, we introduce a third approximation of the optimal proposal distribution in (15). Our new proposal distribution consists of two components; the first relates to the endogenous state variables and the second to the exogenous state variables.

Starting with the endogenous state variables, we suggest sending the sample of particles with uniform weights $\left\{\hat{\mathbf{x}}_{0:t}^{(i)}\right\}_{i=1}^{N}$ through the transition function as in the standard PF. This is done to preserve the singularity of the transition function in the proposal distribution. Hence, the proposal draws for the endogenous state variables are given by

$$\mathbf{x}_{1,t+1}^{(i)} = \mathbf{h}_1\left(\hat{\mathbf{x}}_t^{(i)}; \boldsymbol{\theta}\right) \quad \text{for } i = 1, ..., N. \tag{19}$$

The second component of our proposal distribution uses information from the CDKF when generating proposal draws for the exogenous state variables. The basic idea is to use the posterior mean estimate $\hat{\mathbf{x}}_{2,t+1}^{CDKF}$ and the square root of the covariance matrix for this estimate $\hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}(t+1)$ to form a Gaussian proposal distribution. That is, draws for $\mathbf{x}_{2,t+1}$ are generated by

$$\mathbf{x}_{2,t+1}^{(i)} = \hat{\mathbf{x}}_{2,t+1}^{CDKF} + \gamma_{t+1}\hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}(t+1)\,\boldsymbol{\epsilon}_{t+1}^{(i)} \quad \text{for } i = 1, ..., N. \tag{20}$$

We first note that $\hat{\mathbf{x}}_{2,t+1}^{CDKF}$ and $\hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}(t+1)$ contain information from new observables and this should bring the proposal distribution close to the posterior state distribution. Contrary to the EKPF and SPPF, information from new observables is here incorporated in a very efficient way because only one iteration of the CDKF is needed in each time period to compute $\hat{\mathbf{x}}_{2,t+1}^{CDKF}$ and $\hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}(t+1)$.

The novel feature in (20) is the scalar $\gamma_{t+1}$ which is a free parameter. The role of this parameter is to account for fat tails and other deviations from normality in the posterior state distribution by increasing the variance of the proposal distribution. We suggest to determine

the value of $\gamma_{t+1}$ by numerically minimizing the variance of the importance sampling weights in each time period. The optimal proposal distribution $p\left(\mathbf{x}_{t+1}|\,\mathbf{x}_t, \mathbf{y}_{t+1}; \boldsymbol{\theta}\right)$ implies a variance of zero for these sampling weights, and our choice of $\gamma_{t+1}$ can therefore be interpreted as minimizing the distance between our proposal distribution and the optimal proposal distribution. We recommend to restrict $\gamma_{t+1} \geq 1$ when determining $\gamma_{t+1}$, because values of $\gamma_{t+1}$ less than 1 may easily lead to the undesirable property of thinner tails in the proposal distribution than in the posterior state distribution.

It is convenient to express the minimization problem for $\gamma_{t+1}$ in terms of the effective sample size $N_{eff,t+1}$ in order to evaluate whether the variance in the importance sampling weights is sufficiently low. The effective sample size provides a measure of how many particles effectively approximate the posterior state distribution and is given by (Arulampalam, Maskell, Gordon & Clapp (2002))

$$N_{eff,t+1} = \frac{N}{1 + Var_{\pi}\left(w_{t+1}\left(\gamma_{t+1}\right)\right)} = \frac{N}{E_{\pi}\left(w_{t+1}\left(\gamma_{t+1}\right)^2\right)} \tag{21}$$

where $w_{t+1}\left(\gamma_{t+1}\right)$ denotes the importance sampling weight as a function of $\gamma_{t+1}$. The standard estimate of $N_{eff,t+1}$ is

$$\hat{N}_{eff,t+1} = \frac{1}{\sum_{i=1}^{N}\left(\tilde{w}\left(\gamma_{t+1}\right)_{t+1}^{(i)}\right)^2}, \tag{22}$$

where $\tilde{w}\left(\gamma_{t+1}\right)_{t+1}^{(i)}$ is the normalized importance sampling weight. For instance, $\hat{N}_{eff,t+1} = 1$ if only one particle gets a positive weight in approximating the desired distribution, whereas $\hat{N}_{eff,t+1} = N$ in the optimal case where all particles get a positive weight. Hence, minimizing the variance of the importance sampling weights is equivalent to maximizing the effective sample size, and we can therefore let

$$\gamma_{t+1}^{*} = \arg\max \frac{1}{\sum_{i=1}^{N}\left(\tilde{w}\left(\gamma_{t+1}\right)_{t+1}^{(i)}\right)^2} \tag{23}$$

The computational cost of determining $\gamma_{t+1}$ through optimization can be greatly reduced by using only $N_{opt} << N$ particles in the optimization. We suggest to choose these $N_{opt}$ particles

from $\left\{\hat{\mathbf{x}}_{0:t}^{(i)}\right\}_{i=1}^{N}$ by selecting every $n$'th particle where $n \equiv \lceil N/N_{opt} \rceil$. That is, the particles at positions $\mathcal{J} = \{1, 1+n, 1+2n, ...\}$.[5] This implies

$$\gamma_{t+1}^{*} = \arg\max \frac{1}{\sum_{j \in \mathcal{J}} \left( \tilde{w} \left( \gamma_{t+1} \right)_{t+1}^{(j)} \right)^2}, \tag{24}$$

which often is a good approximation to (23). However, if $\hat{N}_{eff,t+1}$ is considered too low after this optimization, then we suggest to redo the optimization using all particles in a second attempt to improve the proposal distribution.

An overview of the OPF is given in Appendix A. Note here that we use $\left\{\hat{\mathbf{x}}_{0:t+1}^{(i)}\right\}_{i=1}^{N}$ to compute an improved estimate of the mean posterior state estimate which is used in the CDKF. The estimate of $\hat{\mathbf{S}}_{\mathbf{x}}^{CDKF}(t+1)$ could be improved in a similar way but is omitted because it requires much additional computation.

Prediction of the observables in the OPF is straightforward, and smoothing can be undertaken alone the lines suggested by Simon J. Godsill & West (2004). The values for the unknown structural parameters in the economy $\boldsymbol{\theta}$ can be estimated by Maximum Likelihood or Bayesian methods (see Fernández-Villaverde & Rubio-Ramírez (2007), An & Schorfheide (2007) and Flury & Shephard (2008))

# 4 A New Keynesian DSGE model

This section presents the DSGE model which we will use in the Monte Carlo study. This model has the same basic structure as the models in Altig, Christiano, Eichenbaum & Linde (2005), Christiano, Eichenbaum & Evans (2005), and Smets & Wouters (2007). We use the standard notation from the macroeconomic literature when presenting our model, and the notation in this section is therefore unrelated to that used for the filters in the previous section.

[5] We use the notation $\lceil x \rceil$ to denote the nearest integer greater than or equal to $x \in \mathbb{R}$. If $N/n \neq N_{opt}$, then we select the last $N/n - N_{opt}$ particles as $2, 2+n, 2+2n, ....$

**The households:** A representative household has preferences specified by

$$U_t = E_t \sum_{l=0}^{\infty} \beta^l \frac{\left( \left( \frac{c_{t+l} - b x_{t-1+l}}{z^*_{t+l}} \right)^{1-\phi_2} (1 - h_{t+l})^{\phi_2} \right)^{1-\phi_1} - 1}{1-\phi_1},$$

where $c_t$ is consumption and $h_t$ is labor supply. The variable $z^*_t$ is a measure of technological progress and controls the overall trend in consumption. The parameter $b$ specifics external habit formation in the consumption good $c_t$ which is constructed from

$$c_t = \left[ \int_0^1 c_{i,t}^{\frac{\eta-1}{\eta}} di \right]^{\frac{\eta}{\eta-1}}.$$

The habit stock $x_t$ evolves according to $x_{t+1} = \rho_x x_t + (1 - \rho_x) c_t$.

The first constraint on the household is the law of motion for the physical capital stock $k_t$ given by

$$k_{t+1} = (1 - \delta) k_t + i_t \left( 1 - \frac{\kappa}{2} (\frac{i_t}{i_{t-1}} - \mu_i)^2 \right).$$

where $i_t$ is gross investments. The value of $\mu_i$ is determined such that there are no adjustment costs along the economy's balanced growth path.

The second constraint is the household's real period by period budget constraint

$$E_t D_{t,t+1} x^h_{t+1} + c_t + (e_t \Upsilon_t)^{-1} i_t = \frac{x^h_t}{\pi_t} + w_t h_t + \phi_t. \tag{25}$$

The left hand side of (25) is the household's total expenditures in period $t$ which consists of i) state-contingent claims $E_t D_{t,t+1} x^h_{t+1}$, ii) consumption $c_t$, and iii) investments $(e_t \Upsilon_t)^{-1} i_t$. Changes in $e_t \Upsilon_t$ are investment specific shocks, which evolves as an exogenous AR(1) process along a deterministic trend, i.e. $\ln \Upsilon_{t+1} = \ln \Upsilon_t + \ln \mu_{\Upsilon,ss}$ and

$$\ln e_{t+1} = \rho_e \ln e_t + \epsilon_{e,t+1}.$$

We let $\epsilon_{e,t+1} \sim \mathcal{IID}(0, Var(\epsilon_{e,t+1}))$.

The right hand side of (25) is the household's total wealth in period $t$. It consists of: i) pay-off from state-contingent assets purchased in the previous period $x_t^h/\pi_t$, ii) real labor income $w_t h_t$, and iii) dividends received from firms $\phi_t$. Note that $\pi_t$ is the gross inflation rate.

**The firms:** Production is carried out by a continuum of firms indexed by $i \in [0,1]$. They supply a differentiable good $y_{i,t}$ to the goods market which is characterized by monopolistic competition. All firms have access to the same technology

$$y_{i,t} = \begin{cases} k_{i,t}^{\theta} \left(a_t z_t h_{i,t}\right)^{1-\theta} - \psi z_t^* & \text{if } k_{i,t}^{\theta} \left(a_t z_t h_{i,t}\right)^{1-\theta} - \psi z_t^* > 0 \\ 0 & \text{else} \end{cases} \tag{26}$$

where $k_{i,t}$ and $h_{i,t}$ denote physical capital and labor services, respectively. The variable $a_t$ represents stationary technology shocks, and we let

$$\ln a_{t+1} = \rho_a \ln a_t + \epsilon_{a,t+1},$$

where $\epsilon_{a,t+1} \sim \mathcal{IID}\left(0, Var\left(\epsilon_{a,t+1}\right)\right)$. The variable $z_t$ in (26) denotes non-stationary technology shocks. We let $\mu_{z,t} \equiv z_t/z_{t-1}$ and assume

$$\ln \mu_{z,t+1} = \ln \mu_{z,ss} + \epsilon_{z,t+1},$$

where $\epsilon_{z,t+1} \sim \mathcal{IID}\left(0, Var\left(\epsilon_{z,t+1}\right)\right)$. The shocks $a_t$ and $\mu_{z,t}$ are mutually independent, and so are all other shocks in the model. Following Altig et al. (2005), we define $z_t^* \equiv \Upsilon^{\theta/(1-\theta)} z_t$. The amount of fixed production costs $\psi$ is set to ensure a steady state profit of zero.

All firms maximize the present value of their nominal dividend payments, denoted $d_{i,t}$:

$$d_{i,t} \equiv E_t \sum_{l=0}^{\infty} D_{t,t+l} P_{t+l} \phi_{i,t+l},$$

where $D_{t,t+l}$ is the stochastic discount factor and the expression for real dividend payments from the $i$'th firm $\phi_{i,t}$ is given below in (27). The firms face a number of constraints when maximizing

$d_{i,t}$. The first is related to the good produced by the $i$'th firm. The total amount of good $i$ is allocated to consumption and investments, which implies

$$y_t = c_t + (e_t \Upsilon_t)^{-1} i_t.$$

The second constraint is the budget restriction which gives rise to the expression for real dividends from firm $i$ in period $t$

$$\phi_{i,t} = (P_{i,t}/P_t)\, y_{i,t} - r_t^k k_{i,t} - w_t h_{i,t}. \tag{27}$$

The first term in (27) denotes the real revenue from sales of the $i$'th good. The next two terms in (27) are the firm's expenditures on capital services $r_t^k k_{i,t}$ and payments to workers $w_t h_{i,t}$.

The third constraint introduces staggered price adjustments. We make the standard assumption that in each period a fraction $\alpha \in [0, 1[$ of randomly selected firms are not allowed to set the optimal nominal price of the good they produce. Instead, these firms set the current prices equal to the prices in the previous period, i.e. $P_{i,t} = P_{i,t-1}$ for all $i \in [0, 1]$

**The central bank:** We let the central bank determine the gross one period nominal interest rate $R_t$ according to the rule

$$\ln\left(\frac{R_t}{R_{ss}}\right) = \rho_R \ln\left(\frac{R_{t-1}}{R_{ss}}\right) + \beta_\pi \ln\left(\frac{\pi_t}{\pi_{ss}}\right) + \beta_y \ln\left(\frac{y_t}{y_{ss} z_t^*}\right) + \epsilon_{R,t+1},$$

where $\epsilon_{R,t+1} \sim \mathcal{IID}\left(0, Var\left(\epsilon_{R,t+1}\right)\right)$.

### 4.1 The approximated state space representation

The exact solution to our model is given by

$$\mathbf{y}_t = \mathbf{g}\left(\mathbf{x}_t\right) \tag{28}$$

$$\mathbf{x}_{t+1} = \mathbf{h}\left(\mathbf{x}_t\right) + \boldsymbol{\eta}\mathbf{w}_{t+1} \tag{29}$$

where $\mathbf{y}_t$ contains the non-predetermined variables and $\mathbf{x}_t$ is the predetermined state vector. Both variables are expressed in deviation from the deterministic steady state. The functions $\mathbf{g}(\cdot)$ and $\mathbf{h}(\cdot)$ are unknown, and we therefore approximate them up to second order and apply the pruning scheme to this approximation (Schmitt-Grohé & Uribe (2004), Kim, Kim, Schaumburg & Sims (2008)).

Five macro variables are chosen for the Monte Carlo study: i) the quarterly nominal interest rate, ii) the quarterly inflation rate, and the quarterly real growth rates in iii) consumption, iv) investments, and v) output. These series are placed in the vector $\mathbf{y}_t^{obs}$. We allow for measurement errors $\mathbf{v}_t$ in the series for $\mathbf{y}_t^{obs}$ and assume $\mathbf{v}_t \sim \mathcal{NID}(\mathbf{0}, \mathbf{R_v})$ where $\mathbf{R_v}$ is a diagonal matrix. This gives the following state space system

$$\mathbf{y}_t^{obs} = \mathbf{M}_1\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta}) - \mathbf{M}_2\mathbf{g}(\mathbf{x}_{t-1}, \boldsymbol{\theta}) + \mathbf{v}_t \tag{30}$$

$$\mathbf{x}_{t+1} = \mathbf{h}(\mathbf{x}_t) + \boldsymbol{\eta}\mathbf{w}_{t+1} \tag{31}$$

where $\mathbf{M}_1$ and $\mathbf{M}_2$ are selection matrices with appropriate dimensions. The presence of $\mathbf{x}_{t-1}$ in (30) is due to the three growth rates in $\mathbf{y}_t^{obs}$ as shown in Appendix B.

### 4.2 Model calibration

The model is calibrated to US data from 1956Q4 to 2009Q2 in order to make the Monte Carlo study as realistic as possible. We focus on matching i) mean values, ii) standard deviations, iii) skewness, and iv) kurtosis for the five series in $\mathbf{y}_t^{obs}$. All calibrated coefficients are fairly standard and summarized in Table 1. A comment is in order in relation to the size of the measurement errors in $\mathbf{y}_t^{obs}$. We conjecture that the quarterly interest and inflation rates are measured quite accurately, and we therefore let these measurement errors have a standard deviation of 10 basis points. The three growth rates are assumed to be measured less accurately, and we therefore let measurement errors for these variables have a standard deviation of 20 basis points.

< Table 1 about here >

All structural shocks are assumed to be normally distributed in our benchmark case. The empirical and simulated moments are reported in Table 2 which shows that our model is successful at matching all mean values and standard deviations. The model also generates sizeable deviations from normality in the nominal interest rate and the inflation rate. This is evident from the values of skewness and kurtosis for these series which are larger than 0 and 3, respectively.

We also consider the case where all structural shocks have a Laplace distribution. This is a symmetric distribution with thicker tails than the normal distribution. Table 2 shows that these alternative shocks increase kurtosis for all series and thus bring these moments closer to the empirical moments.

< Table 2 about here >

# 5 A Monte Carlo study

This section uses the DSGE model presented above to test the performance of the OPF compared to the standard PF and the SPPF. We focus on the ability of the three filters to reduce the Monte Carlo variation in the log-likelihood function because this is the most important statistic to consider when estimating non-linear DSGE models using particle filters.

The outline for this section is as follows. We start in section 5.1 by describing the details related to the implementation of the Monte Carlo study. Section 5.2 compares the performance of the OPF to the standard PF and the SPPF for the benchmark case with normally distributed shocks. Robustness analysis is undertaken in section 5.3 which examines the effects of having small measurement errors in the observables and Laplace distributed shocks.

## 5.1 Study design

We deal with the presence of $\mathbf{x}_{t-1}$ in the measurement equations by extending the state vector to $\tilde{\mathbf{x}}_t \equiv \begin{bmatrix} \mathbf{x}_t' & \mathbf{x}_{t-1}' \end{bmatrix}'$. It is unnecessary to evaluate $\left\{ \mathbf{g}\left( \mathbf{x}_{t-1}^{(i)}, \boldsymbol{\theta} \right) \right\}_{i=1}^{N}$ in each time period as in Fernández-Villaverde & Rubio-Ramírez (2007) if we store these values and resample

$\left\{\mathbf{g}\left(\mathbf{x}_t^{(i)}, \boldsymbol{\theta}\right)\right\}_{i=1}^{N}$ along with $\left\{\mathbf{x}_t^{(i)}\right\}_{i=1}^{N}$ in the resampling step. The transition equations for $\tilde{\mathbf{x}}_t$ is

$$\begin{bmatrix} \mathbf{x}_{t+1} \\ \mathbf{x}_t \end{bmatrix} = \begin{bmatrix} \mathbf{h}\left(\mathbf{x}_t\right) \\ \mathbf{x}_t \end{bmatrix} + \begin{bmatrix} \boldsymbol{\eta}\mathbf{w}_{t+1} \\ \mathbf{0} \end{bmatrix}.$$

For a given number of particles $N$, the Monte Carlo variability in the log-likelihood function is estimated from 20 simulated time series $i = 1, 2, ..., 20$ each with a sample length of $T = 200$. For each of these time series, we then run the filters 100 times using different draws from the respective probability distributions to compute the log-likelihood function denoted $\mathcal{L}_{i,j}^N$ for $j = 1, 2, ..., 100$. These 100 evaluations enable us to compute the standard deviation in the estimated log-likelihood function $std\left(\mathcal{L}_i^N\right)$ for the $i'$th simulated time series. Averaging across the 20 simulated time series, we then estimate the Monte Carlo variation in the log-likelihood by

$$std\left(\mathcal{L}^N\right) = \frac{1}{20}\sum_{i=1}^{20} std\left(\mathcal{L}_i^N\right),$$

where $std\left(\mathcal{L}^N\right)$ denotes the standard deviation in the estimated log-likelihood function using $N$ particles. The initial state vector $\tilde{\mathbf{x}}_0$ is assumed known and we initialize $\mathbf{P}_{\mathbf{xx}}\left(t = 0\right)$ to $10^{-6}$.

The optimization step for $\gamma_t$ in the OPF is implemented by a simple line search algorithm. We set the number of particles for the optimization to $N_{opt} = \min\left(N/10, 1000\right)$. If the effective sample size after this optimization is below 100, then we redo the optimization using all particles. The importance sampling weights for the OPF are derived in Appendix C.

Finally, all particle filters are implemented with systematic resampling as in Fernández-Villaverde & Rubio-Ramírez (2007).

### 5.2 The benchmark case with normally distributed shocks

We begin by comparing the performance of the OPF to the standard PF for the benchmark case with normally distributed shocks. The performance of the OPF when $\gamma_t = 1$ in all time periods is also reported to examine the effect of the optimization step. Output from this filter is denoted OPF($\gamma = 1$). We study the performance of each filter along three dimensions: i) the

Monte Carlo variation in the log-likelihood function, ii) the effective sample size, and iii) the computing time. We also examine how the performance of the various filters relate to the number of shocks driving the economy. This is done by gradually increasing the number of shocks from one to four. That is, we let $\sqrt{Var\left(\epsilon_{k,t}\right)} = 0$; first for $k = \{a, e, R\}$, then for $k = \{e, R\}$, $k = \{R\}$, and finally for $k = \varnothing$.

The performance of the OPF compared to the SPPF is deferred to section 5.2.2.

< Figure 1 about here >

The first chart in the upper left corner of Figure 1 shows the variation in the log-likelihood function when only one shock drives the economy. This variation is much larger in the standard PF than in both versions of the OPF. The difference is reduced as the number of particles increases from 500 to 5,000, but the OPF clearly dominates the performance of the standard PF for all considered choices of particles. We also note that the OPF($\gamma = 1$) performs just as well as the OPF when there is one shock to the economy, and this must imply that the optimal value of $\gamma_t$ is close to 1. An inspection of the time series for $\gamma_t$ across the sample length and across all $20 \times 100 = 2000$ evaluations of the filter verifies this conjecture, because this time series has a mean value of 1.04 and a standard deviation of 0.02.[6]

The OPF also dominates the performance of the standard PF when two shocks drives the economy. This is seen from the upper right corner of Figure 1. We also note that the OPF has a lower Monte Carlo variation than OPF($\gamma = 1$), meaning that the performance of the OPF is improved when $\gamma_t$ is optimally determined. The importance of letting $\gamma_t$ vary is also evident from the time series of $\gamma_t$ which has a mean of 1.25 and a standard deviation of 0.14 when two shocks drive the economy. It is also interesting to note that the OPF displays a remarkable low variation in the log-likelihood function with just 1,000 particles. As a result, increasing the number of particles does not induce a similar large absolute reduction in the variation of the log-likelihood function as in the standard PF. However, when we measure the gain in accuracy from more particles in relative terms, that is as $std\left(\mathcal{L}^{N=1,000}\right) / std\left(\mathcal{L}^{N=10,000}\right)$, then the OPF has a relative gain of 2.78 which is similar to the relative gain of 2.97 for the standard PF.

[6] These numbers vary slightly with the number of particles.

The same conclusions hold when three and four shocks drive the economy. That is, the OPF outperforms the standard PF, and optimally determining the value of $\gamma_t$ improves the performance of the OPF. Again, the OPF has a remarkably low variation in the log-likelihood function with relative few particles.

< Figure 2 about here >

The only difference between the standard PF and the OPF is the choice of proposal distribution which therefore explains the improved performance of the OPF. As mentioned in section 3.4, the effective sample size measures the quality of a proposal distribution, and the OPF should therefore have a larger effective sample size than the standard PF. We examine this hypothesis by computing the average effective sample size across the sample length and across all 2000 evaluations of a given filter, i.e.

$$\bar{N}_{eff} = \frac{\sum_{i=1}^{20} \sum_{j=1}^{100} \sum_{t=1}^{T} N_{eff,t}^{(i,j)}}{20 \times 100 \times T}.$$

Figure 2 shows that the average effective sample size in the OPF, as expected, is orders of magnitudes higher than the average effective sample size for the standard PF. Hence, the new proposal distribution suggested in this paper has a much better overlap with the posterior state distribution than the state transition distribution used in the standard PF. We also note from Figure 2 that optimally determining the value of $\gamma_t$ greatly increases the average effective sample size when two or more shocks drive the economy.

It is also interesting to look at the minimum value of the effective sample size because it shows whether the approximation of the posterior state distribution is close to a collapse around a single particle at some point during the filtering. Such a collapse is unfortunate and may lead to large variation in the estimated log-likelihood function even thought the average effective

sample size is high.[7] We therefore construct the following statistic

$$N_{eff}^{\min} = \frac{\sum_{i=1}^{20}\sum_{j=1}^{100} \min\left(\left\{N_{eff,t}^{(i,j)}\right\}_{t=1}^{T}\right)}{20 \times 100}$$

which is an average of the minimum effective sample size across all 2000 evaluations of a given filter. Figure 3 shows a very low value of $N_{eff}^{\min}$ in the standard PF, meaning that the approximation of the state distribution is close to a collapse.[8] On the other hand, the value of $N_{eff}^{\min}$ in the OPF is much higher and shows no sign of a collapse. Note also that with three and four shocks to the economy, this is due to the optimization step for $\gamma_t$ which has a sizeable effect on the value of $N_{eff}^{\min}$.

< Figure 3 about here >

To summarize, the new proposal distribution suggested here greatly reduces the Monte Carlo variation in the log-likelihood function compared to the variation in the standard PF. Inspection of the effective sample size shows that this improvement arises because the unknown state distribution is better approximated by our proposal distribution than the state transition distribution used in the standard PF.

### 5.2.1 Comparing the efficiency gain of the OPF

Proposal distributions which condition on new observables are typical very time consuming to implement, and it is therefore interesting to compare the computing time for the OPF to the standard PF. This is done in Figure 4 which reports the average number of seconds it takes to evaluate the filters. We first note that the OPF without the optimization step, i.e. OPF($\gamma = 1$), is more time consuming to compute than the standard PF. This is due to the more elaborate expression for the importance sampling weights. Adding the optimization step to the OPF increases the computing time even further.

[7] We thanks an annonumous referee for making this point.

[8] We did not experience any cases where the standard PF or any other filter lost track of the state distribution and diverged.

< Figure 4 about here >

A natural question is whether the improved accuracy of the OPF outweights its higher computing time compared to the standard PF. In other words, should a researcher use the somewhat slower OPF with relative few particles or the faster standard PF with a very large number of particles? One way to answer this question is to estimate the number of particles needed in the standard PF to achieve the same accuracy as in the OPF, and then compare the computing time for the two alternatives. The OPF should be preferred if it is faster to compute than the standard PF with the required number of particles, and visa versa. The first step in doing this comparison is to estimate how the number of particles affect i) the accuracy and ii) the computing time in the two filters. We deal with each of these relations in turn.

The relationship between the accuracy of a filter $std\left(\mathcal{L}_i^N\right)$ and the number of particles $N$ as displayed in Figure 1 seems to be approximately given by a power function

$$std\left(\mathcal{L}_i^N\right) = \beta_1 N^{-\beta_2}, \tag{32}$$

where $\beta_1, \beta_2 > 0$. That is, the Monte Carlo variation decreases gradually with an increasing number of particles. Note also that this power function has desirable limiting properties because $std\left(\mathcal{L}_i^N\right) \longrightarrow 0$ for $N \longrightarrow \infty$ and $std\left(\mathcal{L}_i^N\right) \longrightarrow \infty$ for $N \longrightarrow 0$. Applying a log-transformation and adding an error term $\epsilon_i \sim \mathcal{IID}\left(0, \sigma_\epsilon^2\right)$ gives

$$\ln std\left(\mathcal{L}_i^N\right) = \ln \beta_1 - \beta_2 \ln N + \epsilon_i. \tag{33}$$

This regression model is straightforward to estimate by pooled OLS using the 20 simulated time series $(i = 1, 2, ..., 20)$ and the various number of particles for each of the filters. The fitted values of $std\left(\mathcal{L}_i^N\right)$ from this regression model are very close to the averages $std\left(\mathcal{L}^N\right)$ reported in Figure 1, which leaves support for the chosen specification in (33).

The relationship between the number of particles and the computing time is well-approximated

by a linear function. We therefore let

$$sec_{i,N} = \alpha_1 + \alpha_2 N + \xi_i$$

where $\xi_i \sim \mathcal{IID}\left(0, \sigma_\xi^2\right)$ and $sec_{i,N}$ denotes the average number of seconds for evaluating a filter from one of the 20 simulated time series. This regression model is estimated by pooled OLS in a similar manner as (33).

< Figure 5 about here >

Given the estimates of $(\beta_1, \beta_2, \alpha_1, \alpha_2)$, it is straightforward to compute the following efficiency measure

$$\Psi^k(x) = \frac{\text{number of seconds for the standard PF to give } std\left(\mathcal{L}^N\right)^{PF} = x}{\text{number of second for filter k to give } std\left(\mathcal{L}^N\right)^k = x} \tag{34}$$

where $k = \{\text{OPF}(\gamma = 1), \text{OPF}\}$ and $x$ is a real number. To understand how we display this efficiency measure in Figure 5, consider the case for OPF with 1 shock to the economy. The accuracy of the OPF with 500 particles is 0.41, and the standard PF is estimated to require 18,100 particles to achieve the same level of accuracy. Computing the standard PF with this number of particles is estimated to take 4.62 seconds. This is about 18 times slower than the OPF which only uses 0.26 seconds. Hence, values above the dotted line at 1 indicate that the OPF is preferred to the standard PF, and visa verse for points below the dotted line.

We draw three conclusions from Figure 5. Firstly, the OPF is always more efficient than the standard PF. Secondly, the biggest efficiency gain appear with two shocks and 1,000 particles where the OPF is 35 times more efficient than the standard PF. The case where the efficiency gain is lowest is with three and four shocks to the economy and 50,000 particles. Here, the OPF is 3 times more efficient than the standard PF. Thirdly, with three and four shocks to the economy, the OPF is only more efficient than the standard PF because of the optimization step for $\gamma_t$. Again, this shows the importance of our novel optimization step in the proposal distribution.

### 5.2.2 Comparing the OPF to the SPPF

An alternative to the standard PF is the SPPF as mentioned in section 3.3. Although this filter is very time consuming to compute, it might be that relative few particles in the SPPF could be sufficient to give a low Monte Carlo variation in the log-likelihood function. If so, the SPPF could be an attractive alternative to our OPF. We therefore compare the performance of the OPF to the SPPF with relatively few particles in Figure 6. The SPPF is surprisingly found to be only marginally better than the OPF, and the OPF even outperforms the SPPF when two shocks drive the economy.

< Figure 6 about here >

The two filters are, however, very different in terms of their computing time. The time it takes to compute the SPPF is increasing linearly from 18 seconds with 400 particles to 45 seconds with 1000 particles. The corresponding numbers for the OPF are 0.5 seconds and 1.5 seconds, respectively. This implies that the OPF displays exceptional high efficiency gains compared to the SPPF as shown in Figure 7. For instance, the OPF is about 400 times more efficient than the SPPF when two shocks drive the economy. We therefore conclude that the OPF is also preferred to the SPPF.

< Figure 7 about here >

## 5.3 Robustness analysis

This section examines the robustness of the previous findings in a setting with i) small measurement errors in the observations and ii) non-normal structural shocks. Only the robustness of the OPF to the standard PF is examined given the relatively less favorable results for the SPPF shown in the previous section.

### 5.3.1 Small measurement errors

A modification of the Monte Carlo study to small measurement errors in the observables is interesting because such errors lead to a very peaked probability distribution of $\mathbf{y}_t^{obs}$ given the states, and this makes it more demanding for particle filters to approximate the posterior state distribution. We model small measurement errors as a 50% reduction in the standard deviation of all measurement errors as given in Table 1. This implies that the quarterly interest and inflation rates have measurement errors with a standard deviation of just 5 basis points, and the errors in the quarterly growth rates have a standard deviation of 10 basis points. These small measurement errors are not unusual in the literature. For instance, annualized interest rates are often found to be measured with errors having a standard deviation of $4 \times 5 = 20$ basis points (see for instance Graeve, Emiris & Wouters (2009)).

Figure 8 shows the variation in the log-likelihood function with small measurement errors. We first note that with two or more shocks to the economy, the performance of the standard PF deteriorates compared to the benchmark case. This weaker performance is also evident from the effective sample size for the standard PF (not reported), which is even lower compared to the benchmark case. Hence, omitting information from new observables in the proposal distribution is costly with small measurement errors because few draws from the proposal distribution apparently hit the very peaked probability distribution of $\mathbf{y}_t^{obs}$ given the states.

The variation in the log-likelihood function from the OPF is largely unchanged with small measurement errors when compared to the benchmark case with larger measurement errors in Figure 1. The effective sample size for the OPF is also unchanged (not reported), and this shows the great value of conditioning on new observables in the proposal distribution when measurement errors are small.

< Figure 8 about here >

The computing time for the two filters with small measurement errors are similar to those in Figure 4. This implies that the efficiency measure in Figure 9 provides even further support for the OPF than in the benchmark case.

< Figure 9 about here >

### 5.3.2 Non-normal shocks

The second modification of our Monte Carlo study is to have non-normal shocks driving the economy. We consider the case where shocks are generated from the Laplace distribution which has thicker tails than the normal distribution. This scenario is interesting because it implies that large shocks occur more often than with normally distributed shocks, and this allow us to study whether the OPF is more robust to state outliers than the standard PF. The standard deviation for the measurement errors are as in the benchmark case.

Figure 10 shows that Laplace distributed shocks increase the variation in the log-likelihood function for the standard PF compared to the benchmark case, and we thus confirm that the standard PF is sensitive to state outliers. The performance of the OPF is unaffected compared to the benchmark case, and the variation in the log-likelihood function remains low regardless of the number of shocks and particles.

< Figure 10 about here >

The computing time for the two filters with Laplace distributed shocks are similar to those reported in Figure 4. This implies that the efficiency measure in Figure 11 provides an even stronger argument in favor of the OPF compared to the benchmark case. For instance, with two shocks to the economy and 1,000 particles, the OPF is now 105 times more efficient than the standard PF.

< Figure 11 about here >

## 6 Conclusion

This paper improves the accuracy and speed of particle filtering for non-linear DSGE models by introducing the OPF. The defining feature of this particle filter is a new proposal distribution

which incorporates information from new observables and has a small optimization step that minimizes the distance to the optimal proposal distribution. We show that our proposal distribution is a good approximation of the unknown state distribution, and this leads to much lower Monte Carlo variation in the log-likelihood function compared to the standard PF. The OPF is further shown to be 3 to 105 times more efficient than the standard PF. These substantial efficiency gains arise from the fact that the OPF has a low variation in the log-likelihood function even with relative few particles. This result holds even with small measurement errors in the observables and non-normal shocks to the economy.

Although we have focused on applying the OPF to non-linear DSGE models, the filter may also be useful for estimating other economic models such as dynamic term structure models and stochastic volatility models. In a purely filter theoretical context, our new proposal distribution may also be useful in various extensions of the general algorithm for particle filters. Obvious applications are in i) the Marginal Particle Filter by Klaas, Freitas & Doucet (2005), ii) the Adaptive Particles Filters by Fox (2001) and Soto (2005), which estimate the number of particles to be used each time period, and iii) in particle filters which include a MCMC step (Gilks & Berzuini (2001)) or a kernel smoothing procedure (Musso, Oudjane & LeGland (2001)) to generate more variation in the approximated posterior state distribution.

# A The algorithm for the OPF

We use the notation $[\mathbf{x}_t, \mathbf{S}_{\mathbf{x}}(t)] = CDKF(\mathbf{S}_{\mathbf{x}}(t-1), \mathbf{x}_{t-1}, \mathbf{y}_t)$ to denote one iteration of the CDKF from time point $t-1$ to time point $t$ based on $\mathbf{S}_{\mathbf{x}}(t-1)$, $\mathbf{x}_{t-1}$, and $\mathbf{y}_t$.

**The Optimized Particle Filter (OPF)**

- Initialization: $t = 0$

  For $i = 1, ..., N$ draw particles $\hat{\mathbf{x}}_0^{(i)}$ from $p(\mathbf{x}_0)$ and let $w_0^{(i)} = \frac{1}{N}$ for all $i$. The posterior state estimate: $\hat{\mathbf{x}}_t = \frac{1}{N}\sum_{i=1}^{N} \hat{\mathbf{x}}_t^{(i)}$

- For $t > 0$

  1. Importance sampling step
     - $\left[\hat{\mathbf{x}}_t^{CDKF}, \hat{\mathbf{S}}_{\mathbf{x}}^{CDKF}(t)\right] = CDKF\left(\hat{\mathbf{S}}_{\mathbf{x}}^{CDKF}(t-1), \hat{\mathbf{x}}_{t-1}, \mathbf{y}_t\right)$
     - Let $\mathbf{x}_{1,t}^{(i)} = \mathbf{h}_1\left(\hat{\mathbf{x}}_{t-1}^{(i)}; \boldsymbol{\theta}\right)$ for $i = 1, ..., N$
     - Determine the value of $\gamma_t$ by numerical optimization
     - Draw particles $\mathbf{x}_{2,t}^{(i)}$ from $\mathcal{N}\left(\mathbf{x}_{2,t}^{(i)} \middle| \hat{\mathbf{x}}_{2,t}^{CDKF}, \gamma_t \hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}(t)\right)$ for $i = 1, ..., N$
     - Evaluate the importance sampling weights:
       $w_t^{(i)} = w_{t-1}^{(i)} \frac{p\left(\mathbf{y}_t \middle| \mathbf{x}_t^{(i)}; \boldsymbol{\theta}\right) p\left(\mathbf{x}_t^{(i)} \middle| \hat{\mathbf{x}}_{t-1}^{(i)}; \boldsymbol{\theta}\right)}{\mathcal{N}\left(\mathbf{x}_t^{(i)} \middle| \hat{\mathbf{x}}_{2,t}^{CDKF}, \gamma_t \hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}(t)\right)}$ for $i = 1, ..., N$
     - The contribution to the log-likelihood function: $L_t = L_{t-1} + \log(\sum_{i=1}^{N} w_t^{(i)})$
     - For $i = 1, ..., N$ compute $\tilde{w}_t^{(i)} = w_t^{(i)} / \sum_{i=1}^{N} w_t^{(i)}$
  2. Resampling step:
     - Resample with replacement from $\left\{\mathbf{x}_{0:t}^{(i)}\right\}_{i=1}^{N}$ with probabilities $\left\{\tilde{w}_t^{(i)}\right\}_{i=1}^{N}$ to obtain a samples of size $N$ approximately distributed according to $p(\mathbf{x}_{0:t} | \mathbf{y}_{1:t}; \boldsymbol{\theta})$. This new sample is denoted by $\left\{\hat{\mathbf{x}}_{0:t}^{(i)}\right\}_{i=1}^{N}$
     - In $\left\{\hat{\mathbf{x}}_t^{(i)}\right\}_{i=1}^{N}$ we have $w_t^{(i)} = \frac{1}{N}$ for all $i = 1, ..., N$
  3. State estimates
     - The posterior state estimate: $\hat{\mathbf{x}}_t = \frac{1}{N}\sum_{i=1}^{N} \hat{\mathbf{x}}_t^{(i)}$

# B Calculating the five observables for the Monte Carlo study

This section shows how to calculate the five observables in the Monte Carlo study. The presence of non-stationary shocks ($z_t$ and $\Upsilon_t$) imply that variables such as $c_t$, $i_t$, and $y_t$ are non-stationary, and this fact must be taken into account when solving the DSGE model. We adopt the standard method to deal with this feature by approximating the model's solution around the economy's balanced growth path. This is done by scaling the non-stationary variables such that they

become stationary. For instance, $c_t$ and $y_t$ are scaled by $1/z_t^*$ and $i_t$ is scaled by $1/\Upsilon_t z_t^*$, and this implies that $C_t \equiv c_t/z_t^*$, $Y_t \equiv y_t/z_t^*$, and $I_t \equiv i/\Upsilon_t z_t^*$ are stationary variables.

Applying a log-transformation, the output from the approximated DSGE model is then

$$\mathbf{y}_t \equiv \begin{bmatrix} \hat{R}_t \\ \hat{\pi}_t \\ \hat{C}_t \\ \hat{I}_t \\ \hat{Y}_t \end{bmatrix} = \mathbf{g}\left(\mathbf{x}_t\right),$$

where we use the standard notation that a hat denotes deviation from the deterministic steady state, i.e. $\hat{v}_t = \ln \frac{v_t}{v_{ss}}$. The elements in $\mathbf{y}_t$ must be transformed to make them comparable to empirical data series. We now show how this transformation is done.

For the nominal interest rate we have

$$R_t^{obs} = \ln R_{ss} + \hat{R}_t,$$

where $R_t^{obs}$ is the observed quarterly net interest rate. The expression for observed quarterly inflation rate is similar, i.e.

$$\pi_t^{obs} = \ln \pi_{ss} + \hat{\pi}_t.$$

The expressions for real quarterly consumption growth is given by

$$\begin{aligned} \ln \mu_{c,t}^{obs} &\equiv \ln \frac{c_t}{c_{t-1}} = \ln \frac{C_t z_t^*}{C_{t-1} z_{t-1}^*} = \ln \frac{C_t}{C_{ss}} - \ln \frac{C_{t-1}}{C_{ss}} + \ln \frac{z_t^*}{z_{t-1}^*} \\ &= \hat{C}_t - \hat{C}_{t-1} + \ln \mu_{z^*,ss} + \hat{\mu}_{z^*,t} \end{aligned}$$

For the quarterly growth rate in investments we have

$$\begin{aligned} \ln \mu_{i,t}^{obs} &\equiv \ln \frac{i_t}{i_{t-1}} = \ln \frac{I_t z_t^* \Upsilon_t}{I_{t-1} z_{t-1}^* \Upsilon_{t-1}} = \ln \frac{I_t}{I_{ss}} - \ln \frac{I_{t-1}}{I_{ss}} + \ln \frac{z_t^*}{z_{t-1}^*} + \ln \frac{\Upsilon_t}{\Upsilon_{t-1}} \\ &= \hat{I}_t - \hat{I}_{t-1} + \ln \mu_{z^*,ss} + \hat{\mu}_{z^*,t} + \ln \mu_{\Upsilon,ss} + \hat{\mu}_{\Upsilon,t} \end{aligned}$$

Finally, for the real quarterly growth rate in output

$$\begin{aligned} \ln \mu_{y,t}^{obs} &\equiv \ln \frac{y_t}{y_{t-1}} = \ln \frac{Y_t z_t^*}{Y_{t-1} z_{t-1}^*} = \ln \frac{Y_t}{Y_{ss}} - \ln \frac{Y_{t-1}}{Y_{ss}} + \ln \frac{z_t^*}{z_{t-1}^*} \\ &= \hat{Y}_t - \hat{Y}_{t-1} + \ln \mu_{z^*,ss} + \hat{\mu}_{z^*,t} \end{aligned}$$

Adding measurement errors $\mathbf{v}_t$ to the observables, we then have

$$\mathbf{y}_t^{obs} \equiv \begin{bmatrix} R_t^{obs} \\ \pi_t^{obs} \\ \ln \mu_{c,t}^{obs} \\ \ln \mu_{i,t}^{obs} \\ \ln \mu_{y,t}^{obs} \end{bmatrix} + \mathbf{v}_t.$$

# C The importance sampling weights in the OPF

$\pi\left(\mathbf{x}_t|\,\mathbf{x}_{0:t-1},\mathbf{y}_{1:t}^{obs}\right):\left[\begin{array}{c}\mathbf{x}_{1,t}\\ \mathbf{x}_{2,t}\end{array}\right]=\left[\begin{array}{c}\mathbf{h}_1\left(\mathbf{x}_{t-1}\right)\\ \hat{\mathbf{x}}_t^{CDKF}\end{array}\right]+\left[\begin{array}{c}\mathbf{0}\\ \gamma_t\hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}\left(t\right)\boldsymbol{\epsilon}_t\end{array}\right]$ with $\mathbf{v}_t\sim\mathcal{NID}\left(\mathbf{0},\mathbf{R}_{\mathbf{v}}\right)$ and $\boldsymbol{\epsilon}_t\sim\mathcal{NID}\left(\mathbf{0},\mathbf{I}\right)$. We let $n_e=\dim\left(\boldsymbol{\epsilon}_t\right)=\dim\left(\mathbf{w}_{2,t}\right)$ and $n_{x_2}=n_e$.

**Normally distributed structural shocks**

We have $\boldsymbol{\eta}\mathbf{w}_t\sim\mathcal{NID}\left(\mathbf{0},\left[\begin{array}{cc}\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{R}_{\mathbf{w}_2}\end{array}\right]\right)$ which implies

$w_t=w_{t-1}\frac{p\left(\mathbf{y}_t^{obs}|\tilde{\mathbf{x}}_t;\boldsymbol{\theta}\right)p\left(\tilde{\mathbf{x}}_t|\tilde{\mathbf{x}}_{t-1};\boldsymbol{\theta}\right)}{\pi\left(\tilde{\mathbf{x}}_t|\tilde{\mathbf{x}}_{0:t-1},\mathbf{y}_{1:t}^{obs}\right)}$

$$\begin{aligned}
&=w_{t-1}\times\left(2\pi\right)^{-n_y/2}\det\left(\mathbf{R}_{\mathbf{v}}\right)^{-1/2}\exp\left\{-0.5\left(\mathbf{y}_t^{obs}-\mathbf{g}\left(\tilde{\mathbf{x}}_t;\boldsymbol{\theta}\right)\right)'\mathbf{R}_{\mathbf{v}}^{-1}\left(\mathbf{y}_t^{obs}-\mathbf{g}\left(\tilde{\mathbf{x}}_t;\boldsymbol{\theta}\right)\right)\right\}\\
&\times\left(2\pi\right)^{-n_{x_2}/2}\det\left(\mathbf{R}_{\mathbf{w}_2}\right)^{-1/2}\exp\left\{-0.5\left(\mathbf{x}_{2,t}-\mathbf{h}_2\left(\mathbf{x}_{2,t-1};\boldsymbol{\theta}\right)\right)'\mathbf{R}_{\mathbf{w}_2}^{-1}\left(\mathbf{x}_{2,t}-\mathbf{h}_2\left(\mathbf{x}_{2,t-1};\boldsymbol{\theta}\right)\right)\right\}\\
&\times\left(2\pi\right)^{n_e/2}\det\left(\gamma_t\hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}\left(t\right)\gamma_t\hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}\left(t\right)'\right)^{1/2}\\
&\times\exp\left\{0.5\left(\hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}\left(t\right)\boldsymbol{\epsilon}_t\right)'\left(\hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}\left(t\right)\hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}\left(t\right)'\right)^{-1}\left(\hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}\left(t\right)\boldsymbol{\epsilon}_t\right)\right\}
\end{aligned}$$

$$\begin{aligned}
&=w_{t-1}\times\left(2\pi\right)^{-n_y/2}\det\left(\gamma_t\hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}\left(t\right)\right)\det\left(\mathbf{R}_{\mathbf{v}}\right)^{-1/2}\det\left(\mathbf{R}_{\mathbf{w}}\right)^{-1/2}\\
&\times\exp\left\{-0.5\left(\mathbf{y}_t^{obs}-\mathbf{g}\left(\tilde{\mathbf{x}}_t;\boldsymbol{\theta}\right)\right)'\mathbf{R}_{\mathbf{v}}^{-1}\left(\mathbf{y}_t^{obs}-\mathbf{g}\left(\tilde{\mathbf{x}}_t;\boldsymbol{\theta}\right)\right)\right\}\\
&\times\exp\left\{-0.5\left(\mathbf{x}_{2,t}-\mathbf{h}_2\left(\mathbf{x}_{2,t-1};\boldsymbol{\theta}\right)\right)'\mathbf{R}_{\mathbf{w}_2}^{-1}\left(\mathbf{x}_{2,t}-\mathbf{h}_2\left(\mathbf{x}_{2,t-1};\boldsymbol{\theta}\right)\right)\right\}\\
&\times\exp\left\{0.5\left(\boldsymbol{\epsilon}_t\right)'\boldsymbol{\epsilon}_t\right\}
\end{aligned}$$

**Laplace distributed structural shocks**

We have $\boldsymbol{\eta}\mathbf{w}_t\sim Laplace\left(\mathbf{0},\left[\begin{array}{cc}\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{R}_{\mathbf{w}_2}\end{array}\right]\right)$ which implies

$w_t=w_{t-1}\frac{p\left(\mathbf{y}_t^{obs}|\tilde{\mathbf{x}}_t;\boldsymbol{\theta}\right)p\left(\tilde{\mathbf{x}}_t|\tilde{\mathbf{x}}_{t-1};\boldsymbol{\theta}\right)}{\pi\left(\tilde{\mathbf{x}}_t|\tilde{\mathbf{x}}_{0:t-1},\mathbf{y}_{1:t}^{obs}\right)}$

$$\begin{aligned}
&=w_{t-1}\times\left(2\pi\right)^{-n_y/2}\det\left(\mathbf{R}_{\mathbf{v}}\right)^{-1/2}\exp\left\{-0.5\left(\mathbf{y}_t^{obs}-\mathbf{g}\left(\tilde{\mathbf{x}}_t;\boldsymbol{\theta}\right)\right)'\mathbf{R}_{\mathbf{v}}^{-1}\left(\mathbf{y}_t^{obs}-\mathbf{g}\left(\tilde{\mathbf{x}}_t;\boldsymbol{\theta}\right)\right)\right\}\\
&\times\frac{1}{\left(\sqrt{2}\right)^{n_{x_2}}\prod_{i=1}^{n_{x_2}}\sqrt{R_{w_2}(i,i)}}\exp\left\{-\sqrt{2}\sum_{i=1}^{n_{x_2}}\frac{\left|x_{n_{x_1}+i,t}-h_{n_{x_1}+i}(\mathbf{x}_{2,t-1};\boldsymbol{\theta})\right|}{\sqrt{R_{w_2}(i,i)}}\right\}\\
&\times\left(2\pi\right)^{n_e/2}\det\left(\gamma_t\hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}\left(t\right)\gamma_t\hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}\left(t\right)'\right)^{1/2}\\
&\times\exp\left\{0.5\left(\hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}\left(t\right)\boldsymbol{\epsilon}_t\right)'\left(\hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}\left(t\right)\hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}\left(t\right)'\right)^{-1}\left(\hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}\left(t\right)\boldsymbol{\epsilon}_t\right)\right\}
\end{aligned}$$

$$\begin{aligned}
&=w_{t-1}\times\left(2\pi\right)^{(n_e-n_y)/2}\det\left(\mathbf{R}_{\mathbf{v}}\right)^{-1/2}\exp\left\{-0.5\left(\mathbf{y}_t^{obs}-\mathbf{g}\left(\tilde{\mathbf{x}}_t;\boldsymbol{\theta}\right)\right)'\mathbf{R}_{\mathbf{v}}^{-1}\left(\mathbf{y}_t^{obs}-\mathbf{g}\left(\tilde{\mathbf{x}}_t;\boldsymbol{\theta}\right)\right)\right\}\\
&\times\frac{1}{\left(\sqrt{2}\right)^{n_{x_2}}\prod_{i=1}^{n_{x_2}}\sqrt{R_{w_2}(i,i)}}\exp\left\{-\sqrt{2}\sum_{i=1}^{n_{x_2}}\frac{\left|x_{n_{x_1}+i,t}-h_{n_{x_1}+i}(\mathbf{x}_{2,t-1};\boldsymbol{\theta})\right|}{\sqrt{R_{w_2}(i,i)}}\right\}\\
&\times\det\left(\gamma_t\hat{\mathbf{S}}_{\mathbf{x}_2}^{CDKF}\left(t\right)\right)\exp\left\{0.5\left(\boldsymbol{\epsilon}_t\right)'\boldsymbol{\epsilon}_t\right\}
\end{aligned}$$

**Table 1: Calibration for the DSGE model**

| Label | Parameters | Values |
|---|---|---|
| Discount factor | $\beta$ | 0.9995 |
| Habit degree | $b$ | 0.80 |
| Habit persistence | $\rho_x$ | 0.95 |
| Preference | $\phi_1$ | 6 |
| Preference | $\phi_2$ | 0.88 |
| Adj costs for investments | $\kappa$ | 2.0 |
| Depreciation rate | $\delta$ | 0.025 |
| Cobb-Douglas parameter | $\theta$ | 0.36 |
| Price elasticity | $\eta$ | 6 |
| Degree of price stickiness | $\alpha$ | 0.85 |
| Reaction to lagged interest rate | $\rho_R$ | 0.99 |
| Reaction to inflation | $\beta_\pi$ | 1.65 |
| Reaction to output | $\beta_y$ | 0.15 |
| Inflation rate in steady state | $\pi_{ss}$ | 1.0070 |
| Growth rate in technology shocks | $\mu_{z,ss}$ | 1.0044 |
| Growth rate in investment shocks | $\mu_{\Upsilon,ss}$ | 1.0007 |
| Persistency in stationary technology shocks | $\rho_a$ | 0.9 |
| Persistency in investment shocks | $\rho_e$ | 0.9 |
| std. of nonstationary technology shocks | $\sqrt{Var\left(w_{z,t}\right)}$ | 0.008 |
| std. of stationary technology shocks | $\sqrt{Var\left(w_{a,t}\right)}$ | 0.012 |
| std. of investment shocks | $\sqrt{Var\left(w_{e,t}\right)}$ | 0.030 |
| std. of shocks to interest rate rule | $\sqrt{Var\left(w_{R,t}\right)}$ | 0.001 |
| std. of errors in the interest rate | $\sqrt{Var\left(v_{R,t}\right)}$ | 0.001 |
| std. of errors in inflation | $\sqrt{Var\left(v_{\pi,t}\right)}$ | 0.001 |
| std. of errors in the growth rate for consumption | $\sqrt{Var\left(v_{\Delta c,t}\right)}$ | 0.002 |
| std. of errors in the growth rate for investments | $\sqrt{Var\left(v_{\Delta i,t}\right)}$ | 0.002 |
| std. of errors in growth rate for output | $\sqrt{Var\left(v_{\Delta y,t}\right)}$ | 0.002 |

**Table 2: Empirical and simulated moments**

Data is from the Federal Reserve Bank of St. Louis covering the period 1956Q4-2009Q2 in the US. The quarterly interest rate is measured by the rate in the secondary market (TB3MS). The quarterly inflation rate is for consumer prices. The growth rate in consumption is calculated from real consumption expenditures (PCECC96). The series for real private fixed investments (FPIC96) is used to calculate the growth rate in investments. The growth rate in output is calculated from real GDP (GDPC96). All growth rates are expressed in quarterly terms and in per capita based on the total population in the US. Simulated moments are calculated based on a simulated time series of 1,000,000 observations.

| | $R_t$ | $\pi_t$ | $\Delta c_t$ | $\Delta i_t$ | $\Delta y_t$ |
|---|---|---|---|---|---|
| Empirical moments | | | | | |
| Mean | 0.0131 | 0.0088 | 0.0055 | 0.0056 | 0.0048 |
| Standard deviation | 0.0070 | 0.0063 | 0.0071 | 0.0254 | 0.0092 |
| Skewness | 1.0787 | 0.9700 | -0.6719 | -1.1384 | -0.4525 |
| Kurtosis | 4.8934 | 4.7352 | 4.9795 | 6.8528 | 4.5435 |
| Simulated moments using Normal distributed shocks | | | | | |
| Mean | 0.0134 | 0.0083 | 0.0048 | 0.0055 | 0.0048 |
| Standard deviation | 0.0088 | 0.0063 | 0.0084 | 0.0270 | 0.0118 |
| Skewness | 0.5664 | 1.1100 | 0.0337 | -0.1646 | 0.0495 |
| Kurtosis | 3.9029 | 5.0441 | 3.0840 | 3.0413 | 3.0406 |
| Simulated moments using Laplace distributed shocks | | | | | |
| Mean | 0.0134 | 0.0083 | 0.0048 | 0.0055 | 0.0048 |
| Standard deviation | 0.0088 | 0.0062 | 0.0084 | 0.0269 | 0.0118 |
| Skewness | 0.5969 | 1.1653 | 0.0217 | -0.2488 | 0.0167 |
| Kurtosis | 4.1262 | 5.4804 | 4.4121 | 4.1642 | 3.7135 |

## References


Altig, D., Christiano, L. J., Eichenbaum, M. & Linde, J. (2005), 'Firm-specific capital, nominal rigidities and the business cycle', *NBER Working Paper* **no 11034**, 1–50.

Amisano, G. & Tristani, O. (2007), 'Euro area inflation persistence in an estimated nonlinear DSGE model', *Working Paper* .

An, S. (2005), 'Bayesian estimation of DSGE-models: Lessons from second-order approximations', *Working Paper* .

An, S. & Schorfheide, F. (2007), 'Bayesian analysis of DSGE models', *Econometric Reviews* **26**(2-4), 113–172.

Arulampalam, M. S., Maskell, S., Gordon, N. & Clapp, T. (2002), 'A tutorial on particle filters for online Nonlinear/Non-gaussian bayesian tracking', *IEEE Transactions on signal processing* **50**(2), 174–188.

Berzuini, C., Best, N. G., Gilks, W. R. & Larizza, C. (1997), 'Dynamic conditional independence models and markov chain monte carlo methods', *American Statistical Association* **92**(440), 1403–1412.

Christiano, L. J., Eichenbaum, M. & Evans, C. L. (2005), 'Nominal rigidities and the dynamic effects of a shock to monetary policy', *Journal of Political Economy* **113**, 1–45.

Doh, T. (2007), 'Yield curve in an estimated nonlinear macro model', *Working Paper* .

Doucet, A., de Freitas, N. & Gordon, N. (2001), 'Sequential monte carlo methods in practice', *Springer* .

Doucet, A., Godsill, S. & Andrieu, C. (2000), 'On sequential monte carlo sampling methods for bayesian filtering', *Statistics and Computing* **10**, 197–208.

Fernández-Villaverde, J. & Rubio-Ramírez, J. F. (2005), 'Estimating dynamic equilibrium economies: Linear versus nonlinear likelihood', *Journal of Applied Econometrics* **20**, 891–910.

Fernández-Villaverde, J. & Rubio-Ramírez, J. F. (2007), 'Estimating macroeconomic models: A likelihood approach', *Review of Economic Studies* **74**, 1–46.

Flury, T. & Shephard, N. (2008), 'Bayesian inference based only on simulated likelihood: Particle filter analysis of dynamic economic models', *Forthcoming in Econometric Theory* .

Fox, D. (2001), 'KLD-sampling: Adaptive particle filters', *Advances in Neural Information Processing Systems* **14**, 713–720.

Gilks, W. R. & Berzuini, C. (2001), 'Following a moving target - monte carlo inference for dynamic bayesian models', *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **63**(1), 127–146.

Graeve, F. D., Emiris, M. & Wouters, R. (2009), 'A structural decomposition of the US yield curve', *Journal of Monetary Economics* **56**, 545–559.

Jazwinski, A. H. (1970), 'Stochastic processes and filtering theory', *Academic Press, Inc. (London) Ltd.* .

Kim, J., Kim, S., Schaumburg, E. & 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.

Klaas, M., Freitas, N. D. & Doucet, A. (2005), 'Towards practical n-squared monte carlo: The marginal particle filter', *Working Paper* .

Liu, J. S. & Chen, R. (1998), 'Sequential monte carlo methods for dynamic systems', *Journal of the American Statistical Association* **93**(443), 1032–1044.

Merwe, R. V. D., Doucet, A., de Freitas, N. & Wan, E. (2000), 'The unscented particle filter', *Working Paper* .

Merwe, R. V. D. & Wan, E. (2003), 'Sigma-point kalman filters for probabilistic inference in dynamic state-space models', *Working Paper* .

Musso, C., Oudjane, N. & LeGland, F. (2001), 'Improving regularised particle filters', *Sequential Monte Carlo Methods in Practice* .

Norgaard, M., Poulsen, N. K. & Ravn, O. (2000), 'Advances in derivative-free state estimation for nonlinear systems', *Automatica* **36:11**, 1627–1638.

Pitt, M. K. & Shephard, N. (1999), 'Filtering via simulation: Auxiliary particle filters', *Journal of the American Statistical Association* **94**(446), 590–599.

Schmitt-Grohé, S. & Uribe, M. (2004), 'Solving dynamic general equilibrium models using a second-order approximation to the policy function', *Journal of Economic Dynamics and Control* **28**, 755–775.

Simon J. Godsill, A. D. & West, M. (2004), 'Monte carlo smoothing for nonlinear time series', *Journal of the American Statistical Association* **99**(465), 156–168.

Smets, F. & Wouters, R. (2007), 'Shocks and frictions in US business cycles: A bayesian DSGE approach', *American Economic Review* **97**(3), 586–606.

Soto, A. (2005), 'Self adaptive particle filter', *Working Paper* .

Strid, I. (2006), 'Parallel particle filters for likelihood evaluation in DSGE models: An assessment', *Working Paper* .

Thomas F. Cooley, E. (1995), 'Frontiers of business cycle research', *Princeton University Press, New Jersey* .

**Figure 1: Variation in log-likelihood function**

The x-axis shows the number of particles, and the y-axis shows the standard deviation in the estimated log-likelihood function for normally distributed shocks.

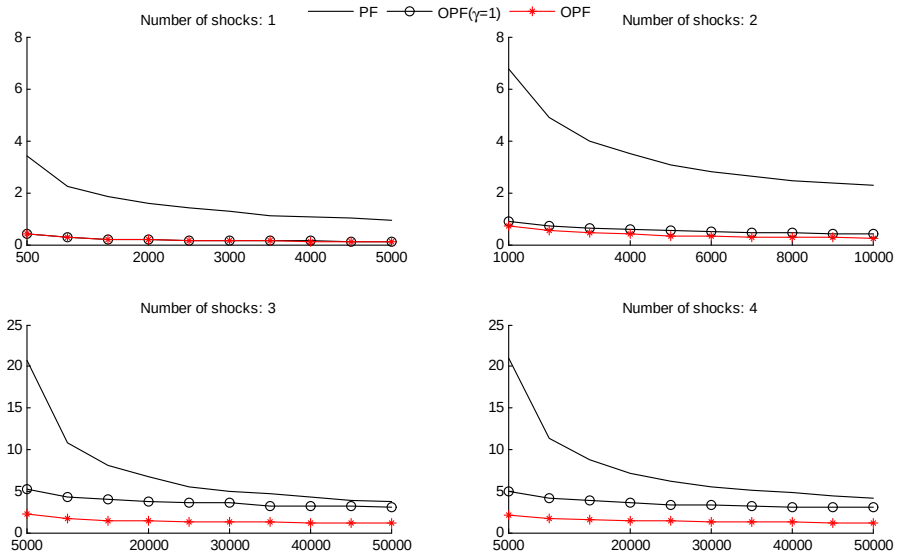


**Figure 2: The average effective sample size**

The x-axis shows the number of particles, and the y-axis shows average effective sample size for normally distributed shocks.

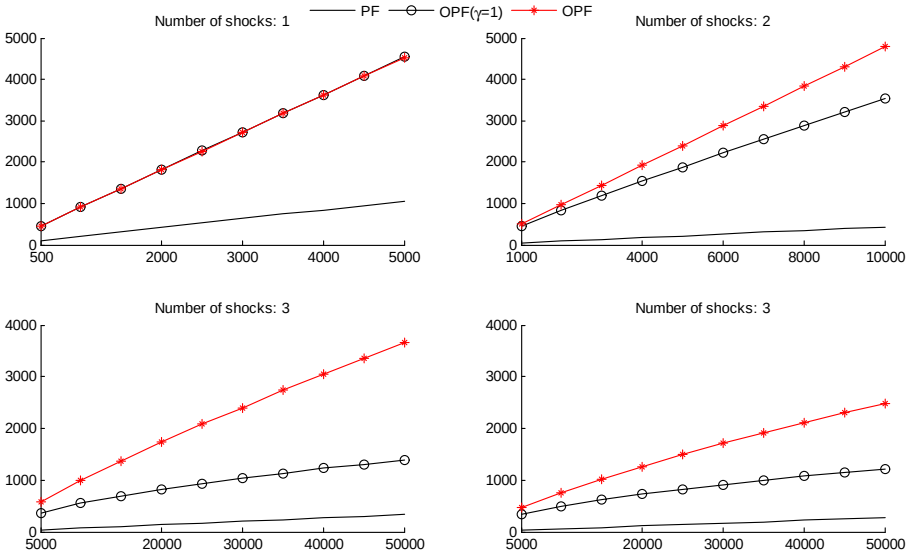

**Figure 3: The minimum effective sample size**

The x-axis shows the number of particles, and the y-axis shows the statistic for the minimum effective sample size in the case of normally distributed shocks.

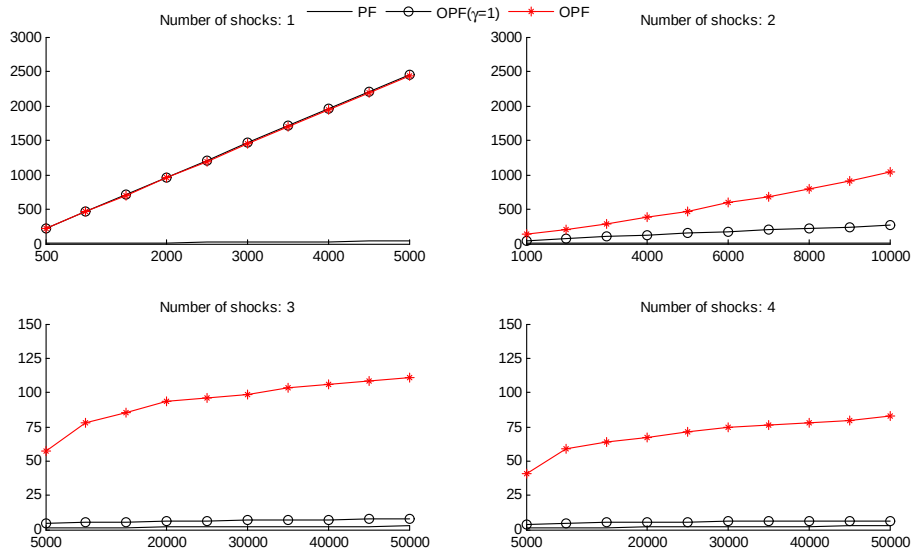


**Figure 4: The average number of seconds to compute the filters**

The x-axis shows the number of particles, and the y-axis shows the average number of seconds for running the filters with normally distributed shocks. All calculatings are done in Fortran 90 on Dell SC1435 compute-nodes, each with 2 dualcore Opteron 2.6 GHz, 8 GB memory, and 250 GB disk.

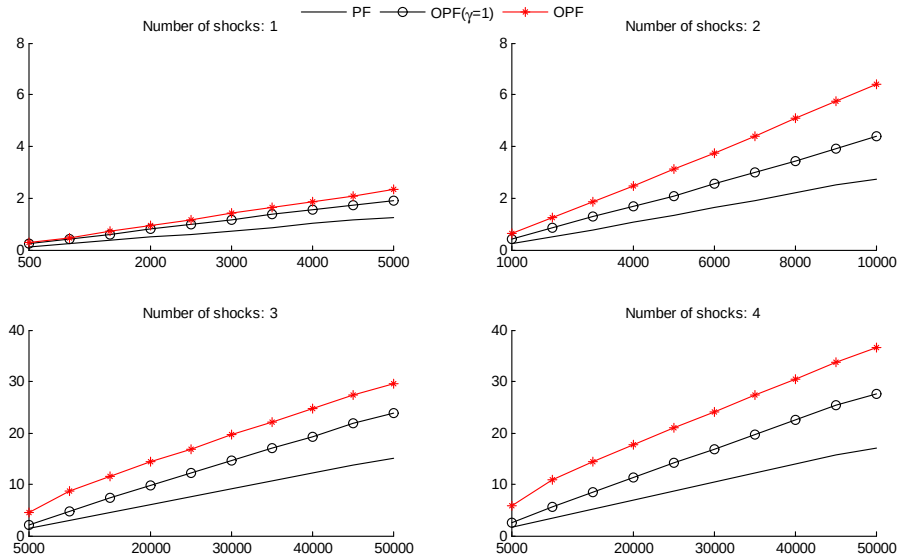

**Figure 5: Efficiency gain compared to the standard PF**

The x-axis shows the number of particles for the two versions of the OPF, and the y-axis shows the efficiency gain compared to the standard PF.

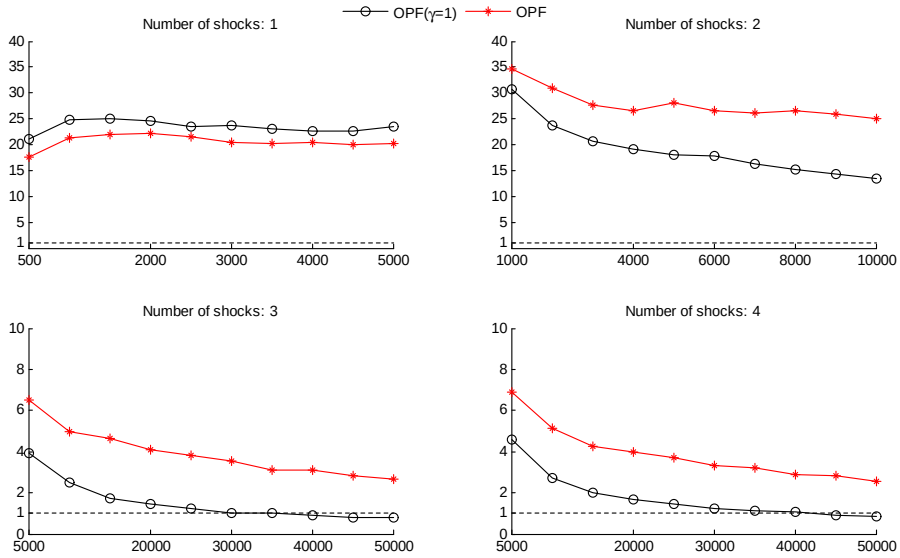


**Figure 6: Variation in log-likelihood function: the SPPF**

The x-axis shows the number of particles, and the y-axis shows the standard deviation in the estimated log-likelihood function for normally distributed shocks.

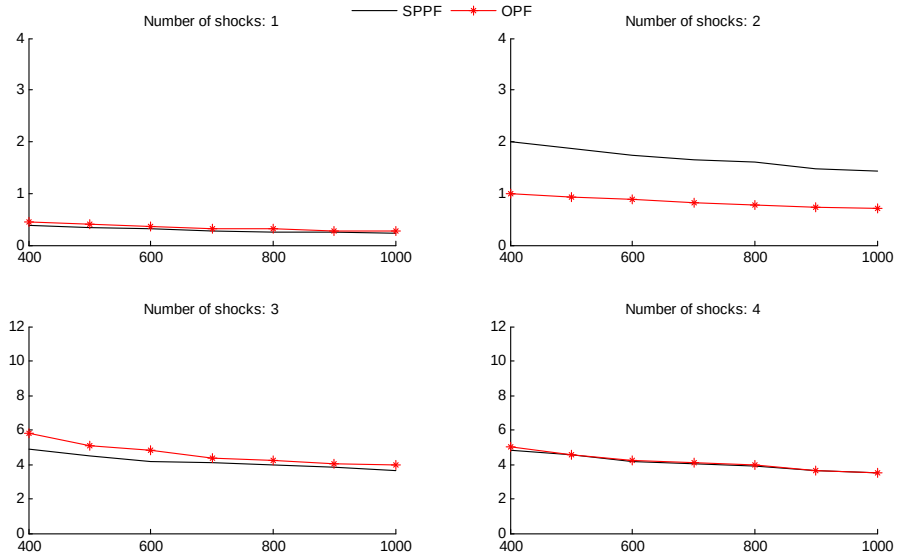

**Figure 7: Efficiency gain compared to the SPPF**

The x-axis shows the number of particles for the OPF, and the y-axis shows the efficiency gain compared to the SPPF.

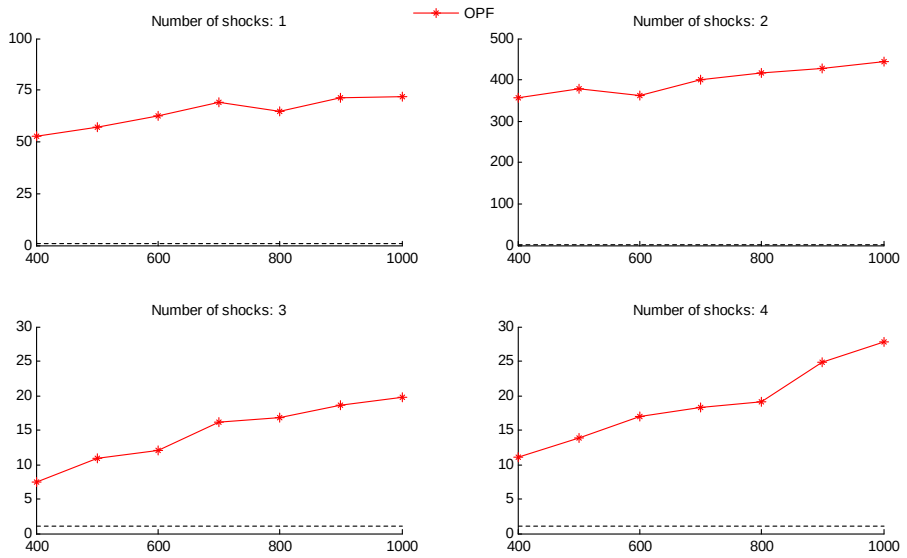


**Figure 8: Variation in log-likelihood function: Small measurement errors**

The x-axis shows the number of particles, and the y-axis shows the standard deviation in the estimated log-likelihood function for normally distributed shocks and small measurement errors in the observables.

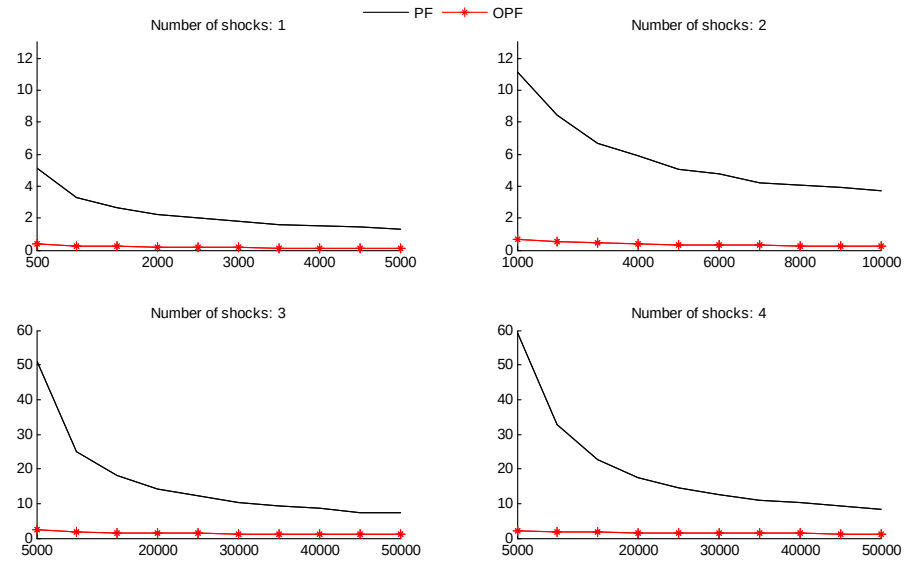

**Figure 9: Efficiency gain compared to the standard PF: Small measurement errors**
The x-axis shows the number of particles for the OPF, and the y-axis shows the efficiency gain compared to the standard PF for normally distributed shocks and small measurement errors in the observables.

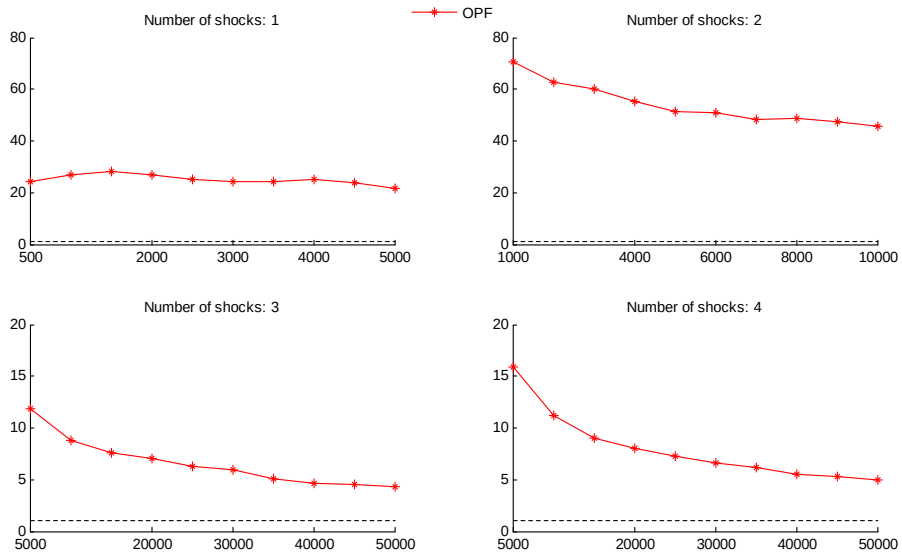


**Figure 10: Variation in log-likelihood function: Laplace distributed shocks**
The x-axis shows the number of particles, and the y-axis shows the standard deviation in the estimated log-likelihood function for Laplace distributed shocks.

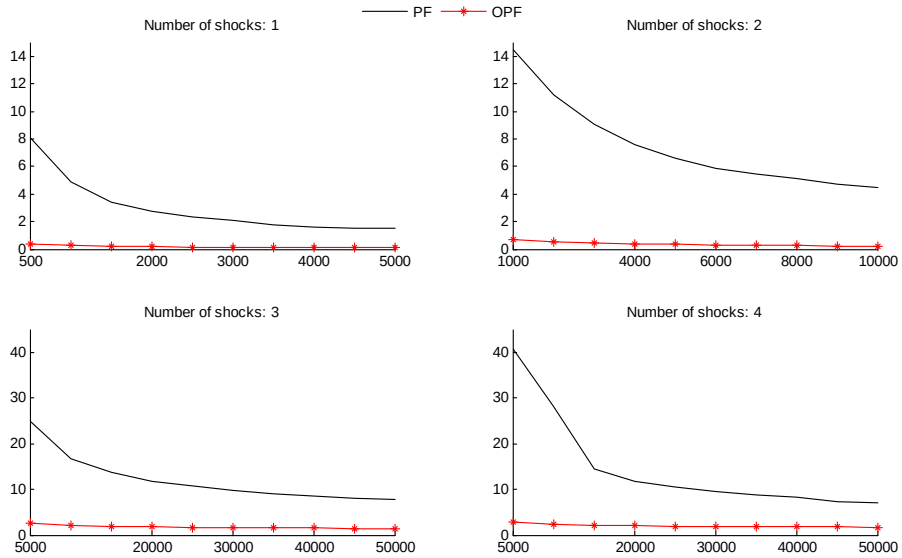

**Figure 11: Efficiency gain compared to the standard PF: Laplace distributed shocks**
The x-axis shows the number of particles for the OPF, and the y-axis shows the efficiency gain compared to the standard PF with Laplace distributed shocks.