CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2008-33

Non-linear DSGE Models, The Central Difference Kalman Filter, and The Mean Shifted Particle Filter

Martin Møller Andreasen

School of Economics and Management
University of Aarhus
Building 1322, DK-8000 Aarhus C
Denmark

Aarhus School of Business
University of Aarhus
Handelshøjskolen
Aarhus Universitet

UNIVERSITY OF
COPENHAGEN

Non-linear DSGE Models, The Central Difference Kalman Filter, and The Mean Shifted Particle Filter

Martin Møller Andreasen*
School of Economics and Management
University of Aarhus and CREATES

June 17, 2008

Abstract

This paper shows how non-linear DSGE models with potential non-normal shocks can be estimated by Quasi-Maximum Likelihood based on the Central Difference Kalman Filter (CDKF). The advantage of this estimator is that evaluating the quasi log-likelihood function only takes a fraction of a second. The second contribution of this paper is to derive a new particle filter which we term the Mean Shifted Particle Filter (MSPF_b). We show that the MSPF_b outperforms the standard Particle Filter by delivering more precise state estimates, and in general the MSPF_b has lower Monte Carlo variation in the reported log-likelihood function.

Keywords: Multivariate Stirling interpolation, Particle filtering, Non-linear DSGE models, Non-normal shocks, Quasi-maximum likelihood.

JEL: C13, C15, E10, E32

*Email: mandreasen@econ.au.dk. Telephone number: +45 8942 2138. I greatly acknowledge financial support from the Danish Center for Scientific Computation (DCSC) and assistance from Niels Carl W. Hansen. 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. Also a thanks to Magnus Norregaard for making codes of the CDKF available on his homepage. Finally, I would like to thank Stephanie Schmitt-Grohe, Bent Jesper Christensen, Torben M. Andersen, and Henning Bunzel for assistence. First version of the paper: May 23, 2008.

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 (2005b)). The fascinating work by Fernández-Villaverde & Rubio-Ramírez (2007a) shows how to do likelihood based inference for non-linear DSGE modes with potential non-normal shocks. They use particle filtering techniques for this purpose, and approximate the conditional state distributions through repeated use of importance sampling and resampling. Fernándes-Villaverde and Rubio-Ramirez (2005b, 2007a) use a particle filter which we will refer to as the standard Particle Filter (PF). Following their lead Justiniano & Primiceri (2008), An (2005), Strid (2006), Doh (2007), and An & Schorfheide (2007) have successfully applied this filter and estimated 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 in the filter. 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 about the current observables is ignored. A simple way to reduce the sample depletion problem is to increase the number of particles. This comes at the cost of increasing the computational requirements which are already severe when using the standard PF to estimate DSGE models. As we show in the present paper, with more than two shocks in a DSGE model a great deal of inaccuracy may still remain in the filter, even when a very large number of particles is used. Hence, faster and more precise filtering methods are needed in relation to DSGE models which typically have more than two shocks. Developing and testing such methods is the purpose of the present paper.

Recently, Norgaard, Poulsen & Ravn (2000) introduced the Central Difference Kalman Filter (CDKF) for state estimation in a general non-linear and non-normal state space system. Norgaard et al. (2000) show that this filter outperforms the Extended Kalman Filter and the second order Kalman Filter. The updating rule for the state vector in the CDKF is restricted to be linear, and the recursive equations for the state estimator and its covariance matrix are therefore only functions of first and second moments. The procedure adopted in the CDKF is to approximate these moments up to second-order accuracy by a deterministic sampling approach based on multivariate Stirling interpolations. This approximation method is computationable very fast and reasonably accurate. The sampling approach implies that no derivatives are required in the CDKF, and this makes the filter very robust and easy to implement. Furthermore, the CDKF propagates and updates the square root of the state covariance matrix over time. This feature increases the numerical stability of the filter, and it ensures that all covariance matrices in the filter are symmetric and positive semidefinite as desired.

However, the CDKF does not rely on likelihood based methods. This implies that the CDKF in general is a sub-optimal filter and that unknown parameters in the DSGE model cannot be estimated by Maximum Likelihood (ML) or Bayesian methods. The first contribution of this paper is to show how parameters in non-linear DSGE models with potentially non-normal shocks can be estimated by Quasi-Maximum Likelihood (QML) based on the CDKF. We focus on the case where measurement errors are present in the observables, and we argue that the QML estimator is consistent and asympotically normal when DSGE models are approximated up to

second order. These properties also hold for DSGE models approximated up to third order if i) third order terms are small or if ii) the state vector, the shocks, and the measurement errors all are normally distributed. We also discuss the case where DSGE models are approximated up to fourth or higher order. Here consistency and asympotic normality is harder to ensure because it requires that all third and higher order terms are insignificant. The main advantage of the QML estimator is that evaluating the quasi log-likelihood function only takes a fraction of a second. In our case, evaluating this function is approximately 120 times faster than approximating the likelihood function by the standard PF with 60,000 particles.

The second contribution of this paper is to show how the CDKF can be used in an efficient way to improve the proposal distribution in particle filters. That is, we focus on the weakest component of the standard PF and derive a new particle filter which we refer to as the Mean Shifted Particle Filter (MSPF_b). We show that the MSPF_b outperforms the standard PF by delivering more precise state estimates, and in general the MSPF_b has lower Monte Carlo variation in the reported log-likelihood function. This is the case even if a low number of particles is used in the MSPF_b and a large number of particles is used in the standard PF. As a result, the MSPF_b is faster to compute than the standard PF.

We test the performance of i) the CDKF, ii) the standard PF, and iii) the MSPF_b in a Monte Carlo study. A more or less standard New Keynesian DSGE is chosen for this purpose because the model is widely used in the literature (see Christiano, Eichenbaum & Evans (2005), Smets & Wouters (2003), Altig, Christiano, Eichenbaum & Linde (2005), Fernández-Villaverde & Rubio-Ramirez (2007*b*), among others). The solution to this model is here approximated up to second order. In addition to the findings mentioned above, we highlight the following results from this Monte Carlo study. First, with one or two shocks to the non-linear DSGE model, both particle filters outperform the CDKF in terms of root mean squared errors for the state vector. Second, with three or more shocks, the CDKF clearly outperforms the standard PF and marginally the MSPF_b. Third, the finite sample distributions for the QML estimator are shown to be well approximated by the asymptotic distribution, and standard errors for the QML estimator can be estimated by well-known techniques due to the smooth nature of the quasi log-likelihood function. These results hold both with normal *and* non-normal shocks driving the economy.

Thus, the CDKF and the MSPF_b are useful tools when taking non-linear DSGE models to the data. For instance, the QML estimator based on the CDKF can be used in itself or simply as a fast first check of a model's ability to match the data. If additional efficiency is called for, the QML estimates can be used as good starting values for conducting full ML estimation by a particle filter. Such a maximisation should be easier to implement by the MSPF_b compared to using the standard PF, because the MSPF_b requires fewer particles and in general has a lower Monte Carlo variation in the reported log-likelihood function. For the same reasons, a Bayesian researcher may also benefit greatly from the MSPF_b when performing the MCMC analysis. Note finally that the Bayesian researcher may also find the QML estimator useful because i) the QML estimates can be used as good starting points for the Markov chain, and ii) the Hessian matrix at the QML estimates can be used to specify the proposal distribution in the random walk Metropolis algorithm for the MCMC analysis.

The rest of the paper is organized as follows. We present the state space representation of DSGE models in section 1. Section 2 describes the CDKF and shows how to estimate the parameters in a non-linear DSGE model by QML. Particle filters are discussed in section 3, where

we describe the standard PF and derive the new MSPF$_b$. We set up a DSGE model in section 5 and construct a sequence of "test economies" for the Monte Carlo study. The performance of the various filters and the QML estimator are then examined in a Monte Carlo study in section 6. Section 7 concludes.

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 at time 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 $\mathbf{v}_t$. We let $\mathbf{x}_t$ have dimension $n_x \times 1$ and $\mathbf{v}_t$ have dimension $n_v \times 1$. More formally,

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

The function $\mathbf{g}\left(\cdot\right)$ is determined by i) the parameters $\boldsymbol{\theta} \in \Theta$ in the economic model and ii) the equilibrium conditions describing the economy. The equations in (1) are known as the set of measurement equations.

The law of motion for the state vector is given by

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

where $\mathbf{w}_t$ 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 if we increase the dimension of the state vector. In general, the state vector $\mathbf{x}_t$ is not observable. Note however, that observable state variables can handled in this framework by letting one or more elements in $\mathbf{g}\left(\cdot\right)$ be the identity mapping. 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

$$\mathbf{x}_t \equiv \left[\begin{array}{c} \mathbf{x}_{1,t} \\ \mathbf{x}_{2,t} \end{array} \right], \tag{3}$$

where $\mathbf{x}_{1,t}$ denotes the endogenous state variables and $\mathbf{x}_{2,t}$ denotes 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

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

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

We do not impose any specific probability distributions for $\mathbf{v}_t$ or $\mathbf{w}_t$. That is, we let $\mathbf{v}_t \sim p\left(\mathbf{v}_t; \boldsymbol{\theta}\right)$ and $\mathbf{w}_t \sim p\left(\mathbf{w}_t; \boldsymbol{\theta}\right)$ and denote their individual covariance matrices by $\mathbf{R}_{\mathbf{v}}\left(t\right)$ and $\mathbf{R}_{\mathbf{w}}\left(t\right)$, respectively. However, independence between $\mathbf{v}_t$ and $\mathbf{w}_t$ is assumed.

3 The Central Difference Kalman Filter

This section presents the CDKF. We start by deriving recursive equations for estimating the unknown state vector $\mathbf{x}_t$ in (1) and (2) based on a two-step procedure of prediction and updating. Following Norgaard et al. (2000), the updating rule for the state vector is in this section restricted to be a linear function of the observables. This implies that only first and second moments from the state space system in (1) and (2) are needed during the state estimation. We then show how these moments are approximated in the CDKF. Finally, a QML estimator is suggested for estimating structural parameters in non-linear DSGE models.

3.1 A linear updating rule for the state vector

We use the standard notation that a "bar" denotes a prior estimates and a "hat" denotes posterior estimates. For instance, $\bar{\mathbf{x}}_{t+1} \equiv E_t[\mathbf{x}_{t+1}]$ and $\hat{\mathbf{x}}_{t+1} \equiv E_{t+1}[\mathbf{x}_{t+1}]$. Here, $E_t[\cdot]$ is the conditional expectation given the observations $\mathbf{y}_{1:t} \equiv \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_t\}$.[1]

The a priori state estimator follows directly from (2) and is given by

$$\bar{\mathbf{x}}_{t+1} \equiv E_t[\mathbf{h}(\mathbf{x}_t, \mathbf{w}_{t+1}; \boldsymbol{\theta})]. \tag{6}$$

The conditional error covariance matrix for this estimator is denoted by

$$\bar{\mathbf{P}}_{\mathbf{xx}}(t+1) \equiv E_t\left[(\mathbf{x}_{t+1} - \bar{\mathbf{x}}_{t+1})(\mathbf{x}_{t+1} - \bar{\mathbf{x}}_{t+1})'\right]. \tag{7}$$

For tractability, the updating rule for the a prior state estimator is restricted to

$$\hat{\mathbf{x}}_{t+1} = \mathbf{b}_{t+1} + \mathbf{K}_{t+1}\mathbf{y}_{t+1}, \tag{8}$$

where $\mathbf{b}_{t+1}$ and $\mathbf{K}_{t+1}$ are determined below. If we choose $\mathbf{b}_{t+1}$ such that the a prior and the posterior state estimators are unbiased, then it follows directly that

$$\mathbf{b}_{t+1} = \bar{\mathbf{x}}_{t+1} - \mathbf{K}_{t+1}\bar{\mathbf{y}}_{t+1}, \tag{9}$$

where

$$\bar{\mathbf{y}}_{t+1} \equiv E_t[\mathbf{g}(\mathbf{x}_{t+1}, \mathbf{v}_{t+1}; \boldsymbol{\theta})]. \tag{10}$$

This gives rise to the well-known updating rule

$$\hat{\mathbf{x}}_{t+1} = \bar{\mathbf{x}}_{t+1} + \mathbf{K}_{t+1}(\mathbf{y}_{t+1} - \bar{\mathbf{y}}_{t+1}). \tag{11}$$

The value of $\mathbf{K}_{t+1}$ is determined such that the conditional error covariance matrix for $\hat{\mathbf{x}}_{t+1}$ is minimized. It is straightforward to show that this criterion implies (see Lewis (1986))

$$\mathbf{K}_{t+1} = \mathbf{P}_{\mathbf{xy}}(t+1)\,\mathbf{P}_{\mathbf{yy}}(t+1)^{-1}, \tag{12}$$

[1] An alternative notation is $\mathbf{x}_{t+1|t} = E_t[\mathbf{x}_{t+1}]$ and $\mathbf{x}_{t+1|t+1} \equiv E_{t+1}[\mathbf{x}_{t+1}]$ as in Hamilton (1994), for instance. We choose $\bar{\mathbf{x}}_{t+1}$ and $\hat{\mathbf{x}}_{t+1}$ because this notation is more parsimonious.

where we have defined

$$\mathbf{P}_{\mathbf{xy}}(t+1) \equiv E_t\left[(\mathbf{x}_{t+1} - \bar{\mathbf{x}}_{t+1})(\mathbf{y}_{t+1} - \bar{\mathbf{y}}_{t+1})'\right] \tag{13}$$

$$\bar{\mathbf{P}}_{\mathbf{yy}}(t+1) \equiv E_t\left[(\mathbf{y}_{t+1} - \bar{\mathbf{y}}_{t+1})(\mathbf{y}_{t+1} - \bar{\mathbf{y}}_{t+1})'\right]. \tag{14}$$

The conditional error covariance matrix for $\hat{\mathbf{x}}_{t+1}$ can be expressed as

$$\begin{aligned} \hat{\mathbf{P}}_{\mathbf{xx}}(t+1) &\equiv E_{t+1}\left[(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1})(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1})'\right] \\ &= \bar{\mathbf{P}}_{\mathbf{xx}}(t+1) - \mathbf{K}_{t+1}\bar{\mathbf{P}}_{\mathbf{yy}}(t+1)\mathbf{K}_{t+1}'. \end{aligned} \tag{15}$$

Thus, the optimal filtering equations for the class of updating rules implied by (8) are given by (6), (7), and (11) - (15).

Two remarks are in order. First, if we are able to accurately evaluate the required first and second moments, then the a prior and the posterior state estimators in (6) and (11) are unbiased by construction. This result holds even thought the state space system is non-linear and no distributional assumptions are assumed for $\mathbf{v}_t$ and $\mathbf{w}_t$. Second, in the case where $\mathbf{g}(\cdot)$ and $\mathbf{h}(\cdot)$ are linear functions, the required first and second moments can be evaluated exactly, and this leads to the standard Kalman Filter. Recall that the standard Kalman Filter has a linear updating rule for the posterior state vector, and the filter can be derived without imposing distributional assumptions for $\mathbf{v}_t$ and $\mathbf{w}_t$ (see for instance Tanizaki (1996)).

However, the non-linearity in (1) and (2) implies that some approximation is needed to calculate the required moments. One way to proceed is to linearize the state space system such that

$$\mathbf{y}_t \approx \mathbf{g}(\bar{\mathbf{x}}_t, \bar{\mathbf{v}}_t; \boldsymbol{\theta}) + \mathbf{G}_{\mathbf{x},t}(\mathbf{x}_t - \bar{\mathbf{x}}_t) + \mathbf{G}_{\mathbf{v},t}(\mathbf{v}_t - \bar{\mathbf{v}}_t) \tag{16}$$

$$\mathbf{x}_{t+1} \approx \mathbf{h}(\hat{\mathbf{x}}_t, \bar{\mathbf{w}}_{t+1}; \boldsymbol{\theta}) + \mathbf{H}_{\mathbf{x},t}(\mathbf{x}_t - \hat{\mathbf{x}}_t) + \mathbf{H}_{\mathbf{w},t}(\mathbf{w}_{t+1} - \bar{\mathbf{w}}_{t+1}) \tag{17}$$

where

$$\mathbf{G}_{\mathbf{x},t} \equiv \left.\frac{\partial \mathbf{g}(\mathbf{x}, \bar{\mathbf{v}}_t; \boldsymbol{\theta})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\bar{\mathbf{x}}_t} \qquad \mathbf{G}_{\mathbf{v},t} \equiv \left.\frac{\partial \mathbf{g}(\bar{\mathbf{x}}_t, \mathbf{v}; \boldsymbol{\theta})}{\partial \mathbf{v}}\right|_{\mathbf{v}=\bar{\mathbf{v}}_t} \tag{18}$$

$$\mathbf{H}_{\mathbf{x},t} \equiv \left.\frac{\partial \mathbf{h}(\mathbf{x}, \bar{\mathbf{w}}_{t+1}; \boldsymbol{\theta})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}_t} \qquad \mathbf{H}_{\mathbf{w},t} \equiv \left.\frac{\partial \mathbf{h}(\hat{\mathbf{x}}_t, \mathbf{w}; \boldsymbol{\theta})}{\partial \mathbf{w}}\right|_{\mathbf{w}=\bar{\mathbf{w}}_{t+1}} \tag{19}$$

Given these approximations, the first and second moments in the filtering equations are easy to evaluate. This is the approach adopted in the Extended Kalman Filter (see for instance Jazwinski (1970)). However, the approximations in (16) and (17) are only accurate up to first-order, and the approximations do not take the probability distribution for the state vector into account because the linearization is done around a single point. Hence, the approximations in (16) and (17) often create significant approximation errors (see Norgaard et al. (2000), Merwe & Wan (2003), among others).

3.2 Multivariate Stirling interpolations

This section shows how the first and second moments in the filtering equations above are approximated in the CDKF. We only describe the version of the CDKF based on a second order approximation.

The idea in the CDKF is to approximate the non-linear expectations in (6), (7), and (11) - (15) by second-order multivariate Stirling interpolations. For this purpose we introduce additional notation. First, four squared and upper triangular Cholesky factorizations $\mathbf{S_w}(t)$, $\mathbf{S_v}(t)$, $\bar{\mathbf{S}}_{\mathbf{x}}(t)$, and $\hat{\mathbf{S}}_{\mathbf{x}}(t)$ are defined by

$$\mathbf{R_w}(t) = \mathbf{S_w}(t)\,\mathbf{S_w}(t)' \qquad \mathbf{R_v}(t) = \mathbf{S_v}(t)\,\mathbf{S_v}(t)' \tag{20}$$

$$\bar{\mathbf{P}}_{\mathbf{xx}}(t) = \bar{\mathbf{S}}_{\mathbf{x}}(t)\,\bar{\mathbf{S}}_{\mathbf{x}}(t)' \qquad \hat{\mathbf{P}}_{\mathbf{xx}}(t) = \hat{\mathbf{S}}_{\mathbf{x}}(t)\,\hat{\mathbf{S}}_{\mathbf{x}}(t)'. \tag{21}$$

For elements in the four matrices we use the notation

$$\mathbf{S_w}(t+1) = \left[\begin{array}{cccc} \mathbf{s}_{\mathbf{w},1} & \mathbf{s}_{\mathbf{w},2} & ... & \mathbf{s}_{\mathbf{w},n_w} \end{array}\right] \qquad \mathbf{S_v}(t) = \left[\begin{array}{cccc} \mathbf{s}_{\mathbf{v},1} & \mathbf{s}_{\mathbf{v},2} & ... & \mathbf{s}_{\mathbf{v},n_v} \end{array}\right] \tag{22}$$

$$\bar{\mathbf{S}}_{\mathbf{x}}(t) = \left[\begin{array}{cccc} \bar{\mathbf{s}}_{\mathbf{x},1} & \bar{\mathbf{s}}_{\mathbf{x},2} & ... & \bar{\mathbf{s}}_{\mathbf{x},n_x} \end{array}\right] \qquad \hat{\mathbf{S}}_{\mathbf{x}}(t) = \left[\begin{array}{cccc} \hat{\mathbf{s}}_{\mathbf{x},1} & \hat{\mathbf{s}}_{\mathbf{x},2} & ... & \hat{\mathbf{s}}_{\mathbf{x},n_x} \end{array}\right]. \tag{23}$$

Here, $\mathbf{s}_{\mathbf{w},j}$ has dimension $n_w \times 1$ for $j = 1, ..., n_w$, $\mathbf{s}_{\mathbf{v},j}$ has dimension $n_v \times 1$ for $j = 1, ..., n_v$, and $\bar{\mathbf{s}}_{\mathbf{x},j}$ and $\hat{\mathbf{s}}_{\mathbf{x},j}$ have dimension $n_x \times 1$ for $j = 1, ..., n_x$. Next, we define four matrices by

$$\underset{(n_x \times n_x)}{\mathbf{S}^{(1)}_{\mathbf{xx}}}(t) = \left\{\left(h_i\left(\hat{\mathbf{x}}_t + h\hat{\mathbf{s}}_{\mathbf{x},j}, \bar{\mathbf{w}}_{t+1}; \boldsymbol{\theta}\right) - h_i\left(\hat{\mathbf{x}}_t - h\hat{\mathbf{s}}_{\mathbf{x},j}, \bar{\mathbf{w}}_{t+1}; \boldsymbol{\theta}\right)\right)/2h\right\} \tag{24}$$

$$\underset{(n_x \times n_w)}{\mathbf{S}^{(1)}_{\mathbf{xw}}}(t) = \left\{\left(h_i\left(\hat{\mathbf{x}}_t, \bar{\mathbf{w}}_{t+1} + h\mathbf{s}_{\mathbf{w},j}; \boldsymbol{\theta}\right) - h_i\left(\hat{\mathbf{x}}_t, \bar{\mathbf{w}}_{t+1} - h\mathbf{s}_{\mathbf{w},j}; \boldsymbol{\theta}\right)\right)/2h\right\} \tag{25}$$

$$\underset{(n_y \times n_x)}{\mathbf{S}^{(1)}_{\mathbf{yx}}}(t) = \left\{\left(g_i\left(\bar{\mathbf{x}}_t + h\bar{\mathbf{s}}_{\mathbf{x},j}, \bar{\mathbf{v}}_t; \boldsymbol{\theta}\right) - g_i\left(\bar{\mathbf{x}}_t - h\bar{\mathbf{s}}_{\mathbf{x},j}, \bar{\mathbf{v}}_t; \boldsymbol{\theta}\right)\right)/2h\right\} \tag{26}$$

$$\underset{(n_y \times n_v)}{\mathbf{S}^{(1)}_{\mathbf{yv}}}(t) = \left\{\left(g_i\left(\bar{\mathbf{x}}_t, \bar{\mathbf{v}}_t + h\mathbf{s}_{\mathbf{v},j}; \boldsymbol{\theta}\right) - g_i\left(\bar{\mathbf{x}}_t, \bar{\mathbf{v}}_t - h\mathbf{s}_{\mathbf{v},j}; \boldsymbol{\theta}\right)\right)/2h\right\}. \tag{27}$$

Here, we use the notation $\mathbf{h}(\cdot) \equiv \left[\begin{array}{cccc} h_1(\cdot) & h_2(\cdot) & ... & h_{n_x}(\cdot) \end{array}\right]'$ and similarly for the function $\mathbf{g}(\cdot)$. The matrices in (24) - (27) contain the first-order effects of the general nonlinear functions and this is denoted by the superscript (1). The corresponding matrices for the second-order effects are:

$$\underset{(n_x \times n_x)}{\mathbf{S}^{(2)}_{\mathbf{xx}}}(t) = \left\{\frac{\sqrt{h^2-1}}{2h^2}\left(h_i\left(\hat{\mathbf{x}}_t + h\hat{\mathbf{s}}_{\mathbf{x},j}, \bar{\mathbf{w}}_{t+1}; \boldsymbol{\theta}\right) + h_i\left(\hat{\mathbf{x}}_t - h\hat{\mathbf{s}}_{\mathbf{x},j}, \bar{\mathbf{w}}_{t+1}; \boldsymbol{\theta}\right) - 2h_i\left(\hat{\mathbf{x}}_t, \bar{\mathbf{w}}_{t+1}; \boldsymbol{\theta}\right)\right)\right\} \tag{28}$$

$$\underset{(n_x \times n_w)}{\mathbf{S}^{(2)}_{\mathbf{xw}}}(t) = \left\{\frac{\sqrt{h^2-1}}{2h^2}\left(h_i\left(\hat{\mathbf{x}}_t, \bar{\mathbf{w}}_{t+1} + h\mathbf{s}_{\mathbf{w},j}; \boldsymbol{\theta}\right) + h_i\left(\hat{\mathbf{x}}_t, \bar{\mathbf{w}}_{t+1} - h\mathbf{s}_{\mathbf{w},j}; \boldsymbol{\theta}\right) - 2h_i\left(\hat{\mathbf{x}}_t, \bar{\mathbf{w}}_{t+1}; \boldsymbol{\theta}\right)\right)\right\} \tag{29}$$

$$\underset{(n_y \times n_x)}{\mathbf{S}^{(2)}_{\mathbf{yx}}}(t) = \left\{\frac{\sqrt{h^2-1}}{2h^2}\left(g_i\left(\bar{\mathbf{x}}_t + h\bar{\mathbf{s}}_{\mathbf{x},j}, \bar{\mathbf{v}}_t; \boldsymbol{\theta}\right) + g_i\left(\bar{\mathbf{x}}_t - h\bar{\mathbf{s}}_{\mathbf{x},j}, \bar{\mathbf{v}}_t; \boldsymbol{\theta}\right) - 2g_i\left(\bar{\mathbf{x}}_t, \bar{\mathbf{v}}_t; \boldsymbol{\theta}\right)\right)\right\} \tag{30}$$

$$\underset{(n_y \times n_v)}{\mathbf{S}^{(2)}_{\mathbf{yv}}}(t) = \left\{\frac{\sqrt{h^2-1}}{2h^2}\left(g_i\left(\bar{\mathbf{x}}_t, \bar{\mathbf{v}}_t + h\mathbf{s}_{\mathbf{v},j}; \boldsymbol{\theta}\right) + g_i\left(\bar{\mathbf{x}}_t, \bar{\mathbf{v}}_t - h\mathbf{s}_{\mathbf{v},j}; \boldsymbol{\theta}\right) - 2g_i\left(\bar{\mathbf{x}}_t, \bar{\mathbf{v}}_t; \boldsymbol{\theta}\right)\right)\right\}. \tag{31}$$

Norgaard et al. (2000) recommend to determine the value of the scalar h based on the distribution of the random variable subject to the multivariate Stirling interpolation. Here, it is optimal to

let h^2 be equal to the kurtosis of this distribution.

As shown by Norgaard et al. (2000), the a priori state estimator in the CDKF is

$$\begin{aligned}\bar{\mathbf{x}}_{t+1} &= \frac{h^2-n_x-n_w}{h^2}\mathbf{h}\left(\hat{\mathbf{x}}_t,\bar{\mathbf{w}}_{t+1};\boldsymbol{\theta}\right) \\ &+\frac{1}{2h^2}\sum_{p=1}^{n_x}\left(\mathbf{h}\left(\hat{\mathbf{x}}_t+h\hat{\mathbf{s}}_{\mathbf{x},p},\bar{\mathbf{w}}_{t+1};\boldsymbol{\theta}\right)+\mathbf{h}\left(\hat{\mathbf{x}}_t-h\hat{\mathbf{s}}_{\mathbf{x},p},\bar{\mathbf{w}}_{t+1};\boldsymbol{\theta}\right)\right) \\ &+\frac{1}{2h^2}\sum_{p=1}^{n_w}\left(\mathbf{h}\left(\hat{\mathbf{x}}_t,\bar{\mathbf{w}}_{t+1}+h\mathbf{s}_{\mathbf{w},p};\boldsymbol{\theta}\right)+\mathbf{h}\left(\hat{\mathbf{x}}_t,\bar{\mathbf{w}}_{t+1}-h\mathbf{s}_{\mathbf{w},p};\boldsymbol{\theta}\right)\right).\end{aligned} \tag{32}$$

The a priori covariance matrix of this estimator is obtained by a Householder transformation of the matrix

$$\left[\begin{array}{cccc}\mathbf{S}_{\mathbf{xx}}^{(1)}(t) & \mathbf{S}_{\mathbf{xw}}^{(1)}(t) & \mathbf{S}_{\mathbf{xx}}^{(2)}(t) & \mathbf{S}_{\mathbf{xw}}^{(2)}(t)\end{array}\right]. \tag{33}$$

We denote the Householder transformation of a rectangular matrix $\mathbf{A}$ by $\boldsymbol{\Phi}(\mathbf{A})$. This transformation produces a squared and upper triangular matrix $\mathbf{S}=\boldsymbol{\Phi}(\mathbf{A})$ such that $\mathbf{AA}'=\mathbf{SS}'$. We refer to Norgaard et al. (2000) and their references for more information on this transformation and how to compute the $\mathbf{S}$ matrix. More formally, we let

$$\bar{\mathbf{S}}_{\mathbf{x}}(t+1)=\Phi\left(\left[\begin{array}{cccc}\mathbf{S}_{\mathbf{xx}}^{(1)}(t) & \mathbf{S}_{\mathbf{xw}}^{(1)}(t) & \mathbf{S}_{\mathbf{xx}}^{(2)}(t) & \mathbf{S}_{\mathbf{xw}}^{(2)}(t)\end{array}\right]\right). \tag{34}$$

From the construction of $\bar{\mathbf{S}}_x(t+1)$ and the definition of the Householder transformation

$$\bar{\mathbf{P}}_{\mathbf{xx}}(t+1)=\mathbf{S}_{\mathbf{xx}}^{(1)}(t)\mathbf{S}_{\mathbf{xx}}^{(1)}(t)'+\mathbf{S}_{\mathbf{xx}}^{(2)}(t)\mathbf{S}_{\mathbf{xx}}^{(2)}(t)'+\mathbf{S}_{\mathbf{xw}}^{(1)}(t)\mathbf{S}_{\mathbf{xw}}^{(1)}(t)'+\mathbf{S}_{\mathbf{xw}}^{(2)}(t)\mathbf{S}_{\mathbf{xw}}^{(2)}(t)'. \tag{35}$$

For instance, $\mathbf{S}_{\mathbf{xx}}^{(1)}(t)\mathbf{S}_{\mathbf{xx}}^{(1)}(t)'+\mathbf{S}_{\mathbf{xx}}^{(2)}(t)\mathbf{S}_{\mathbf{xx}}^{(2)}(t)'$ corresponds to $\mathbf{H}_{\mathbf{x},\mathbf{t}}\hat{\mathbf{P}}_{\mathbf{xx}}(t)\mathbf{H}_{\mathbf{x},\mathbf{t}}'$ in the Extended Kalman Filter. However, the former is a more accurate approximation than the latter.

The a priori estimator for the vector of observables is given by

$$\begin{aligned}\bar{\mathbf{y}}_{t+1} &= \frac{h^2-n_x-n_v}{h^2}\mathbf{g}\left(\bar{\mathbf{x}}_{t+1},\bar{\mathbf{v}}_{t+1};\boldsymbol{\theta}\right) \\ &+\frac{1}{2h^2}\sum_{p=1}^{n_x}\left(\mathbf{g}\left(\bar{\mathbf{x}}_{t+1}+h\bar{\mathbf{s}}_{\mathbf{x},p},\bar{\mathbf{v}}_{t+1};\boldsymbol{\theta}\right)+\mathbf{g}\left(\bar{\mathbf{x}}_{t+1}-h\bar{\mathbf{s}}_{\mathbf{x},p},\bar{\mathbf{v}}_{t+1};\boldsymbol{\theta}\right)\right) \\ &+\frac{1}{2h^2}\sum_{p=1}^{n_v}\left(\mathbf{g}\left(\bar{\mathbf{x}}_{t+1},\bar{\mathbf{v}}_{t+1}+h\mathbf{s}_{\mathbf{v},p};\boldsymbol{\theta}\right)+\mathbf{g}\left(\bar{\mathbf{x}}_{t+1},\bar{\mathbf{v}}_{t+1}-h\mathbf{s}_{\mathbf{v},p};\boldsymbol{\theta}\right)\right).\end{aligned} \tag{36}$$

The covariance matrix of this estimator is calculated based on

$$\bar{\mathbf{S}}_{\mathbf{y}}(t+1)=\boldsymbol{\Phi}\left(\left[\begin{array}{cccc}\mathbf{S}_{\mathbf{yx}}^{(1)}(t+1) & \mathbf{S}_{\mathbf{yv}}^{(1)}(t+1) & \mathbf{S}_{\mathbf{yx}}^{(2)}(t+1) & \mathbf{S}_{\mathbf{yv}}^{(2)}(t+1)\end{array}\right]\right), \tag{37}$$

and the Kalman gain is given by

$$\mathbf{K}_{t+1}=\bar{\mathbf{S}}_{\mathbf{x}}(t+1)\mathbf{S}_{\mathbf{yx}}^{(1)}(t+1)'\left[\bar{\mathbf{S}}_{\mathbf{y}}(t+1)\bar{\mathbf{S}}_{\mathbf{y}}(t+1)'\right]^{-1}. \tag{38}$$

Finally, the covariance matrix of the posterior state estimator follows from

$$\begin{aligned}\hat{\mathbf{S}}_{\mathbf{x}}(t+1) &= \Phi([\ \bar{\mathbf{S}}_{\mathbf{x}}(t+1) - \mathbf{K}_{t+1}\mathbf{S}^{(1)}_{\mathbf{yx}}(t+1) \quad \mathbf{K}_{t+1}\mathbf{S}^{(1)}_{\mathbf{yv}}(t+1) \\ &\qquad \mathbf{K}_{t+1}\mathbf{S}^{(2)}_{\mathbf{yx}}(t+1) \quad \mathbf{K}_{t+1}\mathbf{S}^{(2)}_{\mathbf{yv}}(t+1)\ ]).\end{aligned} \tag{39}$$

Note that $\bar{\mathbf{P}}_{\mathbf{xx}}(t+1)$, $\hat{\mathbf{P}}_{\mathbf{xx}}(t+1)$, and $\bar{\mathbf{P}}_{\mathbf{yy}}(t+1)$ by construction always are symmetric and positive semidefinite in the CDKF. An overview of the CDKF is given in appendix A.

Predictions for the observables are obtained by iterating (32), (34), and (36) forward in time. Särkkä (2008) shows how to derive a Forward-Backward smoother for the Unscented Kalman Filter, and Dunik & Simandl (2006) derive a square root implementation of this smoother.[2] We emphasize that these smoothers are derived based on the additional assumption that the filtered and the smoothed state distributions are multivariate normal. Given the results in Norgaard et al. (2000), it is straightforward to set up the Forward-Backward smoother for the CDKF. This is done in appendix B.

3.3 A QML estimator based on the CDKF

When deriving the CDKF in the previous two subsections, no likelihood methods were used. Hence, estimating the structural parameters ($\boldsymbol{\theta}$) by ML or Bayesian inference is in general not feasible with the CDKF. Instead, the structural parameters can be estimated by GMM or simulation based methods (Hansen (1982), Duffie & Singleton (1993) and Smith (1993)).

Next, consider the typical case where the state space system in (1) and (2) simplifies to

$$\mathbf{y}_t = \mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}) + \mathbf{v}_t \text{ and } \mathbf{v}_t \sim \mathcal{IID}(\mathbf{0}, \mathbf{R}_{\mathbf{v}}(t)) \tag{40}$$

$$\mathbf{x}_t = \mathbf{h}(\mathbf{x}_{t-1};\boldsymbol{\theta}) + \boldsymbol{\eta}\mathbf{w}_t \text{ and } \mathbf{w}_t \sim \mathcal{IID}(\mathbf{0}, \mathbf{R}_{\mathbf{w}}(t)). \tag{41}$$

This situation can occur when i) measurement errors are assumed to be present in the observables and ii) all structural shocks enter additively. The matrix $\boldsymbol{\eta}$ has dimension $n_x \times n_w$ and specifies the endogenous and exogenous state variables. If the distributions for $\mathbf{v}_t$ and $\mathbf{w}_t$ are bell-shaped, then it seems reasonable to assume that $\mathbf{y}_{t+1}|\,\mathbf{y}_{1:t}$ is approximately normally distributed, i.e.

$$\mathbf{y}_{t+1}|\,\mathbf{y}_{1:t} \overset{a}{\sim} \mathcal{N}\left(\bar{\mathbf{y}}_{t+1}, \bar{\mathbf{P}}_{\mathbf{yy}}(t+1);\boldsymbol{\theta}\right) \tag{42}$$

for $t = 1, ..., T$. If we further let the initial state vector $\mathbf{x}_0$ be uncorrelated with $\mathbf{v}_t$ and $\mathbf{w}_t$ for all values of t, then the quasi log-likelihood function for the entire sample is

$$\mathcal{L}(\boldsymbol{\theta}, \mathbf{y}_{1:T}) = \frac{-n_y T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\log\left(\left|\bar{\mathbf{P}}_{\mathbf{yy}}(t)\right|\right) - (\mathbf{y}_t - \bar{\mathbf{y}}_t)'\,\bar{\mathbf{P}}^{-1}_{\mathbf{yy}}(t)\,(\mathbf{y}_t - \bar{\mathbf{y}}_t)\right). \tag{43}$$

Thus, this quasi log-likelihood function is a smooth function in $\boldsymbol{\theta}$ and the function is therefore easy to optimize. This is in contrast to log-likelihood functions in particle filters where the

[2] The Unscented Kalman Filter (UKF) developed by Julier, Uhlmann & Durrant-Whyte (1995) is another derivative free implementation of the filtering equations presented in the previous subsection However, Norgaard et al. (2000) show that the CDKF has marginally higher theoretical accuracy than the UKF.

reported functions in general do not display smoothness in $\boldsymbol{\theta}$ (see Fernández-Villaverde & Rubio-Ramírez (2007*a*) for further details). Another important thing to note in (43) is that the first and second moments of this quasi log-likelihood function are correctly specified up to second-order accuracy. Bollerslev & Wooldrigde (1992) show consistency and asymptotic normality for the QML estimator when: i) only the first and second moments are correctly specified in the quasi log-likelihood function, ii) the quasi log-likelihood function is derived based on the normal distribution, and iii) standard regularity conditions hold (see Bollerslev & Wooldrigde (1992)). That is, for

$$\hat{\boldsymbol{\theta}}_{QML} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\max}\mathcal{L}\left(\boldsymbol{\theta}, \mathbf{y}_{1:T}\right)$$

it holds that

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_{QML} - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{A}_0^{-1}\mathbf{B}_0\mathbf{A}_0^{-1}\right), \tag{44}$$

where the zero subscript denotes the true value of theta ($\boldsymbol{\theta}_0$) and

$$\mathbf{A}_0 \equiv -E\left[\left.\frac{\partial^2\mathcal{L}\left(\boldsymbol{\theta}, \mathbf{y}_{1:T}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right] \tag{45}$$

$$\mathbf{B}_0 \equiv E\left[\mathbf{s}\left(\boldsymbol{\theta}_0, \mathbf{y}_{1:T}\right)\mathbf{s}\left(\boldsymbol{\theta}_0, \mathbf{y}_{1:T}\right)'\right] \tag{46}$$

$$\mathbf{s}\left(\boldsymbol{\theta}_0, \mathbf{y}_{1:T}\right) \equiv \left.\frac{\partial\mathcal{L}\left(\boldsymbol{\theta}, \mathbf{y}_{1:T}\right)}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}. \tag{47}$$

Thus, we can estimate $\boldsymbol{\theta}$ by QML provided that second-order accuracy of the first and second moments in the quasi log-likelihood function is sufficient. To assess whether this is the case, we must compare this degree of precision to the chosen approximation order of the DSGE model. We introduce our way of reasoning by starting with a linerized DSGE model. Here, first and second moments are only accurate up to first and second order, respectively, and the CDKF reduces to the standard Kalman Filter which exactly captures the first and second moments to the desired degree of precision. Thus, we recover the standard result that the QML estimator is consistent and asymptotically normal for a linerized DSGE model.[3]

When a DSGE model is approximated up to second order accuracy, then first and second moments in the model are accurate up to second and third order, respectively. This implies that the second order precision in the CDKF is sufficiently accurate for the first moment, but not for the second moments where approximation errors are present in the third order terms. However, these errors are likely to be insignificant because the second moments of $\mathbf{y}_{t+1}|\,\mathbf{y}_{1:t}$ are often very small. The validity of such an argument can in all cases be verified directly by inspection of $\bar{\mathbf{P}}_{\mathbf{yy}}\left(t\right)$.[4] Thus, when a DSGE model is approximated up to second order, the precision delivered by the CDKF should in all realistic settings be sufficient and the QML estimator can be expected to be consistent and asymptotically normal.

For a DSGE model approximated up to third order, it holds that first and second moments in the model are accurate up to third and fourth order, respectively. Using a similar argument as above, the third and fourth order terms in the second moments are likely to be insignificant. Hence, the precision deliver by the CDKF is sufficient if third order terms in the first moment

[3] Inference is here used in the sense that the approximated DSGE model is the true data generating process.

[4] In our benchmark model specified below, the largest term in $\bar{\mathbf{P}}_{\mathbf{yy}}\left(t\right)$ attains values around 0.0003 and the typical values are around 0.00005.

are small. Alternatively, if the state vector is taken to be normally distributed and the structural shocks and the measurement noise are normally distributed, then the CDKF is actually accurate up to third order (Norgaard et al. (2000)). Given this assumption, the CDKF correctly captures the first moment of $\mathbf{y}_{t+1}|\,\mathbf{y}_{1:t}$, and errors in $\bar{\mathbf{P}}_{\mathbf{yy}}(t)$ are only present in fourth order terms. Thus, if i) third order terms are small or if ii) the state vector, the shocks, and the measurement errors all are normally distributed, then the QML estimator can be expected to be consistent and asymptotically normal even for DSGE models approximated up to third order.

Finally, when a fourth or higher approximation order is used for DSGE models, approximation errors are present in the first and second moment of $\mathbf{y}_{t+1}|\,\mathbf{y}_{1:t}$. The size of the errors in the first moment are difficult to evaluate a priori and this implies that consistency and asymptotic normality for the QML estimator cannot be guarantied.

4 Particle Filters

The objective in particle filters is to recursively estimate the entire probability distribution for the sequence of unknown state vector given all available information at a given point in time ($\mathbf{y}_{1:t}$). This posterior state distribution is denoted by $p(\mathbf{x}_{0:t}|\,\mathbf{y}_{1:t})$ where $\mathbf{x}_{0:t} \equiv \{\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_t\}$, and the unknown state vector is typically estimated by the mean of the marginal distribution $p(\mathbf{x}_t|\,\mathbf{y}_{1:t})$. The recursive estimation of $p(\mathbf{x}_{0:t}|\,\mathbf{y}_{1:t})$ is done based on the probability structure of the state space in (1) and (2) and by sequential use of importance sampling and resampling. Thus, particle filters do not impose restrictions on the updating rule for the posterior state vector as done in the previous section.

The outline for the remaining part of this section is as follows. We proceed by describing the standard PF as presented in Doucet, de Freitas & Gordon (2001*a*). A few extensions of this algorithm are then briefly discussed in order to motivate our own extension of the standard PF.

We adopt the standard notation in terms of the state vector $\mathbf{x}_t$ during the following presentation. Hence, to avoid stochastic singularity it is assumed that the dimension of $\mathbf{v}_t$ is equal to or larger than the number of observables, i.e. $n_v \geq n_y$. 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*a*). We show in appendix C that the CDKF also can be used in this case.

4.1 The standard particle filter

The first requirement for particle filters is that we can evaluate the two conditional probabilities $p(\mathbf{y}_t|\,\mathbf{x}_t;\boldsymbol{\theta})$ and $p(\mathbf{x}_t|\,\mathbf{x}_{t-1};\boldsymbol{\theta})$ for all values of $\mathbf{x}_t$ and $\mathbf{y}_t$ for $t = 1, ..., T$. Using the well-known result we have

$$p(\mathbf{y}_t|\,\mathbf{x}_t;\boldsymbol{\theta}) = p(\mathbf{v}_t;\boldsymbol{\theta}) \left|\det\left(\frac{\partial \mathbf{g}(\mathbf{x}_t, \mathbf{v}_t;\boldsymbol{\theta})}{\partial \mathbf{v}_t}\right)\right|^{-1} \tag{48}$$

$$p(\mathbf{x}_t|\,\mathbf{x}_{t-1};\boldsymbol{\theta}) = p(\mathbf{w}_t;\boldsymbol{\theta}) \left|\det\left(\frac{\partial \mathbf{h}(\mathbf{x}_{t-1}, \mathbf{w}_t;\boldsymbol{\theta})}{\partial \mathbf{w}_t}\right)\right|^{-1}, \tag{49}$$

provided the Jacobian of $\mathbf{g}(\cdot)$ and $\mathbf{h}(\cdot)$ exist and their determinants are different from zero. We henceforth assume that the distributions in (48) and (49) are always well-defined.

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

$$p\left(\mathbf{x}_{0:t+1}|\,\mathbf{y}_{1:t+1};\boldsymbol{\theta}\right)=p\left(\mathbf{x}_{0:t}|\,\mathbf{y}_{1:t};\boldsymbol{\theta}\right)\frac{p\left(\mathbf{y}_{t+1}|\,\mathbf{x}_{t+1};\boldsymbol{\theta}\right)p\left(\mathbf{x}_{t+1}|\,\mathbf{x}_{t};\boldsymbol{\theta}\right)}{p\left(\mathbf{y}_{t+1}|\,\mathbf{y}_{1:t};\boldsymbol{\theta}\right)}. \tag{50}$$

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

$$p\left(\mathbf{x}_{0:t}|\,\mathbf{y}_{1:t};\boldsymbol{\theta}\right)\approx\sum_{i=1}^{N}w_{t}^{(i)}\delta\left(\mathbf{x}_{0:t}-\mathbf{x}_{0:t}^{(i)}\right). \tag{51}$$

Here, $\delta\left(\cdot\right)$ denotes the Dirac delta function and $\left\{\mathbf{x}_{0:t}^{(i)}\right\}_{i=1}^{N}$ is a random draw from $p\left(\mathbf{x}_{0:t}|\,\mathbf{y}_{1:t};\boldsymbol{\theta}\right)$. 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 by $\pi\left(\cdot\right)$ and it is specified by

$$\pi\left(\mathbf{x}_{0:t+1}|\,\mathbf{y}_{1:t+1}\right)=\pi\left(\mathbf{x}_{0:t}|\,\mathbf{y}_{1:t}\right)\pi\left(\mathbf{x}_{t+1}|\,\mathbf{x}_{0:t},\mathbf{y}_{1:t+1}\right). \tag{52}$$

The structure of $\pi\left(\cdot\right)$ implies that $p\left(\mathbf{x}_{0:t+1}|\,\mathbf{y}_{1:t+1};\boldsymbol{\theta}\right)$ is approximated without modifying the past estimated state values, $\mathbf{x}_{0:t}$. The specific form of $\pi\left(\cdot\right)$ should be chosen such that its support includes that of the posterior state distribution, $p\left(\mathbf{x}_{0:t+1}|\,\mathbf{y}_{1:t+1};\boldsymbol{\theta}\right)$. It is straightforward to show that the proposal distribution in (52) gives rise to the following recursive formula for the importance sampling weights $\left(w_{t+1}^{(i)}\right)$

$$w_{t+1}^{(i)}=w_{t}^{(i)}\frac{p\left(\mathbf{y}_{t+1}|\,\mathbf{x}_{t+1}^{(i)};\boldsymbol{\theta}\right)p\left(\mathbf{x}_{t+1}^{(i)}\middle|\,\mathbf{x}_{t}^{(i)};\boldsymbol{\theta}\right)}{\pi\left(\mathbf{x}_{t+1}^{(i)}\middle|\,\mathbf{x}_{0:t}^{(i)},\mathbf{y}_{1:t+1}\right)}\qquad\text{,for }i=1,...,N. \tag{53}$$

The normalized importance samplings weights are given by

$$\tilde{w}_{t+1}^{(i)}=\frac{w_{t+1}^{(i)}}{\sum_{i=1}^{N}w_{t+1}^{(i)}}\qquad\text{,for }i=1,...,N. \tag{54}$$

A random sample from $p\left(\mathbf{x}_{0:t+1}|\,\mathbf{y}_{1:t+1};\boldsymbol{\theta}\right)$ 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 deteriorates in precision, a phenomenon referred to as the "sample depletion problem". The purpose of the resampling step is to mitigate this problem by eleminating particles which are far from the true state vector (i.e. particles with low values of $\tilde{w}_{t+1}^{(i)}$) and multiplying particles which are close to the true state vector (i.e. particles with high values of $\tilde{w}_{t+1}^{(i)}$).

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

$$E\left[f\left(\mathbf{x}_{0:t+1}\right)\right]=\frac{1}{N}\sum_{i=1}^{N}f\left(\hat{\mathbf{x}}_{0:t+1}^{(i)}\right). \tag{55}$$

Here, Berzuini, Best, Gilks & Larizza (1997) provide a central limit theorem for this estimator.[5]

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

$$p\left(\mathbf{y}_{t+1}|\,\mathbf{y}_{1:t};\boldsymbol{\theta}\right)=\sum_{i=1}^{N}w_{t+1}^{(i)}. \tag{56}$$

Doucet, Godsill & Andrieu (2000) show that the following proposal distribution

$$\pi\left(\mathbf{x}_{t+1}|\,\mathbf{x}_{0:t},\mathbf{y}_{1:t+1}\right)=p\left(\mathbf{x}_{t+1}|\,\mathbf{x}_{t},\mathbf{y}_{t+1};\boldsymbol{\theta}\right) \tag{57}$$

is optimal in the sense that it minimizes the variance of the importance weights given $\mathbf{x}_{0:t}$ and $\mathbf{y}_{1:t+1}$. This proposal distribution is in general intractable and approximations are therefore needed. The standard PF approximates $\pi\left(\cdot\right)$ by the transition distribution, i.e.

$$p\left(\mathbf{x}_{t+1}|\,\mathbf{x}_{t},\mathbf{y}_{t+1};\boldsymbol{\theta}\right)\simeq p\left(\mathbf{x}_{t+1}|\,\mathbf{x}_{t};\boldsymbol{\theta}\right). \tag{58}$$

This is an obvious choice because 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 weights reduces to $w_{t+1}^{(i)}=w_{t}^{(i)}p\left(\mathbf{y}_{t+1}|\,\mathbf{x}_{t+1}^{(i)};\boldsymbol{\theta}\right)$ in this case. However, the transition distribution does not use information about new observables ($\mathbf{y}_{t+1}$), and this proposal distribution is therefore said to be "blind". It is in this sense that the proposal distribution in the standard PF is sub-optimal. An overview of the standard PF is presented in appendix D.

4.2 Two extensions of the standard Particle Filter

Omitting information about new observables in the proposal distribution is unfortunate for two reasons. First, if new observations are very informative about the state values, then 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 with a high signal to noise ratio. Second, 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 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 uneven distributed in such a situation, and as a result many particles are needed to get a satisfying approximation of $p\left(\mathbf{x}_{0:t+1}|\,\mathbf{y}_{1:t+1};\boldsymbol{\theta}\right)$.

One way to improve the performance of the standard PF is therefore to include information about new observables in the proposal distribution. Doucet et al. (2000) suggest to do this by using the Extended Kalman Filter (EKF) to generate a Gaussian approximation of the optimal proposal distribution in (57). This is done by sending *each* particle through one iteration in the

[5] Liu & Chen (1998) recommend to do state estimation before the resampling step, i.e. by $E\left[f\left(\mathbf{x}_{0:t+1}\right)\right]=\sum_{i=1}^{N}\tilde{\omega}_{t+1}^{(i)}f\left(\mathbf{x}_{0:t+1}^{(i)}\right)$, because the resampling step introduces additional random variation in the sample of particles. For 5,000 or more particles, the two estimators give almost identical results in our case, and the estimator by Liu & Chen (1998) does not dominate (55). Moreover, the estimator in (55) is computationally faster than the one recommended by Liu & Chen (1998), because $\frac{1}{N}\sum_{i=1}^{N}f\left(\hat{\mathbf{x}}_{0:t+1}^{(i)}\right)=\frac{1}{N}\sum_{i=1}^{n}N_{i}f\left(\hat{\mathbf{x}}_{0:t+1}^{(i)}\right)$ where N_i is the number of repetitions of the i'th particle and n is typical must smaller than N.

EKF to generate a new probability distribution, where the mean and the covariance matrix thus contain information from new observables. The hope is that sampling from these distributions moves particles to areas of high likelihood. More formally,

$$p\left(\mathbf{x}_{t+1}|\,\mathbf{x}_t^{(i)},\mathbf{y}_{t+1}\right)\simeq\mathcal{N}\left(\hat{\mathbf{x}}_{t+1}^{EKF,(i)},\hat{\mathbf{P}}_{\mathbf{xx}}^{EKF,(i)}\left(t+1\right)\right)\quad\text{,for } i=1,...,N. \tag{59}$$

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 PF leads to the Extended Kalman Particle Filter (EKPF). Doucet et al. (2000) and Merwe, Doucet, de Freitas & Wan (2000) show that using the proposal distribution in (59) gives more precise state estimates compared to the state estimates from the standard PF.

However, two drawbacks are related to (59). First, the EKPF is very time consuming to implement, because the mean and the covariance matrix in (59) must be calculated for a large number of particles in each time period.[6] Second, the approximations of the first and second moments in the EKF are only accurate up to first-order, and the approximations do not take the probability distribution for the state vector into account.

Merwe et al. (2000) and Merwe & Wan (2003) suggest to solve the second problem related to the EKPF by using the CDKF to calculate more accurate expressions for the mean and the covariance matrix in (59). That is, they replace (59) with

$$p\left(\mathbf{x}_{t+1}|\,\mathbf{x}_t^{(i)},\mathbf{y}_{t+1}\right)\simeq\mathcal{N}\left(\hat{\mathbf{x}}_{t+1}^{CDKF,(i)},\hat{\mathbf{P}}_{\mathbf{xx}}^{CDKF,(i)}\left(t+1\right)\right)\quad\text{,for } i=1,...,N \tag{60}$$

and this new filter is called the Sigma Point Particle Filter (SPPF).[7] In an application Merwe et al. (2000) show that the SPPF clearly outperforms the EKPF. However, the SPPF is like the EKPF very time consuming to calculate. For instance, Merwe & Wan (2003) report that the SPPF is 20 times slower to compute than the standard PF for a dynamic state space model with only one state variable and one observable! They use 500 particles in both cases. Thus, the computational requirement for the SPPF is so severe that the filter in practice is infeasible in the context of DSGE models.

4.3 The Mean Shifted Particle Filter

This section addresses the computational issue for merging the CDKF with the PF. We do so by suggesting a third approximation of the optimal proposal distribution in (57).

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

[7] Merwe (2004) shows that the CDKF and the Unscented Kalman Filter can be nested in the class of Sigma Point Kalman Filters. Thus, one only needs to choose a Sigma Point Kalman Filter to calculate the mean and the covariance matrix in this proposal distribution. This motivates the name, Sigma Point Particle Filter. However, Merwe & Wan (2003) prefer to use the CDKF and this motivate our presentation.

Note first that the proposal distribution in the SPPF corresponds to

$$\mathbf{x}_{t+1}^{(i)} = \hat{\mathbf{x}}_t^{(i)} + \boldsymbol{\mu}_{t+1}^{CDKF,(i)} + \hat{\mathbf{S}}_{\mathbf{x}}^{CDKF,(i)}(t+1)\,\boldsymbol{\epsilon}_{t+1}^{(i)} \qquad \text{for } i = 1, ..., N, \tag{61}$$

where $\boldsymbol{\epsilon}_{t+1} \sim \mathcal{NID}(\mathbf{0}, \mathbf{I})$ and

$$\boldsymbol{\mu}_{t+1}^{(i)} \equiv \hat{\mathbf{x}}_{t+1}^{CDKF,(i)} - \hat{\mathbf{x}}_t^{(i)} \qquad \text{for } i = 1, ..., N. \tag{62}$$

Thus, the proposal distribution in the SPPF can be interpreted as a mean correction $\left(\boldsymbol{\mu}_{t+1}^{(i)}\right)$ of draws from the previous posterior state distribution $\left(\hat{\mathbf{x}}_t^{(i)}\right)$ plus a Gaussian noise component $\left(\mathbf{S}_{\mathbf{x}}^{CDKF,(i)}(t+1)\,\boldsymbol{\epsilon}_{t+1}\right)$. We emphasize that $\boldsymbol{\mu}_{t+1}^{(i)}$ and $\hat{\mathbf{S}}_{\mathbf{x}}^{CDKF,(i)}(t+1)$ are functions of $\mathbf{y}_{t+1}$ and that both functions may differ for different particles in (61). The latter feature is what makes the SPPF filter so time consuming to calculate. Note also that the reason the SPPF sends *each* particle, and not just one or a small subset of particles, through the CDKF is to preserve characteristics such as i) multimodal features, ii) thick tails, etc. in the previous posterior state distribution. These characteristics may be important when estimating the current posterior state distribution.

To reduce the computational cost of including information about current observables in the proposal distribution, we suggest to replace (61) by

$$\mathbf{x}_{t+1}^{(i)} = \hat{\mathbf{x}}_t^{(i)} + \boldsymbol{\mu}_{t+1} + \hat{\mathbf{S}}_{\mathbf{x}}^{CDKF}(t+1)\,\boldsymbol{\epsilon}_{t+1}^{(i)} \quad \text{for } i = 1, ..., N. \tag{63}$$

That is, $\boldsymbol{\mu}_{t+1}$ and $\hat{\mathbf{S}}_{\mathbf{x}}^{CDKF}(t+1)$ are here the same for all particles. Thus, our new proposal distribution corresponds to using the jittering method proposed by Gordon, Salmond & Smith (1993) on the previous estimated state distribution with a mean correction. The idea is now to calculate the values of $\boldsymbol{\mu}_{t+1}$ and $\hat{\mathbf{S}}_{\mathbf{x}}^{CDKF}$ by sending the posterior state estimate in the previous period ($\hat{\mathbf{x}}_t$) through one iteration in the CDKF where information about $\mathbf{y}_{t+1}$ is used. Then we let

$$\boldsymbol{\mu}_{t+1} \equiv \hat{\mathbf{x}}_{t+1}^{CDKF} - \hat{\mathbf{x}}_t. \tag{64}$$

In this way we still incorporate information about new observables in the proposal distribution, and it is done in a very efficient way, because only one iteration in the CDKF is needed in each time period. Moreover, all features from the previous posterior state distribution are preserved in (63).

The defining feature of our new proposal distribution in (63) is the mean shifting operation. We therefore refer to a particle filter with this proposal distribution as the Mean Shifted Particle Filter (MSPF). In our implementation of the MSPF we choose to use one iteration in the standard PF to estimate the first value of $\hat{\mathbf{S}}_{\mathbf{x}}(t)$. An overview of the MSPF is given in appendix E.

Prediction for the observables in the MSPF is straightforward, and smoothing can be done alone the lines described by Simon J. Godsill & West (2004). The values for the unknown structural parameters in the economy ($\boldsymbol{\theta}$) can be estimated by standard ML or Bayesian methods as described by Fernández-Villaverde & Rubio-Ramírez (2007*a*).

5 A New Keynesian DSGE model

This section presents the DSGE model which we will use in the subsequent Monte Carlo study. To avoid making the results in this Monte Carlo study dependent on the specific parameterization of our DSGE model a sequence of "test economies" is constructed. That is, we generate a sequence of DSGE models with different combinations of the structural parameters. The idea is then to simulate data series from these economies, and based on this data to measure the precision of the filters. The construction of the test economies is the topic in the next three subsections.

5.1 The Model

The DSGE model we use has the same basic structure as the models developed by Christiano et al. (2005), Altig et al. (2005), and Schmitt-Grohé & Uribe (2006). We refer to these papers for additional details. When presenting our model, we use the notation from the macroeconomic literature. Hence, the notation in this section is unrelated to the notation used for the filters in previous sections.

The households: We start by assuming that the behavior of the households may be described by a representative household. The household's preferences are specified by a utility function defined over real per capita consumption (c_t) and per capita labor supply (h_t)

$$U_t = E_t \sum_{l=0}^{\infty} \beta^l u\left(c_{t+l} - bc_{t-1+l}, h_{t+l}\right). \tag{65}$$

Here, E_t is the conditional expectation given information available at time t and $\beta \in [0,1[$. The function $u(\cdot,\cdot)$ is a period utility index which we assume has the standard form

$$u\left(c_t - bc_{t-1}, h_t\right) = \frac{\left(\left(c_t - bc_{t-1}\right)^{1-\phi_4}\left(1-h_t\right)^{\phi_4}\right)^{1-\phi_3} - 1}{1-\phi_3},$$

where $b \in [0,1]$, $\phi_3 \in]0,1[\cup]1,\infty[$, and $\phi_4 \in]0,1[$. The parameter b specifics the degree of the internal habit effect in the consumption good. This good is constructed from a continuum of differentiated goods ($c_{i,t}, i \in [0,1]$) and the aggregation function

$$c_t = \left[\int_0^1 c_{i,t}^{\frac{\eta-1}{\eta}} di\right]^{\frac{\eta}{\eta-1}}. \tag{66}$$

Here, $\eta > 1$ is the intratemporal elasticity of substitution across the differentiated goods.

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

$$k_{t+1} = (1-\delta)k_t + i_t\left(1 - S\left(\frac{i_t}{i_{t-1}}\right)\right). \tag{67}$$

The parameter $\delta \in [0,1]$ is the depreciation rate for the capital stock and i_t is gross investments. The function $S\left(\frac{i_t}{i_{t-1}}\right) = \frac{\kappa}{2}(\frac{i_t}{i_{t-1}} - \mu_i)^2$ with $\kappa \geq 0$ adds investment adjustment costs to the

economy based on changes in the growth rate of investments. The value of μ_i is determined in such a way 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

$$E_t r_{t,t+1} x_{t+1}^h + c_t \left(1 + l\left(v_t\right)\right) + \left(e_t \Upsilon_t\right)^{-1} i_t + m_t^h = \frac{x_t^h + m_{t-1}^h}{\pi_t} + w_t h_t + \phi_t \tag{68}$$

The function $l\left(v_t\right) = \phi_1 v_t + \phi_2 / v_t - 2\left(\phi_1 \phi_2\right)^{0,5}$ determines the transactional costs imposed on the household based on the velocity $v_t \equiv c_t / m_t^h$, where $\phi_1 \geq 0$ and ϕ_2 are subject to the constraint that $l\left(v_t\right) \geq 0$. The left hand side of (68) is the household's total expenditures in period t which include: i) state-contingent claims $\left(E_t r_{t,t+1} x_{t+1}^h\right)$, ii) consumption including transaction costs $\left(c_t\left[1 + l\left(v_t\right)\right]\right)$, iii) investments $\left(e_t \Upsilon_t\right)^{-1} i_t$, and iv) the real money holdings $\left(m_t^h\right)$. Changes in $e_t \Upsilon_t$ are investment specific shocks, which we specify as an exogenous AR(1) process with a deterministic trend, i.e.

$$\ln \Upsilon_{t+1} = \ln \Upsilon_t + \ln\left(\mu_{\Upsilon,ss}\right) \tag{69}$$

$$\ln\left(e_{t+1}\right) = \rho_e \ln\left(e_t\right) + \epsilon_{e,t+1}. \tag{70}$$

We let $\epsilon_{e,t+1}$ be independent and identically distributed according to the Generalized Error Distribution (GED) with: i) mean zero, ii) variance $Var\left(\epsilon_{e,t+1}\right)$, and iii) tail thickness parameter $\xi_e \in (0, \infty)$. This is denoted by $\epsilon_{e,t+1} \sim \mathcal{GID}\left(0, Var\left(\epsilon_{e,t+1}\right), \xi_e\right)$. For $\xi_e = 2$ the GED distribution reduces to the normal distribution, and for $\xi_e < 2$ the GED distribution has thicker tails than the normal distribution, and vice versa for $\xi_e > 2$. We also require that $\rho_e \in]-1, 1[$.

The right hand side of (68) is the household's total wealth in period t which consists of: i) pay-off from state-contingent assets purchased in period $t-1$ $\left(x_t^h / \pi_t\right)$, ii) the real money holdings from the previous period $\left(m_{t-1}^h / \pi_t\right)$, iii) real labor income $\left(w_t h_t\right)$, and iv) dividends received from the firms $\left(\phi_t\right)$. Note that π_t is the gross inflation rate. The dividend payments are restricted to be zero in steady state.

The firms: The production in the economy is assumed to be undertaken by a continuum of firms, indexed by $i \in [0, 1]$. Here, we adopt the standard assumption that each firm supplies a differentiable good $\left(y_{i,t}^s\right)$ to the goods market which is characterized by monopolistic competition with no exit or entry. Furthermore, all firms have access to the same technology given as follows

$$y_{i,t}^s = \begin{cases} a_t F\left(k_{i,t}, z_t h_{i,t}\right) - \psi_t z_t^* & \text{if } F\left(k_{i,t}, z_t h_{i,t}\right) - \psi_t z_t^* > 0 \\ 0 & \text{else} \end{cases} \tag{71}$$

where $F\left(\cdot\right) \equiv k_{i,t}^{\theta}\left(z_t h_{i,t}\right)^{1-\theta}$ with $\theta \in]0, 1[$. Here, $k_{i,t}$ and $h_{i,t}$ denote physical capital and labor services used by the i'th firm, respectively. The variable a_t denotes stationary technology shocks, and we let

$$\ln a_{t+1} = \rho_a \ln a_t + \epsilon_{a,t+1}, \tag{72}$$

where $\epsilon_{a,t+1} \sim \mathcal{GID}\left(0, Var\left(\epsilon_{a,t+1}\right), \xi_a\right)$ and $\rho_a \in]-1, 1[$. The variable z_t in (71) denotes a non-stationary technology shock. For these shocks, we let $\mu_{z,t} \equiv z_t / z_{t-1}$ and assume

$$\ln\left(\frac{\mu_{z,t+1}}{\mu_{z,ss}}\right) = \rho_z \ln\left(\frac{\mu_{z,t}}{\mu_{z,ss}}\right) + \epsilon_{z,t+1}, \tag{73}$$

where $\epsilon_{z,t+1} \sim \mathcal{GID}\left(0, Var\left(\epsilon_{z,t}\right), \xi_z\right)$ and $\rho_z \in]-1,1[$. We emphasize that a_t and $\mu_{z,t}$ are mutually independent shocks, and so are all the other exogenous shocks in our DSGE model. Following Altig et al. (2005), we define z_t^* by the relation $z_t^* \equiv \Upsilon_t^{\theta/(1-\theta)} z_t$ which implies that we may interpret z_t^* as an overall measure of technological progress in the economy.

Smets & Wouters (2003) introduce real supply shocks in this framework by letting firms' markup rates be subject to random shocks. With Calvo price contracts such markup shocks prevent an exact recursive representation of our economy, which is needed for a non-linear approximation of the economy. Instead, we introduce real supply shocks by letting the firms' fixed costs be time-varying beyond the variation in z_t^*. The inclusion of these real supply shocks can be motivated by variation in firms' fixed costs due to changes in: i) oil prices, ii) maintenance costs, iii) firms' subsidies, etc. We let

$$\ln\left(\frac{\psi_{t+1}}{\psi_{ss}}\right) = \rho_\psi \ln\left(\frac{\psi_t}{\psi_{ss}}\right) + \epsilon_{\psi,t+1}, \tag{74}$$

where $\epsilon_{\psi,t+1} \sim \mathcal{GID}\left(0, Var\left(\epsilon_{\psi,t+1}\right), \xi_\psi\right)$ and $\rho_\psi \in]-1,1[$.

All firms are assumed to maximize the present value of their nominal dividend payments, denoted $d_{i,t}$. That is, each firm maximizes

$$d_{i,t} \equiv E_t \sum_{l=0}^{\infty} r_{t,t+l} P_{t+l} \phi_{i,t+l}, \tag{75}$$

where $r_{t,t+l}$ is the stochastic discount factor and the expression for real dividend payments from the i'th firm $\left(\phi_{i,t}\right)$ is given below in (77). 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 including transaction costs and investments. We make the standard assumption that the aggregation function for the latter component coincide with the aggregation function for consumption in (66). Hence, the restriction on the aggregate demand can be written as

$$y_t^d = c_t\left(1 + l\left(v_t\right)\right) + \left(e_t \Upsilon_t\right)^{-1} i_t. \tag{76}$$

In addition, we assume that the firms satisfy demand, i.e. $y_{i,t}^s \geq y_{i,t}^d$ for all $i \in [0,1]$.

The second restriction is a cash-in-advance constraint on a fraction ν of the firms' payments to workers. Thus, the money demanded by the i'th firm is $m_{i,t}^f = \nu w_t h_{i,t}$.

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

$$\phi_{i,t} = \left(P_{i,t}/P_t\right) y_{i,t}^d - r_t^k k_{i,t} - w_t h_{i,t} - m_{i,t}^f \left(1 - R_{t,1}^{-1}\right) \tag{77}$$

$$-E_t r_{t,t+1} x_{i,t+1}^f + m_{i,t}^f - \pi_t^{-1}\left(x_{i,t}^f + m_{i,t-1}^f\right).$$

The first term in (77) denotes the real revenue from sales of the i'th good. The next terms in (77) are the firm's expenditures which are allocated to: i) purchase of capital services $\left(r_t^k k_{i,t}\right)$, ii) payments to the workers $\left(w_t h_{i,t}\right)$, and iii) opportunity costs of holding money due to the cash-in-advance constraint $\left(m_{i,t}^f \left(1 - R_{t,1}^{-1}\right)\right)$. The final terms in (77) denote the change in the

firm's real financial wealth.

The fourth constraint introduces staggered price adjustments. We make the standard assumption that in each period a fraction $\alpha \in [0,1[$ of randomly picked 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 assume that the central bank determines the nominal interest rate according to a forward looking Taylor rule of the form

$$\ln\left(\frac{R_{t,1}}{R_{ss,1}}\right) = \alpha_R \ln\left(\frac{R_{t-1,1}}{R_{ss,1}}\right) + \alpha_\pi \ln E_t\left(\frac{\pi_{t+1}}{\pi^*_{t+1}}\right) + \alpha_y \ln E_t\left(\frac{y^d_{t+1}/y^d_t}{\mu_y}\right), \tag{78}$$

where π^*_t is a time-varying inflation rate target. The parameters in this rule are subject to the constraints: i) $\alpha_R \in [0,1]$, ii) $\alpha_\pi \geq 0$, and iii) $\alpha_y \geq 0$. We use a standard specification of the inflation rate target by letting π^*_t be a weighted sum of all previous inflation rates and a noise component (see Bekaert, Cho & Moreno (2005))

$$\pi^*_{t+1} = (1-\omega^*)\sum_{j=0}^{\infty}(\omega^*)^j\left(\pi_{t-j} + \frac{\epsilon_{\pi^*,t+1-j}}{1-\omega^*}\right). \tag{79}$$

The restriction $\omega^* \in [0,1)$ is imposed to get smooth changes in the target, and we let $\epsilon_{\pi^*,t} \sim \mathcal{GID}\left(0, Var\left(\epsilon_{\pi^*,t}\right), \xi_{\pi^*}\right)$.

5.2 Solving the DSGE model

It is straightforward to show that market clearing conditions and the first order conditions for the household and the firms can be written in the following way

$$E_t\left[\mathbf{f}\left(\mathbf{y}_{t+1}, \mathbf{y}_t, \mathbf{x}_{t+1}, \mathbf{x}_t\right)\right] = \mathbf{0}, \tag{80}$$

where i) $\mathbf{f}\left(\cdot,\cdot,\cdot,\cdot\right)$ is a real-valued function, ii) $\mathbf{y}_t$ is a 7 dimensional control vector, and iii) $\mathbf{x}_t$ is a 11 dimensional state vector.[8] Following Schmitt-Grohé & Uribe (2004), we introduce σ as a perturbation parameter scaling the matrix $\boldsymbol{\eta}$ containing standard deviations for the structural shocks. Given this assumption, the solution to this class of economies is then given by

$$\mathbf{y}_t = \mathbf{g}\left(\mathbf{x}_t, \sigma\right) \tag{81}$$

$$\mathbf{x}_{t+1} = \mathbf{h}\left(\mathbf{x}_t, \sigma\right) + \sigma\boldsymbol{\eta}\mathbf{w}_{t+1}. \tag{82}$$

The important thing to notice is that the structural shocks ($\mathbf{w}_{t+1}$) enter linearly in the state equations regardless of the chosen approximation order. The functions $\mathbf{g}\left(\cdot\right)$ and $\mathbf{h}\left(\cdot\right)$ 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 (2003)).[9]

[8] We refer to the paper's technical appendix for the proof of this statement and the elements of $\mathbf{y}_t$ and $\mathbf{x}_t$. The technical appendix is avaliable on request.

[9] We refer to the papers technical appendix for additional details.

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

$$\mathbf{y}_t^{obs} = \mathbf{M}_1 \mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta}) - \mathbf{M}_2 \mathbf{g}(\mathbf{x}_{t-1}, \boldsymbol{\theta}) + \mathbf{v}_t \tag{83}$$

$$\mathbf{x}_{t+1} = \mathbf{h}(\mathbf{x}_t, \sigma) + \sigma \boldsymbol{\eta} \mathbf{w}_{t+1} \tag{84}$$

where $\mathbf{M}_1$ and $\mathbf{M}_2$ are selection matrices with appropriate dimensions.

5.3 Determining the structural parameters

The final step in constructing the test economies is to determine the values for the structural parameters. To make our test economies as representative as possible for the DSGE models in the literature, we adopt the following strategy: i) select appropriate sample intervals for the structural parameters based on estimation and calibration results in the literature, and ii) draw uniformly from these intervals to generate 100 test economies. Where possible, the mean values for these sampling intervals are determined based on the results in Altig et al. (2005), Schmitt-Grohé & Uribe (2006), and Fernández-Villaverde & Rubio-Ramirez (2007*b*). The size of the sampling intervals is determined to capture much of the uncertainty about the true value of the parameters. Alternatively, these sampling intervals can also be considered as the area where we would search for the most likely structural parameters if we were to estimate our DSGE model. During such a search, we still need to be able to estimate the state variables accurately, even though we may be far away from the most likely parameter values. The intervals are shown in table 1.

< Table 1 about here >

In setting up the sampling intervals in table 1, we deviate slightly in the following cases from the calibrated or estimated values in the papers by Altig et al. (2005), Schmitt-Grohé & Uribe (2006), and Fernández-Villaverde & Rubio-Ramirez (2007*b*). First, the mean value for the interval of the household's discount parameter (β) is set equal to 0.9992, which gives an annual interest rate of 5.4% in the steady state. This level corresponds to the average three month interest rate in the postwar US economy. Schmitt-Grohé & Uribe (2006) let $\beta = 1.03^{-1/4}$ and Fernández-Villaverde & Rubio-Ramirez (2007*b*) estimate β to 0.9999. Given the other parameters of our model, these alternative values of β lead to a too high ($R_{ss} = 8.2\%$) or too low ($R_{ss} = 5.1\%$) value of the annual interest rate in the steady state, respectively. Second, we let the interval for the degree of internal habit formation (b) have a mean value of 0.1 to get a satisfying degree of persistency in the simulated series for the interest rate in our model. Values of b around 0.69 or higher as in Altig et al. (2005), Schmitt-Grohé & Uribe (2006), and Fernández-Villaverde & Rubio-Ramirez (2007*b*) imply too volatile interest rates given the other parameters in our model. Third, the mean value in the interval for α_π (the central banks reaction to deviations from the inflation target) is set equal to 12 to ensure a unique and stable solution in our model. Given this value of α_π, we let the mean value of the interval for α_y be

8. The mean values of the sampling intervals for the parameters in the five exogenous processes are determined such that the simulated series look similar (to the best of our ability) to the corresponding series for the post-war US economy.

Finally, determining the intervals for the five measurement errors based on empirical evidence is truly very difficult. We conjecture that the annual three month interest rate is measured quite accurately and the interval for the standard error is therefore chosen to $[0.0005, 0.0009]$. The quarterly inflation rate is taken to be measured less precisely, so we let the corresponding interval be $[0.0010, 0.0020]$. For the three real growth rates we choose quite wide intervals for the standard errors, ranging between 0.0015 to 0.0035.

As pointed out by Andreasen (2008*a*), certain requirements need to be fulfilled in DSGE models with stochastic and deterministic trends in order to ensure that the objective functions of the household and the firms are finite. Our DSGE model has the same key properties as the model in Andreasen (2008*a*), and hence the results in Andreasen (2008*a*) also apply to our DSGE model. Using proposition 1(a) in Andreasen (2008*a*), we get the following condition

$$E\left\{\frac{F_z \epsilon_{z,t}}{(1-\rho_z)}\right\} \beta \mu_{z,ss}^{F_z} \mu_{\Upsilon,ss}^{\frac{\theta}{1-\theta}F_z} < 1 \tag{85}$$

$$F_z \equiv (1-\phi_4)(1-\phi_3) \tag{86}$$

besides a boundedness condition. Andreasen (2008*a*) shows that this boundedness condition is satisfied if we assume that all variables in the economy are never too far away from the economy's growth path. Given this assumption, all our test economies have finite objective functions provided that (85) hold. We therefore impose (85) when generating our 100 test economies.

6 A Monte Carlo study

This section conducts an extensive Monte Carlo study of the following filters: i) the CDKF, ii) the standard PF, and iii) the MSPF. Following some details for the implementation of the Monte Carlo study, we start by testing the ability of the three filters to estimate the state variables in our sequence of test economies. The ability of the CDKF filter to estimate the structural parameters ($\boldsymbol{\theta}$) by QML is examined afterwards.[10]

6.1 Study design

For each of the 100 test economies, 50 data sets are simulated each with a sample length of $T = 200$. Hence, we conduct a total of 5000 repetitions of each filter. In all filters using multivariate Stirling interpolations, the value of the step size h is set equal to the optimal value for the normal distribution $\left(h = \sqrt{3}\right)$ even though the distribution may be non-normal.

[10] All calculatings are made in Fortran 90 on Dell SC1435 compute-nodes, each with 2 dualcore Opteron 2.6 GHz, 8 GB memory, and 250 GB disk. User-friendly Matlab versions of the filters tested in this Monte Carlo study are available from the author's homepage.

We deal with the problem of having values of $\mathbf{x}_{t-1}$ in the set of measurement equation (83) in the following way. In the CDKF, the state vector is defined to be $\tilde{\mathbf{x}}_t \equiv \left[\begin{array}{cc} \mathbf{x}_t' & \mathbf{x}_{t-1}' \end{array} \right]'$ with the following expanded set of transition equations

$$\left[\begin{array}{c} \mathbf{x}_{t+1} \\ \mathbf{x}_t \end{array} \right] = \left[\begin{array}{c} \mathbf{h}\left(\mathbf{x}_t, \sigma\right) \\ \mathbf{x}_t \end{array} \right] + \left[\begin{array}{c} \sigma \boldsymbol{\eta} \mathbf{w}_{t+1} \\ \mathbf{0} \end{array} \right]. \tag{87}$$

In the standard PF and the MSPF, we simply store the values of $\left\{ \mathbf{g}\left(\mathbf{x}_t^{(i)}, \boldsymbol{\theta} \right) \right\}_{i=1}^N$ and resample $\left\{ \mathbf{x}_t^{(i)}, \mathbf{g}\left(\mathbf{x}_t^{(i)}, \boldsymbol{\theta} \right) \right\}_{i=1}^N$ with probabilities $\left\{ \tilde{w}_t^{(i)} \right\}_{i=1}^N$. This procedure is computationally equivalent to defining $\tilde{\mathbf{x}}_t \equiv \left[\begin{array}{cc} \mathbf{x}_t' & \mathbf{x}_{t-1}' \end{array} \right]'$ with (87) and computing $\left\{ \mathbf{g}\left(\mathbf{x}_t^{(i)}, \boldsymbol{\theta} \right), \mathbf{g}\left(\mathbf{x}_{t-1}^{(i)}, \boldsymbol{\theta} \right) \right\}_{i=1}^N$ each time period as done in Fernández-Villaverde & Rubio-Ramírez (2007*a*). The computational advantage of resampling $\left\{ \mathbf{x}_t^{(i)}, \mathbf{g}\left(\mathbf{x}_t^{(i)}, \boldsymbol{\theta} \right) \right\}_{i=1}^N$ is that we do not need to evaluate $\left\{ \mathbf{g}\left(\mathbf{x}_{t-1}^{(i)}, \boldsymbol{\theta} \right) \right\}_{i=1}^N$ in each time period.

The estimated state values reported in the Monte Carlo study are for $\mathbf{x}_t$ at time point t. The precision of these estimates are measured by the total root mean squared errors (RMSE) given by

$$RMSE_i = \sum_{j=1}^{n_x} \sqrt{\frac{\sum_{t=1}^T \left(x_{j,t} - \hat{x}_{j,t} \right)^2}{T}}, \tag{88}$$

for the i'th run of a given filter. Note that we consider the initial state ($\mathbf{x}_0$) to be known in all experiments.[11] In the interest of space, we only report the average RMSE across the 5000 runs, i.e.

$$RMSE = \frac{\sum_{i=1}^{5000} RMSE_i}{5000}. \tag{89}$$

Both particle filters are implemented with residual resampling as proposed by Liu & Chen (1998). This resampling procedure is computationally very fast to implement and it has a lower sampling variance than simple random resampling.[12]

6.2 State estimation

We begin by considering the case where all five shocks to the 100 test economies are normally distributed. That is $\xi_i = 2$ for $i = \{\pi^*, a, e, \psi, z\}$.

< Figure 1 about here >

The graph in the upper left panel of figure 1 reports the RMSE for the three filters. We first note that the CDKF (denoted by a line with a cicle) has a very low RMSE of 0.0720, and more surprisingly, the CDKF clearly outperforms the standard PF (denoted by an unmarked line) even with a large number of particles. For instance, with 60,000 particles the RMSE is equal to 0.0810 for the standard PF. This finding indicates that a linear updating rule and

[11] We initialize the CDKF by letting $\mathbf{P}_{\mathbf{xx}}\left(t = 0\right) = 10^{-6}$.

[12] According to Merwe (2004), the specific choice of the resampling scheme does not significantly affect the performance of particle filters.

the multivariate Stirling interpolations used in the CDKF are reasonable approximations for our class of test economies. Of course, greater accuracy can be achieved in the standard PF if an even higher number of particles is used. For instance, with 120,000 particles the RMSE is reduced to 0.0776, but this is still higher than the RMSE for the CDKF.

Another surprising result is the impressive performance of the MSPF (denoted by a line with a star), and even with a very low number of particles. For instance, with 5,000 particles the MSPF has a RMSE of 0.0752. Thus, the MSPF with 5,000 particles performs better than the standard PF with 60,000 particles. However, the CDKF also marginally outperforms the MSPF which is surprising.

The upper right panel in figure 1 displays the average number of seconds it takes to evaluate each of the three filters. Here we note that the CDKF is extremely fast to compute (0.14 seconds). This makes the CDKF approximately 120 times faster to compute than the standard PF with 60,000 particles (17.09 seconds), and at the same time the CDKF delivers more precise state estimates. For the two particles filter, we see that the MSPF is only marginally slower to compute than the standard PF for the same number of particles. Thus, the MSPF with 5,000 particles (taking 1.78 seconds) is approximately 9.6 times faster to compute than the standard PF with 60,000 particles, and in addition the MSPF delivers more precise state estimates.

The quality of a proposal distribution in a particle filter can be measured by calculating the effective sample size

$$N_{eff,i,t} = \frac{1}{\sum_{j=1}^{N} \left(w_t^{(j)} \right)^2} \tag{90}$$

for the i'th run of a given particle filter at time point t. If only one particle gets a positive weight $N_{eff,i,t} = 1$, and in the optimal case of perfect adaptation, where $w_t^{(i)} = 1/N$ for all i, it holds that $N_{eff,i,t} = N$. Hence, large values of $N_{eff,i,t}$ indicate that a given proposal distribution is good, and vice versa for low values of $N_{eff,i,t}$. In the interest of space, figure 1 only reports the average effective sample size across the sample length and across all runs of a given particle filter, i.e.

$$N_{eff} = \frac{1}{T5000} \sum_{i=1}^{5000} \left(\sum_{t=1}^{T} N_{eff,i,t} \right). \tag{91}$$

The graphs in the lower left panel of figure 1 clearly show that the effective sample size for the MSPF is orders of magnitudes higher than the effective sample size for the standard PF. Hence, the new proposal distribution we suggest in (63) has a much better overlap with the posterior state distribution than the transition distribution used in the standard PF.

The lower right panel in figure 1 reports the number of times, where the filters completely lose track of the underlying state vector and hence diverge during the 5,000 runs. The CDKF never diverges, and this shows that the filter is indeed very robust as claimed in the introduction of this paper. In line with common intuition, increasing the number of particles in the standard PF significantly reduces the number of cases with divergence, from 419 cases with 5,000 particles to only 18 cases with 60,000 particles. Surprisingly, a similar sharp reduction in the number of filter divergences is not present for the MSPF, although the initial level of filter divergences is very low. The number of filter divergence for the MSPF only decreases from 66 cases with 5,000 particles to 48 cases with 60,000 particles.

This somewhat unsatisfying performance of the MSPF is probably not due to imprecise estimates of $\boldsymbol{\mu}_t$ because the CDKF estimates the state vector quite accurately as shown above.

However, momentarily too large values of $\hat{\mathbf{S}}_{\mathbf{x}}^{CDKF}(t)$ could easily generate a too diverge sample of particles and result in low values of the log-likelihood function (L_t) or even divergence of the filter. We test this hypothesis by introducing a so-called backup proposal distribution in the MSPF when the filter experiences very low values of the log-likelihood function. The specific backup distribution we use is given by (63), where the Cholesky factor of the co-variance matrix for the structural shocks hitting the economy ($\mathbf{S}_{\mathbf{w}}$) is used instead of $\hat{\mathbf{S}}_{\mathbf{x}}^{CDKF}$. That is, our backup proposal distribution is

$$\mathbf{x}_{t+1}^{(i)} = \hat{\mathbf{x}}_t^{(i)} + \boldsymbol{\mu}_{t+1} + \mathbf{S}_{\mathbf{w}}(t+1)\,\boldsymbol{\epsilon}_{t+1}^{(i)} \qquad \text{for } i = 1, ..., N. \tag{92}$$

We refer to the MSPF with this backup distribution as MSPF$_b$. An overview of the MSPF$_b$ is given in appendix E. The lower right panel in figure 1 clearly shows that this minor adjustment significantly reduces the number of cases with filter divergence to around 2 for MSPF$_b$ (denoted by a line with a diamond) regardlessly of the number of particles. Thus, these results confirm our hypothesis from above. The RMSE, the number of seconds, and effective sample size for the MSPF$_b$ are almost identical to the corresponding statistics for the MSPF. Hence, we do not report these statistics for the MSPF$_b$ in figure 1. Based on these findings, we prefer the MSPF$_b$ to the MSPF, and we therefore focus on the performance of the MSPF$_b$ in the remaining part of this Monte Carlo study.

Another metric for comparing the performance of different particle filters is to examine the size of the Monte Carlo variation in the reported values for the log-likelihood function. We can measure this variation by solely increasing the number of particles, and then calculate the change in the reported log-likelihoods values. Denoting the log-likelihood value at time point T, based on N particles, and for the i'th run of a filter by $L_T^i(N)$, the standard deviation of the Monte Carlo error in the log-likelihood can be measured by

$$std_{MC}(j) = \sqrt{\frac{1}{5000}\sum_{i=1}^{5000}\left(L_T^i(5000 \times j) - L_T^i(5000 \times (j-1))\right)^2} \tag{93}$$

for $j = 2, 3, ..., 12$. Figure 2 shows the degree of Monte Carlo variation in the standard PF and the MSPF$_b$, and as expected, this variation decreases with an increasing number of particles in both filters. More striking is the low Monte Carlo variation in the MSPF$_b$, which is orders of magnitude lower than the Monte Carlo variation in the standard PF. Therefore, also along this metric does the MSPF$_b$ outperform the standard PF.

< Figure 2 about here >

To summarize the results from our first experiment, we find that a linear updating rule and the multivariate Stirling interpolations used in the CDKF is a more accurate approximation than using particles combined with importance sampling and resampling. The next two experiments examine the robustness of this result.

The setup for the first robostness analysis is motivated by the following observation: if the observables ($\mathbf{y}_t$), the state vector ($\mathbf{x}_t$), and all shocks to the state space system are normally distributed, then a linear updating rule for the posterior state vector is optimal (Merwe & Wan (2003)). Hence, it might be the case that the good performance of the CDKF compared to the two particle filters is related to the fact that normally distributed shocks are used in

the previous experiment. We examine this possibility in our next experiment, where we let $\xi_i = 1$ for $i = \{e, a, z, \psi, \pi^*\}$ such that the generalized error distribution reduces to the Laplace distribution. Recall that the Laplace distribution has thicker tails than the normal distribution, meaning that large shocks occur more often when the Laplace distribution is used instead of the normal distribution. Therefore, our second experiment also examines whether the CDKF and the MSPF$_b$ are more robust to state outliers than the standard PF.

< Figure 3 about here >

The results from our second experiment are shown in figure 3. Again, the CDKF is seen to clearly outperform the standard PF, and the CDKF still performs marginally better than the MSPF$_b$. The number of seconds for each of the filters and the effective sample sizes are very similar to the results from the first experiment. The lower right panel in figure 3 shows the number of cases where each of the filters diverge during the 5,000 runs. Again, this number is zero for the CDKF. The number of cases with filter divergence for the standard PF is 150 with 60,000 particles, which is much higher than the 18 cases when shocks are normally distributed. Hence, we confirm that the standard PF is sensitive to state outliers. Again, we see the benefit of introducing the backup distribution in our new particle filter. For instance, with 5,000 particles the number of filter divergences is reduced from 384 cases in the MSPF to 156 cases in the MSPF$_b$. The degree of Monte Carlo variation in the standard PF and the MSPF$_b$ is similar to the level with normal shocks driving the economy, and therefore we do not report these statistics in the case with Laplace distributed shocks.

Thus, the presence of normally distributed shocks does not explain the good performance of the CDKF compared to the two particle filters, because the CDKF performs equally well with non-normal shocks driving the economy. The second robustness analysis we perform is motivated by the following observation. Fernández-Villaverde & Rubio-Ramírez (2005b) and Fernández-Villaverde & Rubio-Ramírez (2005a) use the standard PF to illustrate the effect of non-linear terms when estimating DSGE models by likelihood inference. Their findings indicate that the standard PF is sufficiently precise in the context of DSGE models. On the other hand, our results in figure 1 and figure 3 show that the approximation of the posterior state distribution in the standard PF may be quite inaccurate, because the simple CDKF clearly outperforms the standard PF. One explanation for these seemingly opposite findings could be that Fernández-Villaverde and Rubio-Ramínez (2005a, 2005b) use the standard PF on a relatively small DSGE model (the neoclassical growth model), which only has one shock and one endogenous state variable. On the other hand, the DSGE model used in this paper is much bigger and has five shocks and six endogenous state variables, and this difference might explain the opposite findings. We test this hypothesis in our third experiment where we gradually increase the number of shocks hitting the economy. That is, we once again use our 100 test economies and use normally distributed shocks to the economy, but now we impose the additional constraints that $\sqrt{Var\left(\epsilon_{i,t}\right)} = 0$; first for $i = \{a, e, \psi, z\}$, then for $i = \{e, \psi, z\}$, $i = \{\psi, z\}$, $i = \{z\}$, and finally for $i = \varnothing$. Thus, this experiment allow us to examine the effect of gradually increasing the number of shocks while keeping the number of endogenous state variables constant.

< Figure 4 about here >

The results in figure 4 confirm our hypothesis. The standard PF clearly outperforms the CDKF with one and two shocks. That is, with few shocks to the economy we recover the standard result that the particle approximation of the state distribution is more accurate than using a linear updating rule and multivariate Stirling interpolations. However, adding the third shock makes the CDKF perform better than the standard PF for the considered interval of particles. With four and five shocks, the gain in performance by using the CDKF compared to the standard PF increases even further. We also see that the MSPF$_b$ constantly outperforms the standard PF, and when the CDKF is better than the standard PF, the performance of the MSPF$_b$ is very similar to the CDKF. In other words, the MSPF$_b$ is either the best filter or very close to the best filter.

Based on these findings we conclude that the standard PF suffers from a curse of dimensionality with respect to the number of shocks hitting the economy. That is, the performance of the standard PF deteriorates rappidly with more than two shocks for the considered interval of particles. On the other hand, the MSPF$_b$ inherits the good performance of the CDKF and the precision of the MSPF$_b$ deteriorates more slowly than the standard PF when the number of shocks is increased.

< Figure 5 about here >

Figure 5 displays the degree of Monte Carlo variation in the log-likelihood function as measured by (93), when we gradually increase the number of shocks. Except for the case with one shock, the MSPF$_b$ is seen to have a much lower amount of Monte Carlo variation in the log-likelihood function than the log-likelihood function based on the standard PF. A further inspection of the graphs for the MSPF$_b$ in figure 5 reveals that the Monte Carlo variation with two shocks surprisingly is lower than the Monte Carlo variation with only one shock. Recall that the case with one shock is the scenario where only shocks to the inflation rate target are present and these shocks have a small variance. Hence, one explanation for the relatively high Monte Carlo variation in the MSPF$_b$ with one shock could be due to small values of $\hat{\mathbf{S}}_{\mathbf{x}}^{CDKF}$, implying smaller tails in the proposal distribution than in the posterior state distribution. If this is the explanation, then increasing $\hat{\mathbf{S}}_{\mathbf{x}}^{CDKF}$ by a given scalar should reduce the Monte Carlo variation. Further simulation studies verify this explanation. For instance, scaling $\hat{\mathbf{S}}_{\mathbf{x}}^{CDKF}$ by 1.5 reduces the Monte Carlo variation in the log-likelihood function from the MSPF$_b$, although this variation is still higher than the Monte Carlo variation in the log-likelihood function from the standard PF.

To summarize, we highlight the three main results from the simulation study. First, with one or two shocks in the non-linear DSGE model, both particles filters outperform the CDKF in terms of RMSE for the state vector. Second, with three or more shocks, the CDKF clearly outperforms the standard PF and marginally the MSPF$_b$. Moreover, the CDKF is at the same time much faster to compute than both particle filters. Third, the MSPF$_b$ outperforms the standard PF by delivering more precise state estimates, and in general the MSPF$_b$ has lower Monte Carlo variation in the reported log-likelihood function. This is the case even if a low number of particles is used in the MSPF$_b$ and a large number of particles is used in the standard PF. As a result, the MSPF$_b$ is faster to compute than the standard PF.

6.3 Parameter estimation

This section examines the finite sample properties of the QML estimator for the structural parameters in our DSGE model. To make this study feasible, we only consider five of the 34 parameters in our DSGE model as unknowns. The five unknown parameters are: i) the preference parameter (ϕ_4), ii) the degree of price stickiness (α), iii) the central banks reaction to deviations from the inflation rate target (α_π), iv) the degree of persistency in stationary technology shocks (ρ_a), and v) the standard deviation for non-stationary technology shocks $\left(\sqrt{Var\left(\varepsilon_{z,t}\right)}\right)$. All the parameters are given their benchmark value as stated in table 1 for the Monte Carlo study, implying that five shocks are hitting the economy.

The objective function is optimized by the CMA-ES routine adapted to DSGE models by Andreasen (2008*b*). Andreasen (2008*b*) shows that the CMA-ES routine is able to optimize likelihood functions for DSGE models.

We begin by examining the properties of the QML estimator in the case where five normally distributed shocks drive the economy.

< Figure 6 about here >

< Table 2 about here >

Figure 6 reports the estimated finite sample distributions for the five parameters based on 990 repetitions in the Monte Carlo study.[13] All five distributions are clearly well approximated by a normal distribution. Table 2 shows that the biases in all five parameters are very small. The $\mathbf{A}$ and $\mathbf{B}$ matrices in the expression for the standard errors to the QML estimator are estimated based on numerical derivatives and the corresponding sample moments. We then note that standard errors in table 2 calculated from $\hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{A}}^{-1}$ have small positive biases, and the actual Type I errors at a 5% significance level are therefore often higher than 5%. Out of curiosity, we also report in table 2 standard errors calculated from $\hat{\mathbf{B}}^{-1}$, i.e. in the case where the information equality is assumed to hold. Surprisingly, this gives more precise estimates of the standard errors and marginally better Type I errors at a 5% significance level. Hence, with normal shocks to the economy, the quasi log-likelihood function approximates the true log-likelihood function well, and the induce errors from imposing the information equality is seen to be smaller than the estimation errors in the Hessian matrix.

To examine whether the QML estimator is robust to non-normal shocks driving the economy, we repeat the Monte Carlo study from above but with five Laplace distributed shocks.

< Figure 7 about here >

< Table 3 about here >

Again, the biases in all five parameter estimates are negligible and the finite sample distributions for the five parameters are well approximated by a normal distribution (see figure 7). The true standard errors for each of the five parameters with Laplace distributed shocks are

[13] We have deleted 10 Monte Carlo repetitions because we were unable to obtain reliable estimates of the standard errors.

seen to be slightly larger than when normal shocks drive the economy, in particular for α_π and $\sqrt{Var\left(\epsilon_{z,t}\right)}$. Estimating these standard errors based on $\hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{A}}^{-1}$ implies in general positive biases, which are rather large for ϕ_4 and α. As expected, all standard errors calculated from $\hat{\mathbf{B}}^{-1}$ have negative biases. A further inspection of the biases in these standard errors reveal that the estimates from $\hat{\mathbf{B}}^{-1}$ once again are more precise than the standard errors $\hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{A}}^{-1}$ as measured by the RMSE for all five estimates.

To summarize, this simulation study shows that the finite sample distributions for the QML estimator are well approximated by the asymptotic normal distributions. Standard errors for the QML estimator can be estimated by well-known techniques due to the smooth nature of the quasi log-likelihood function. The latter result should be seen in contrast to ML estimation by a particle filter, where the missing smoothness of the likelihood function makes it difficult to calculate standard errors (see for instance Fernández-Villaverde & Rubio-Ramírez (2007*a*) and Fernández-Villaverde & Rubio-Ramirez (2007*b*)).

7 Conclusion

The contribution of this paper is twofold. First, we show how structural parameters in non-linear DSGE models can be estimated by QML based on the CDKF. The advantage of this estimator is mainly that the quasi log-likelhood function can be evaluated in a fraction of a second, and standard errors can be estimated by well-known techniques due to the smooth nature of the quasi log-likelihood function.

Our second contribution is to introduce the MSPF$_b$, where the CDKF is used in an efficient way to get a good proposal distribution in the importance sampling step of particle filters. We show that the MSPF$_b$ outperforms the standard PF by delivering more precise state estimates, and in general the MSPF$_b$ has lower Monte Carlo variation in the reported log-likelihood function. This is the case even if a low number of particles is used in the MSPF$_b$ and a large number of particles is used in the standard PF. As a result, the MSPF$_b$ is faster to compute than the standard PF.

Our new mean shifted proposal distribution based on the CDKF can also be used in various extensions of the general framework 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 estimated samples from the posterior state distribution. See the book by Doucet, de Freitas & Gordon (2001*b*) for other interesting extensions of the general framework for particle filters. Testing the performance of our new proposal distribution in these cases is left for future research.

A The algorithm for the Central Difference Kalman Filter

- Initialization: $t = 0$

  Set $\hat{\mathbf{x}}_t$ and $\hat{\mathbf{S}}_{\mathbf{x}}(t)$.

- For $t > 1$

  Prediction step:

  - $\bar{\mathbf{x}}_{t+1} = \frac{h^2 - n_x - n_w}{h^2} \mathbf{h}\left(\hat{\mathbf{x}}_t, \bar{\mathbf{w}}_{t+1}; \boldsymbol{\theta}\right)$
    $+ \frac{1}{2h^2} \sum_{p=1}^{n_x} \left(\mathbf{h}\left(\hat{\mathbf{x}}_t + h\hat{\mathbf{s}}_{\mathbf{x},p}, \bar{\mathbf{w}}_{t+1}; \boldsymbol{\theta}\right) + \mathbf{h}\left(\hat{\mathbf{x}}_t - h\hat{\mathbf{s}}_{\mathbf{x},p}, \bar{\mathbf{w}}_{t+1}; \boldsymbol{\theta}\right)\right)$
    $+ \frac{1}{2h^2} \sum_{p=1}^{n_w} \left(\mathbf{h}\left(\hat{\mathbf{x}}_t, \bar{\mathbf{w}}_{t+1} + h\mathbf{s}_{\mathbf{w},p}; \boldsymbol{\theta}\right) + \mathbf{h}\left(\hat{\mathbf{x}}_t, \bar{\mathbf{w}}_{t+1} - h\mathbf{s}_{\mathbf{w},p}; \boldsymbol{\theta}\right)\right)$
  - $\bar{\mathbf{S}}_{\mathbf{x}}(t+1) = \Phi\left(\left[\begin{array}{cccc} \mathbf{S}^{(1)}_{\mathbf{xx}}(t) & \mathbf{S}^{(1)}_{\mathbf{xw}}(t) & \mathbf{S}^{(2)}_{\mathbf{xx}}(t) & \mathbf{S}^{(2)}_{\mathbf{xw}}(t) \end{array}\right]\right)$

  Updating step:

  - $\bar{\mathbf{y}}_{t+1} = \frac{h^2 - n_x - n_v}{h^2} \mathbf{g}\left(\bar{\mathbf{x}}_{t+1}, \bar{\mathbf{v}}_{t+1}; \boldsymbol{\theta}\right)$
    $+ \frac{1}{2h^2} \sum_{p=1}^{n_x} \left(\mathbf{g}\left(\bar{\mathbf{x}}_{t+1} + h\bar{\mathbf{s}}_{\mathbf{x},p}, \bar{\mathbf{v}}_{t+1}; \boldsymbol{\theta}\right) + \mathbf{g}\left(\bar{\mathbf{x}}_{t+1} - h\bar{\mathbf{s}}_{\mathbf{x},p}, \bar{\mathbf{v}}_{t+1}; \boldsymbol{\theta}\right)\right)$
    $+ \frac{1}{2h^2} \sum_{p=1}^{n_v} \left(\mathbf{g}\left(\bar{\mathbf{x}}_{t+1}, \bar{\mathbf{v}}_{t+1} + h\mathbf{s}_{\mathbf{v},p}; \boldsymbol{\theta}\right) + \mathbf{g}\left(\bar{\mathbf{x}}_{t+1}, \bar{\mathbf{v}}_{t+1} - h\mathbf{s}_{\mathbf{v},p}; \boldsymbol{\theta}\right)\right)$
  - $\bar{\mathbf{S}}_{\mathbf{y}}(t+1) = \boldsymbol{\Phi}\left(\left[\begin{array}{cccc} \mathbf{S}^{(1)}_{\mathbf{yx}}(t+1) & \mathbf{S}^{(1)}_{\mathbf{yv}}(t+1) & \mathbf{S}^{(2)}_{\mathbf{yx}}(t+1) & \mathbf{S}^{(2)}_{\mathbf{yv}}(t+1) \end{array}\right]\right)$
  - $\mathbf{K}_{t+1} = \bar{\mathbf{S}}_{\mathbf{x}}(t+1)\, \mathbf{S}^{(1)}_{\mathbf{yx}}(t+1)' \left[\bar{\mathbf{S}}_{\mathbf{y}}(t+1)\, \bar{\mathbf{S}}_{\mathbf{y}}(t+1)'\right]^{-1}$
  - $\hat{\mathbf{S}}_{\mathbf{x}}(t+1) = \Phi([\ \bar{\mathbf{S}}_{\mathbf{x}}(t+1) - \mathbf{K}_{t+1}\mathbf{S}^{(1)}_{\mathbf{yx}}(t+1) \quad \mathbf{K}_{t+1}\mathbf{S}^{(1)}_{\mathbf{yv}}(t+1)$
    $\mathbf{K}_{t+1}\mathbf{S}^{(2)}_{\mathbf{yx}}(t+1) \quad \mathbf{K}_{t+1}\mathbf{S}^{(2)}_{\mathbf{yv}}(t+1)\ ])$

  Quasi log-likelihood function

- $L_{t+1} = L_t - \frac{n_y}{2}\log(2\pi) - \frac{1}{2}\log\left(\left|\bar{\mathbf{P}}_{\mathbf{yy}}(t+1)\right|\right)$
  $-\frac{1}{2}\left(\mathbf{y}_{t+1} - \bar{\mathbf{y}}_{t+1}\right)' \bar{\mathbf{P}}^{-1}_{\mathbf{yy}}(t+1)\left(\mathbf{y}_{t+1} - \bar{\mathbf{y}}_{t+1}\right)$

B The smoother for the CDKF

The smoothed estimate of $\mathbf{x}_t$ is denoted $\mathbf{x}^s_t$. The covariance matrix of this estimate is denoted $\mathbf{P}^s_{\mathbf{xx}}(t) \equiv E_T\left[\left(\mathbf{x}_t - \mathbf{x}^s_t\right)\left(\mathbf{x}_t - \mathbf{x}^s_t\right)'\right]$. It holds that

$$\mathbf{x}^s_t = \hat{\mathbf{x}}_t + \mathbf{K}_{t+1}\left[\mathbf{x}^s_{t+1} - \bar{\mathbf{x}}_{t+1}\right]$$

$$\mathbf{P}^s_{\mathbf{xx}}(t) = \hat{\mathbf{P}}_{\mathbf{xx}}(t) + \mathbf{K}_{t+1}\left(\mathbf{P}^s_{\mathbf{xx}}(t+1) - \bar{\mathbf{P}}_{\mathbf{xx}}(t+1)\right)\mathbf{K}'_{t+1}$$

where

$$\mathbf{C}_{t+1} = E_t\left[\left(\mathbf{x}_t - \hat{\mathbf{x}}_t\right)\left(\mathbf{x}_{t+1} - \bar{\mathbf{x}}_{t+1}\right)'\right],$$

and the smoothing gain is given by

$$\mathbf{K}_{t+1} = \mathbf{C}_{t+1}\bar{\mathbf{P}}_{\mathbf{xx}}(t+1)^{-1}.$$

Based on the results in Norgaard et al. (2000), we have

$$\mathbf{C}_{t+1} = \hat{\mathbf{S}}_{\mathbf{x}}(t)\,\mathbf{S}_{\mathbf{xx}}^{(1)}(t+1)'.$$

The square root of the covariance matrix for the smoothed state estimate is given by

$$\begin{aligned}\mathbf{S}_{\mathbf{x}}^{s}(t) \;=\; \Phi([\;&\hat{\mathbf{S}}_{\mathbf{x}}(t) - \mathbf{K}_{t+1}\mathbf{S}_{\mathbf{xx}}^{(1)}(t) \quad \mathbf{K}_{t+1}\mathbf{S}_{\mathbf{x}}^{s}(t+1) \quad \mathbf{K}_{t+1}\mathbf{S}_{\mathbf{xw}}^{(1)}(t)\\ &\mathbf{K}_{t+1}\mathbf{S}_{\mathbf{xx}}^{(2)}(t) \quad \mathbf{K}_{t+1}\mathbf{S}_{\mathbf{xw}}^{(2)}(t)\;]).\end{aligned}$$

This smoothing recursion is started at the last time step, because it holds that $\mathbf{x}_T^s = \hat{\mathbf{x}}_T$ and $\mathbf{P}_{\mathbf{xx}}^{s}(T) = \hat{\mathbf{P}}_{\mathbf{xx}}(T)$. Thus, the proceedure is first to calculate the posterior estimates $\hat{\mathbf{x}}_t$ and $\hat{\mathbf{S}}_{\mathbf{x}}(t)$ for $t = 1, 2, ..., T$ by running the CDKF. Then, the smoothing recursion is started in time period T and iterated back in time.

C A transformed state space system

This section considers the case where some of the structural shocks are used to identify the model. That is, some structural shocks are used in a similar manner as $\mathbf{v}_t$ during the filtering. We follow Fernández-Villaverde & Rubio-Ramírez (2007*a*) and partition the exogenous state variables ($\mathbf{x}_{2,t}$) as

$$\mathbf{x}_{2,t} \equiv \begin{bmatrix} \mathbf{x}_{21,t} \\ \mathbf{x}_{22,t} \end{bmatrix} = \begin{bmatrix} \mathbf{h}_{21}(\mathbf{x}_{2,t-1}, \mathbf{w}_{1,t}; \boldsymbol{\theta}) \\ \mathbf{h}_{22}(\mathbf{x}_{2,t-1}, \mathbf{w}_{2,t}; \boldsymbol{\theta}) \end{bmatrix}.$$

The shocks $\mathbf{w}_{2,t}$ are used to identify the model. Recall that $\mathbf{x}_t \equiv \begin{bmatrix} \mathbf{x}_{1,t} \\ \mathbf{x}_{21,t} \\ \mathbf{x}_{22,t} \end{bmatrix}$. Substituting $\mathbf{x}_{22,t}$ into the measurement equations we get

$$\mathbf{y}_t = \mathbf{g}\left(\begin{bmatrix} \mathbf{x}_{1,t} \\ \mathbf{x}_{21,t} \\ \mathbf{h}_{22}(\mathbf{x}_{2,t-1}, \mathbf{w}_{2,t}; \boldsymbol{\theta}) \end{bmatrix}, \mathbf{v}_t; \boldsymbol{\theta}\right)$$

Thus, the transformed state space system reads

$$\mathbf{y}_t = \tilde{\mathbf{g}}\left(\mathbf{x}_{2,t-1}, \begin{bmatrix} \mathbf{x}_{1,t} \\ \mathbf{x}_{21,t} \end{bmatrix}, \begin{bmatrix} \mathbf{v}_t \\ \mathbf{w}_{2,t} \end{bmatrix}; \boldsymbol{\theta}\right)$$

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

$$\mathbf{x}_{21,t} = \mathbf{h}_{21}(\mathbf{x}_{2,t-1}, \mathbf{w}_{1,t}; \boldsymbol{\theta})$$

where $\tilde{\mathbf{g}}(\cdot)$ is a new function. The CDKF can then be used on this new non-linear system to get the state variables.

D The algorithm for the standard Particle Filter

- 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.

- For $t > 1$

  1. Importance sampling step
     - For $i = 1, ..., N$ draw particles $\mathbf{x}_t^{(i)}$ from $p\left(\mathbf{x}_t \left| \hat{\mathbf{x}}_{t-1}^{(i)} ; \boldsymbol{\theta} \right.\right)$
     - For $i = 1, ..., N$ evaluate the importance weights: $w_t^{(i)} = w_{t-1}^{(i)} p\left(\mathbf{y}_t \left| \mathbf{x}_t^{(i)} ; \boldsymbol{\theta} \right.\right)$
     - The contribution to the log-likelihood function: $L_t = L_{t-1} + \log(\sum_{i=1}^N w_t^{(i)})$
     - For $i = 1, ..., N$ normalize the importance weights $\tilde{w}_t^{(i)} = w_t^{(i)} / \sum_{i=1}^N w_t^{(i)}$
  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: $\hat{\mathbf{x}}_t = \frac{1}{N} \sum_{i=1}^N \hat{\mathbf{x}}_t^{(i)}$

E The algorithm for the Mean Shifted Particle Filter

We use the notation $\left\{\hat{\mathbf{x}}_{0:t}^{(i)}\right\}_{i=1}^{N} = PF\left(\left\{\hat{\mathbf{x}}_{0:t-1}^{(i)}\right\}_{i=1}^{N}, \mathbf{y}_t\right)$ to denote one iteration in the standard PF from time point $t-1$ to time point t based on $\left\{\hat{\mathbf{x}}_{0:t-1}^{(i)}\right\}_{i=1}^{N}$ and $\mathbf{y}_t$. Similarly, we use the notation $\left[\mathbf{x}_t^{CDKF}, \hat{\mathbf{S}}_{\mathbf{x}}(t)\right] = CDKF\left(\hat{\mathbf{S}}_{\mathbf{x}}(t-1), \hat{\mathbf{x}}_{t-1}, \mathbf{y}_t\right)$ to denote one iteration in the CDKF from time point $t-1$ to time point t based on $\hat{\mathbf{S}}_{\mathbf{x}}(t-1)$, $\hat{\mathbf{x}}_{t-1}$, and $\mathbf{y}_t$.

The Mean Shifted Particle Filter (MSPF)

- 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 $t = 1$

  $\left\{\hat{\mathbf{x}}_{0:t}^{(i)}\right\}_{i=1}^{N} = PF\left(\left\{\hat{\mathbf{x}}_{0:t-1}^{(i)}\right\}_{i=1}^{N}, \mathbf{y}_t\right)$

  The posterior state estimate: $\hat{\mathbf{x}}_t = \frac{1}{N}\sum_{i=1}^{N}\hat{\mathbf{x}}_t^{(i)}$

  The posterior estimate of $\mathbf{S}_{\mathbf{x}}(t)$ is: $\hat{\mathbf{S}}_{\mathbf{x}}(t) = \Phi\left(\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\mathbf{x}}_t^{(i)} - \hat{\mathbf{x}}_t\right)\left(\hat{\mathbf{x}}_t^{(i)} - \hat{\mathbf{x}}_t\right)'\right)$

- For $t > 1$

  1. Importance sampling step
     - $\left[\mathbf{x}_t^{CDKF}, \hat{\mathbf{S}}_{\mathbf{x}}(t)\right] = CDKF\left(\hat{\mathbf{S}}_{\mathbf{x}}(t-1), \hat{\mathbf{x}}_{t-1}, \mathbf{y}_t\right)$
     - Let $\boldsymbol{\mu}_t = \hat{\mathbf{x}}_t^{CDKF} - \hat{\mathbf{x}}_{t-1}$ be a mean correction term
     - For $i = 1, ..., N$ draw particles $\mathbf{x}_t^{(i)}$ from $\mathcal{N}\left(\mathbf{x}_t^{(i)} \middle| \mathbf{x}_{t-1}^{(i)}, \boldsymbol{\mu}_t, \hat{\mathbf{S}}_{\mathbf{x}}(t)\right)$
     - For $i = 1, ..., N$ evaluate the importance weights: $w_t^{(i)} = w_{t-1}^{(i)} \frac{p\left(\mathbf{y}_t \middle| \mathbf{x}_t^{(i)}; \boldsymbol{\theta}\right) p\left(\mathbf{x}_t^{(i)} \middle| \hat{\mathbf{x}}_{t-1}^{(i)}; \boldsymbol{\theta}\right)}{\mathcal{N}\left(\mathbf{x}_t^{(i)} \middle| \mathbf{x}_{t-1}^{(i)}, \boldsymbol{\mu}_t, \hat{\mathbf{S}}_{\mathbf{x}}(t)\right)}$
     - The contribution to the log-likelihood function: $L_t = L_{t-1} + \log(\sum_{i=1}^{N} w_t^{(i)})$
     - For $i = 1, ..., N$ normalize the importance weights $\tilde{w}_t^{(i)} = w_t^{(i)} / \sum_{i=1}^{N} w_t^{(i)}$
  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: $\hat{\mathbf{x}}_t = \frac{1}{N}\sum_{i=1}^{N}\hat{\mathbf{x}}_t^{(i)}$

The Mean Shifted Particle Filter with backup distribution (MSPF$_b$)

- 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 $t = 1$

  $\left\{\hat{\mathbf{x}}_{0:t}^{(i)}\right\}_{i=1}^{N} = PF\left(\left\{\hat{\mathbf{x}}_{0:t-1}^{(i)}\right\}_{i=1}^{N}, \mathbf{y}_t\right)$

  The posterior state estimate: $\hat{\mathbf{x}}_t = \frac{1}{N}\sum_{i=1}^{N} \hat{\mathbf{x}}_t^{(i)}$

  The posterior estimate of $\mathbf{S}_{\mathbf{x}}(t)$ is: $\hat{\mathbf{S}}_{\mathbf{x}}(t) = \Phi\left(\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\mathbf{x}}_t^{(i)} - \hat{\mathbf{x}}_t\right)\left(\hat{\mathbf{x}}_t^{(i)} - \hat{\mathbf{x}}_t\right)'\right)$

- For $t > 1$

  1. Importance sampling step
     - $\mathbf{x}_t^{CDKF} = CDKF\left(\mathbf{S}_{\mathbf{x},t-1}\hat{\mathbf{x}}_{t-1}, \mathbf{y}_t\right)$
     - Let $\boldsymbol{\mu}_t = \hat{\mathbf{x}}_t^{CDKF} - \hat{\mathbf{x}}_{t-1}$ be a mean correction term
     - For $i = 1, ..., N$ draw particles $\mathbf{x}_t^{(i)}$ from $\mathcal{N}\left(\mathbf{x}_t^{(i)} \middle| \mathbf{x}_{t-1}^{(i)}, \boldsymbol{\mu}_t, \hat{\mathbf{S}}_{\mathbf{x}}(t)\right)$
     - For $i = 1, ..., N$ evaluate the importance weights:
       $w_t^{(i)} = w_{t-1}^{(i)} \frac{p\left(\mathbf{y}_t \middle| \mathbf{x}_t^{(i)}\right) p\left(\mathbf{x}_t^{(i)} \middle| \hat{\mathbf{x}}_{t-1}^{(i)}; \boldsymbol{\theta}\right)}{\mathcal{N}\left(\mathbf{x}_t^{(i)} \middle| \mathbf{x}_{t-1}^{(i)}, \boldsymbol{\mu}_t, \hat{\mathbf{S}}_{\mathbf{x}}(t)\right)}$
     - The contribution to the log-likelihood function: $L_t = L_{t-1} + \log(\sum_{i=1}^{N} w_t^{(i)})$
     - The backup proposal distribution:
       if $\log(\sum_{i=1}^{N} w_t^{(i)}) < 0.2\frac{L_t}{t}$
       For $i = 1, ..., N$ draw particles $\mathbf{x}_t^{(i)}$ from $\mathcal{N}\left(\mathbf{x}_t^{(i)} \middle| \mathbf{x}_{t-1}^{(i)}, \boldsymbol{\mu}_t, \mathbf{S}_{\mathbf{w}}(t)\right)$
       For $i = 1, ..., N$ evaluate the importance weights:
       $w_t^{(i)} = w_{t-1}^{(i)} \frac{p\left(\mathbf{y}_t \middle| \mathbf{x}_t^{(i)}\right) p\left(\mathbf{x}_t^{(i)} \middle| \hat{\mathbf{x}}_{t-1}^{(i)}; \boldsymbol{\theta}\right)}{\mathcal{N}\left(\mathbf{x}_t^{(i)} \middle| \mathbf{x}_{t-1}^{(i)}, \boldsymbol{\mu}_t, \mathbf{S}_{\mathbf{w}}(t)\right)}$
       end if
     - For $i = 1, ..., N$ normalize the importance weights $\tilde{w}_t^{(i)} = w_t^{(i)} / \sum_{i=1}^{N} w_t^{(i)}$
  2. Resampling step:
     - Resample with replacement from $\left\{\mathbf{x}_t^{(i)}\right\}_{i=1}^{N}$ 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 $\left\{\hat{\mathbf{x}}_{0:t}^{(i)}\right\}_{i=1}^{N}$ we have $w_t^{(i)} = \frac{1}{N}$ for all $i = 1, ..., N$
  3. State estimates
     - The posterior state estimate: $\hat{\mathbf{x}}_t = \frac{1}{N}\sum_{i=1}^{N} \hat{\mathbf{x}}_t^{(i)}$

Table 1: Parameters for the test economies

The fourth column named "interval" denotes the intervals we sample uniformly from when generating the test economies.

Label	Parameters	Benchmark	Interval
Discount factor	β	0.9992	$[0.9986, 0.9998]$
Habit degree	b	0	$[0, 0.2]$
Transaction cost	ϕ_1	0.05	$[0.01, 0.09]$
Transaction cost	ϕ_2	0.13	$[0.03, 0.23]$
Preference	ϕ_3	$\lim_{i \to 1} i$	$[0.8, 1.2]$
Preference	ϕ_4	0.5	$[0.25, 0.75]$
Adj costs for investments	κ	2.8	$[0.8, 4.8]$
Depreciation rate	δ	0.025	$[0.01, 0.04]$
Cobb-Douglas parameter	θ	0.36	$[0.30, 0.42]$
CIA	ν	0.6	$[0.4, 0.8]$
Price elasticity	η	6	$[4, 8]$
Degree of price stickiness	α	0.7	$[0.5, 0.9]$
Process for inflation target	ω^*	0.4	$[0.1, 0.7]$
Reaction to lagged interest rate	α_R	0.8	$[0.7, 0.9]$
Reaction to inflation	α_π	12	$[8, 16]$
Reaction to output	α_y	8	$[6, 10]$
Inflation rate in steady state	π_{ss}	1.008	$[1.005, 1.010]$
Growth rate in technology shocks	μ_z	1.0021	$[1.0016, 1.0026]$
Growth rate in investment shocks	μ_Υ	1.0042	$[1.0037, 1.0047]$
Persistency in stationary technology shocks	ρ_a	0.7	$[0.5, 0.9]$
Persistency in investment shocks	ρ_e	0.8	$[0.6, 1.0[$
Persistency in shocks to firms' fixed costs	ρ_ψ	0.9	$[0.8, 1.0[$
Persistency in nonstationary technology shocks	ρ_z	0.2	$[0.0, 0.4]$
std. of shocks to inflation target	$\sqrt{Var\left(\epsilon_{\pi^*,t}\right)}$	0.5	$[0.25, 0.75]$
std. of stationary technology shocks	$\sqrt{Var\left(\epsilon_{a,t}\right)}$	2	$[1.5, 2.5]$
std. of stationary investment shocks	$\sqrt{Var\left(\epsilon_{e,t}\right)}$	4	$[3, 5]$
std. of shocks to firms' fixed costs	$\sqrt{Var\left(\epsilon_{\psi,t}\right)}$	4	$[3, 5]$
std. of nonstationary technology shocks	$\sqrt{Var\left(\epsilon_{z,t}\right)}$	0.5	$[0.25, 0.75]$
Perturbation parameter	σ	0.003	$[0.002, 0.004]$
std. of errors in the interest rate	$\sqrt{Var\left(v_{R,t}\right)}$	0.0007	$[0.0005, 0.0009]$
std. of errors in inflation	$\sqrt{Var\left(v_{\pi,t}\right)}$	0.0015	$[0.0010, 0.0020]$
std. of errors in the growth rate for consumption	$\sqrt{Var\left(v_{c,t}\right)}$	0.0025	$[0.0015, 0.0035]$
std. of errors in the growth rate for investments	$\sqrt{Var\left(v_{i,t}\right)}$	0.0025	$[0.0015, 0.0035]$
std. of errors in growth rate for GDP	$\sqrt{Var\left(v_{y,t}\right)}$	0.0025	$[0.0015, 0.0035]$

Table 2: The QML estimator and 5 normally distributed shocks

The results are based on 990 repetitions in the Monte Carlo study. We have deleted 10 Monte Carlo repetitions because we were unable to obtain reliable estimates of the standard errors. The type I error is calculated at a 5 percentage significance level. Derivatives are approximated based on $\frac{\partial f}{\partial x} = \frac{f(x+h)-f(x-h)}{2h}$. For $\mathbf{A}$ we let h = 0.001 and for $\mathbf{B}$ we let h = 0.00001.

	True values			using $\hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{A}}^{-1}$		using $\hat{\mathbf{B}}^{-1}$	
	Level	SE	Bias in level	Bias in SE	Type I	Bias in SE	Type I
ϕ_4	0.5	0.104	0.000	0.051	0.095	-0.006	0.092
α	0.7	0.035	-0.002	0.014	0.119	-0.001	0.117
α_π	12	0.601	0.057	0.078	0.049	0.014	0.048
ρ_a	0.7	0.051	-0.012	0.004	0.076	-0.001	0.068
$\sqrt{Var\left(\epsilon_{z,t}\right)}$	0.7	0.069	-0.011	0.000	0.033	0.003	0.031

Table 3: The QML estimator and five Laplace distributed shocks

The results are based on 989 repetitions in the Monte Carlo study. We have deleted 11 Monte Carlo repetitions because we were unable to obtain reliable estimates of the standard errors. The type I error is calculated at a 5 percentage significance level. Derivatives are approximated based on $\frac{\partial f}{\partial x} = \frac{f(x+h)-f(x-h)}{2h}$. For $\mathbf{A}$ we let h = 0.001 and for $\mathbf{B}$ we let h = 0.00001.

	True values			using $\hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{A}}^{-1}$		using $\hat{\mathbf{B}}^{-1}$	
	Level	SE	Bias in level	Bias in SE	Type I	Bias in SE	Type I
ϕ_4	0.5	0.107	-0.003	0.128	0.097	-0.011	0.125
α	0.7	0.039	-0.002	0.029	0.130	-0.006	0.164
α_π	12	0.721	0.037	0.052	0.071	-0.150	0.122
ρ_a	0.7	0.053	-0.014	0.006	0.081	-0.004	0.081
$\sqrt{Var\left(\epsilon_{z,t}\right)}$	0.7	0.077	-0.013	-0.003	0.051	-0.009	0.081

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.

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 Review* .

Andreasen, M. M. (2008*a*), 'Ensuring the validity of the micro foundation in DSGE models', *Working Paper* .

Andreasen, M. M. (2008*b*), 'How to maximize the likelihood function for a DSGE model', *Working Paper* .

Bekaert, G., Cho, S. & Moreno, A. (2005), 'New-keynesian macroeconomics and the term structure', *NBER Working Paper* **Nr. 11340**.

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.

Bollerslev, T. & Wooldrigde, J. M. (1992), 'Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances', *Econometrics Reviews* **11**(2), 143–172.

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*a*), 'An introduction to sequential monte carlo methods', *Sequential Monte Carlo Methods in Practice, Springer, editors: Arnaud Doucet, Nando de Freitas and Neil Gordon* .

Doucet, A., de Freitas, N. & Gordon, N. (2001*b*), '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.

Duffie, D. & Singleton, K. J. (1993), 'Simulated moments estimation of markov models of asset prices', *Econometrica* **61**(4), 929–952.

Dunik, J. & Simandl, M. (2006), 'Design of square-root derivative-free smoothers', *Working Paper* .

Fernández-Villaverde, J. & Rubio-Ramírez, J. F. (2005*a*), 'Convergence properties of the likelihood of computed dynamic models', *NBER working paper nr. 315* .

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

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

Fernández-Villaverde, J. & Rubio-Ramirez, J. F. (2007*b*), 'How structural are structural parameters?', *Working Paper* .

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.

Gordon, N. J., Salmond, D. J. & Smith, A. F. M. (1993), 'Novel approach fo nonlinear/non-gaussian bayesian state estimation', *IEE Proceedings-F* **140**(2), 107–113.

Hamilton, J. D. (1994), 'Time series analysis', *Princeton University Press, Princeton New Jersey* .

Hansen, L. P. (1982), 'Large sample properties of generalized method of moments estimators', *Econometrica* **50**(4), 1029–1054.

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

Julier, S. J., Uhlmann, J. K. & Durrant-Whyte, H. F. (1995), 'A new approach for filtering nonlinear systems', *In The Proceedings of the American Control Conference* pp. 1628–1632.

Justiniano, A. & Primiceri, G. E. (2008), 'The time varying volatility of macroeconomic fluctuations', *The American Economic Review* pp. 604–641.

Kim, J., Kim, S., Schaumburg, E. & Sims, C. A. (2003), 'Calculating and using second order accurate solutions of discrete time dynamic quilibrium models', *Working Paper* .

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

Lewis, F. L. (1986), 'Optimal estimation - with an introduction to stochastic control theory', *John Wiley and Sons* .

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. (2004), 'Sigma-point kalman filters for probabilistic inference in dynamic state-space models', *Thesis* .

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.

Särkkä, S. (2008), 'Unscented rauch-tung-striebel smoother', *IEEE Transactions on Automatic Control* .

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.

Schmitt-Grohé, S. & Uribe, M. (2006), 'Optimal inflation stabilization in a medium-scale macroeconomic model', *Working Paper* pp. 1–59.

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. (2003), 'Shocks and frictions in US business cycles: A bayesian DSGE approach', *Working Paper* pp. 1–27.

Smith, J. A. A. (1993), 'Estimating nonlinear time-series models using simulated vector autoregressions', *Journal of Applied Econometrics* **8, Supplement: Special Issue on Econometric Inference Using Simulation Techniques**, S63–S84.

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* .

Tanizaki, H. (1996), 'Nonlinear filters, estimation and applications (second, revised and enlarged edition)', *Springer* .

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

Figure 1: For 5 normally distributed shocks

Notation: i) the unmarked line denotes the standard PF, ii) the line with a circle denotes the CDKF, iii) the line with a star denotes the MSPF, and iv) the line with a diamond denotes the MSPF$_b$. The x-axis shows the number of particles in thousands.

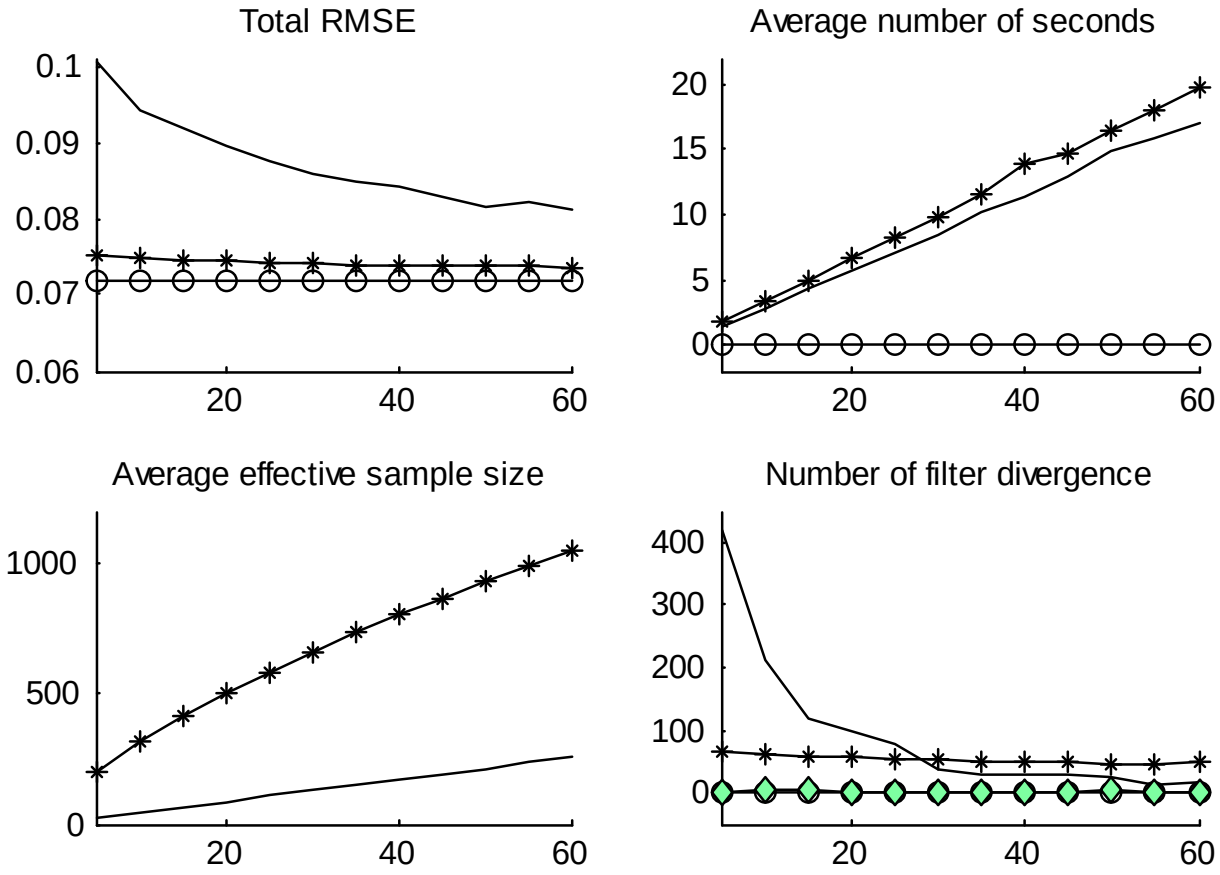


Figure 2: Monte Carlo variation in the reported log-likelihood functions
Notation: i) the unmarked line denotes the standard PF, and ii) the line with a star denotes the MSPF$_b$. Five normally distributed shocks are hitting the test economies in this case. The x-axis shows the number of particles in thousands.

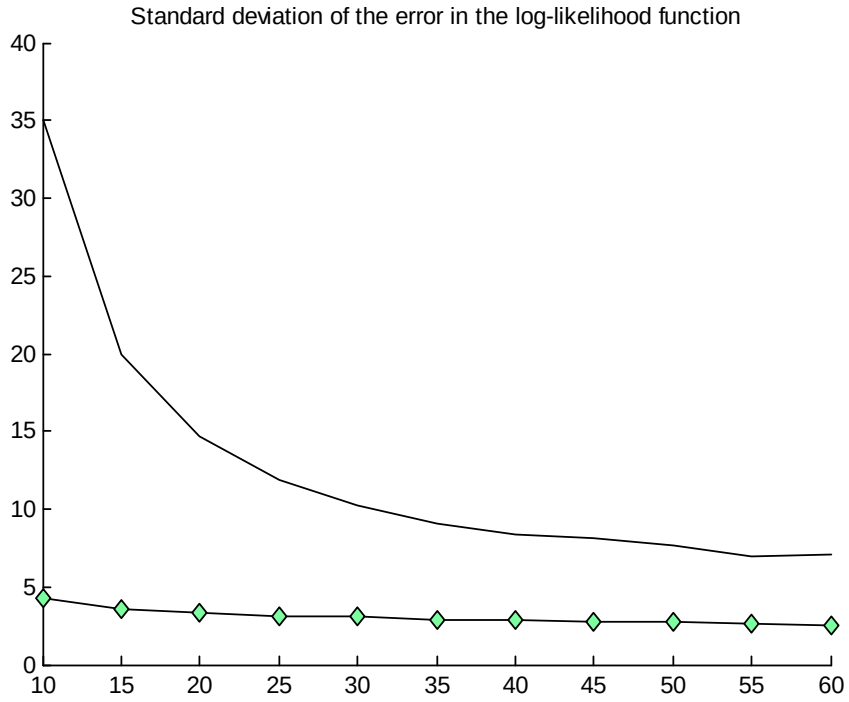


Figure 3: For 5 Laplace distributed shocks

Notation: i) the unmarked line denotes the standard PF, ii) the line with a circle denotes the CDKF, iii) the line with a star denotes the MSPF, and iv) the line with a diamond denotes the MSPF$_b$. The x-axis shows the number of particles in thousands.

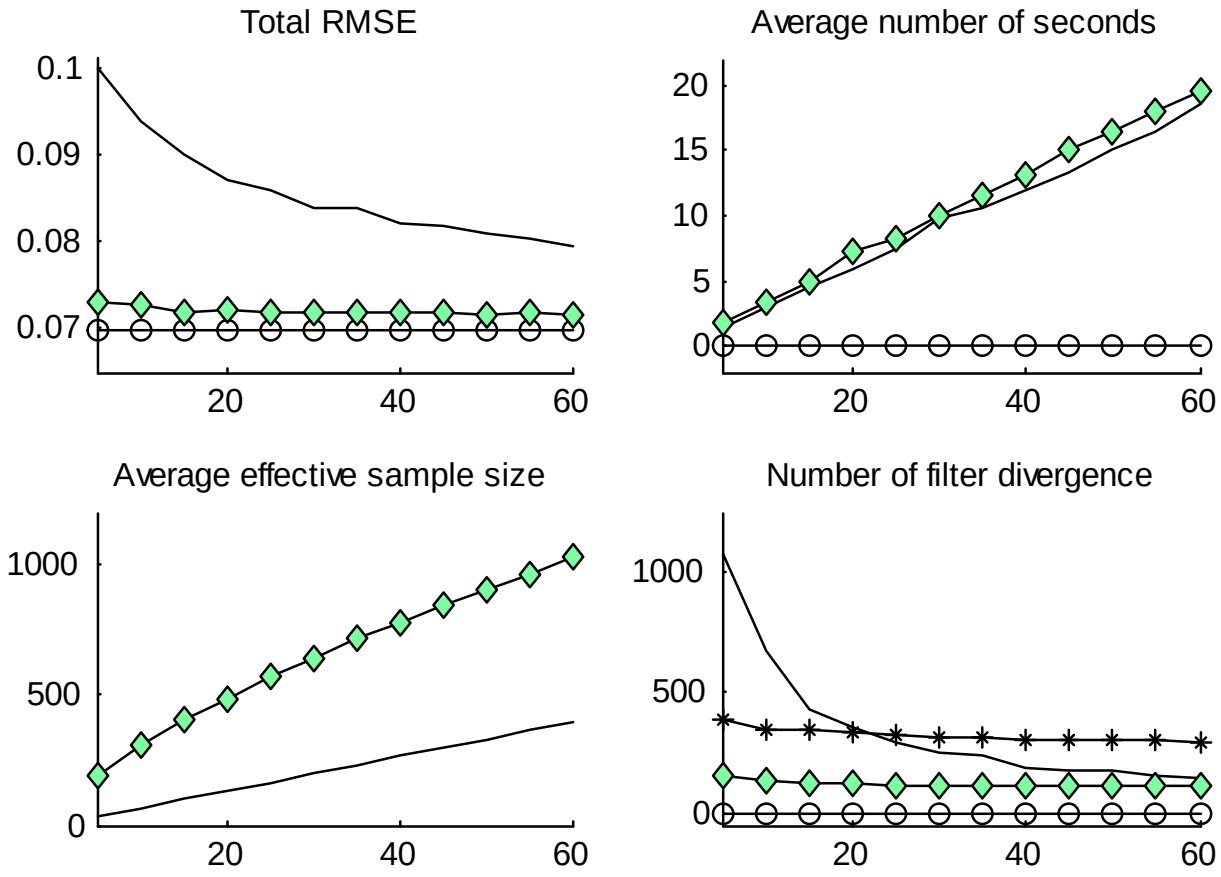


Figure 4: The total RMSE with 1 to 5 normally distributed shocks
Notation: i) the unmarked line denotes the standard PF, ii) the line with a circle denotes the CDKF, and iii) the line with a diamond denotes the MSPF$_b$. The x-axis denotes the number of particles in thousands. The MSPF$_b$ is marginally better than the standard PF with only one shock, but this difference is not visable in our graph.

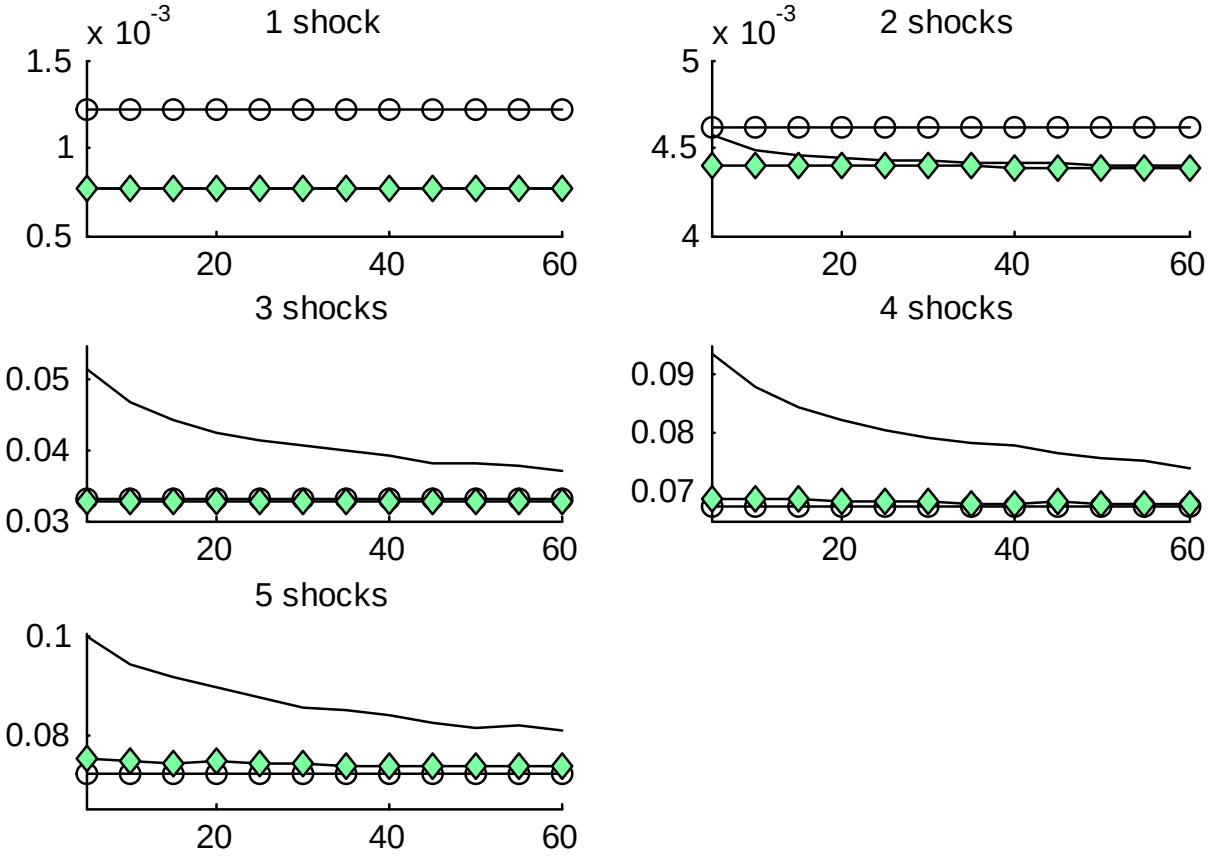


Figure 5: The Monte Carlo variation in the log-likelihood function with 1 to 5 normally distributed shocks

Notation: i) the unmarked line denotes the standard PF, and ii) the line with a diamond denotes the MSPF_b. The variation is measured by the standard deviation of the error in the log-likelihood function. The x-axis denotes the number of particles in thousands.

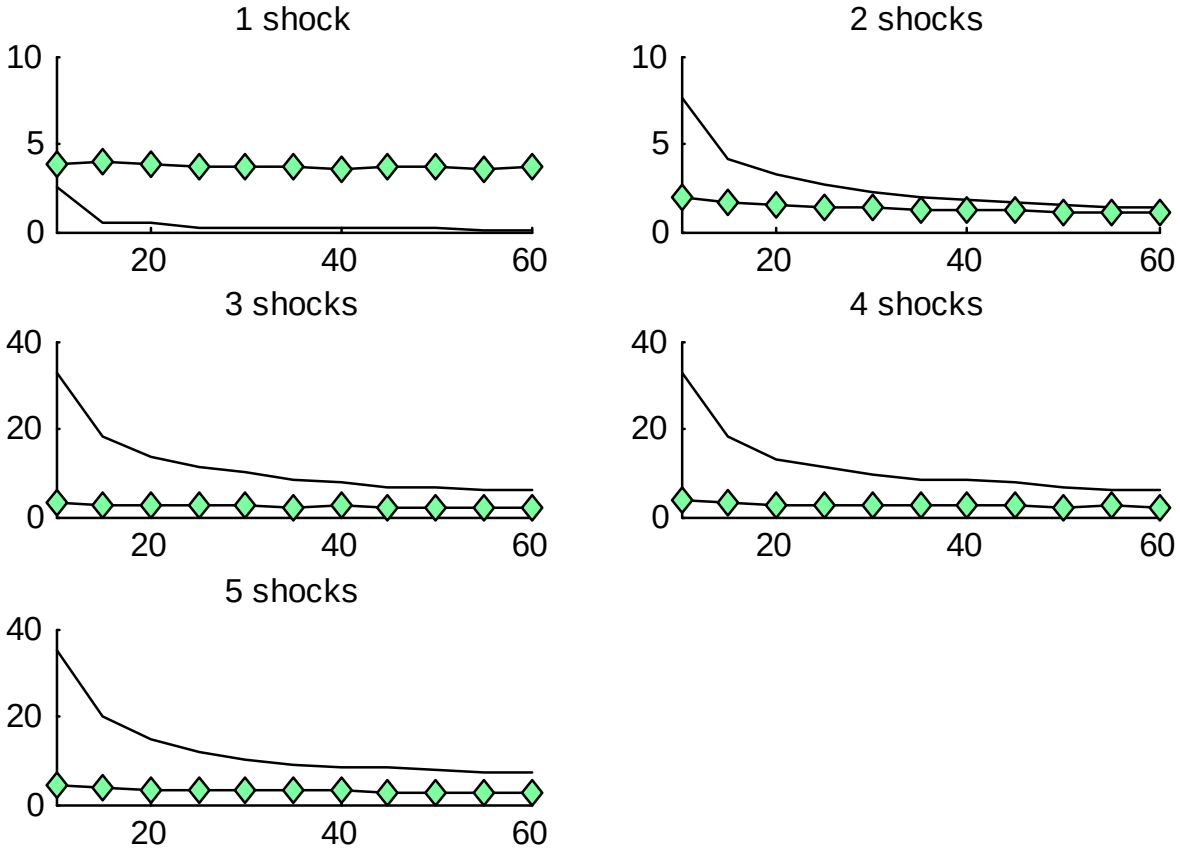


Figure 6: Distributions for the QML estimator with five normally distributed shocks
The results are based on 990 repetitions in the Monte Carlo study. The black line denotes the finite sample distribution for the QML estimator (estimated by kernel methods) and the red (gray) line is the normal distribution.

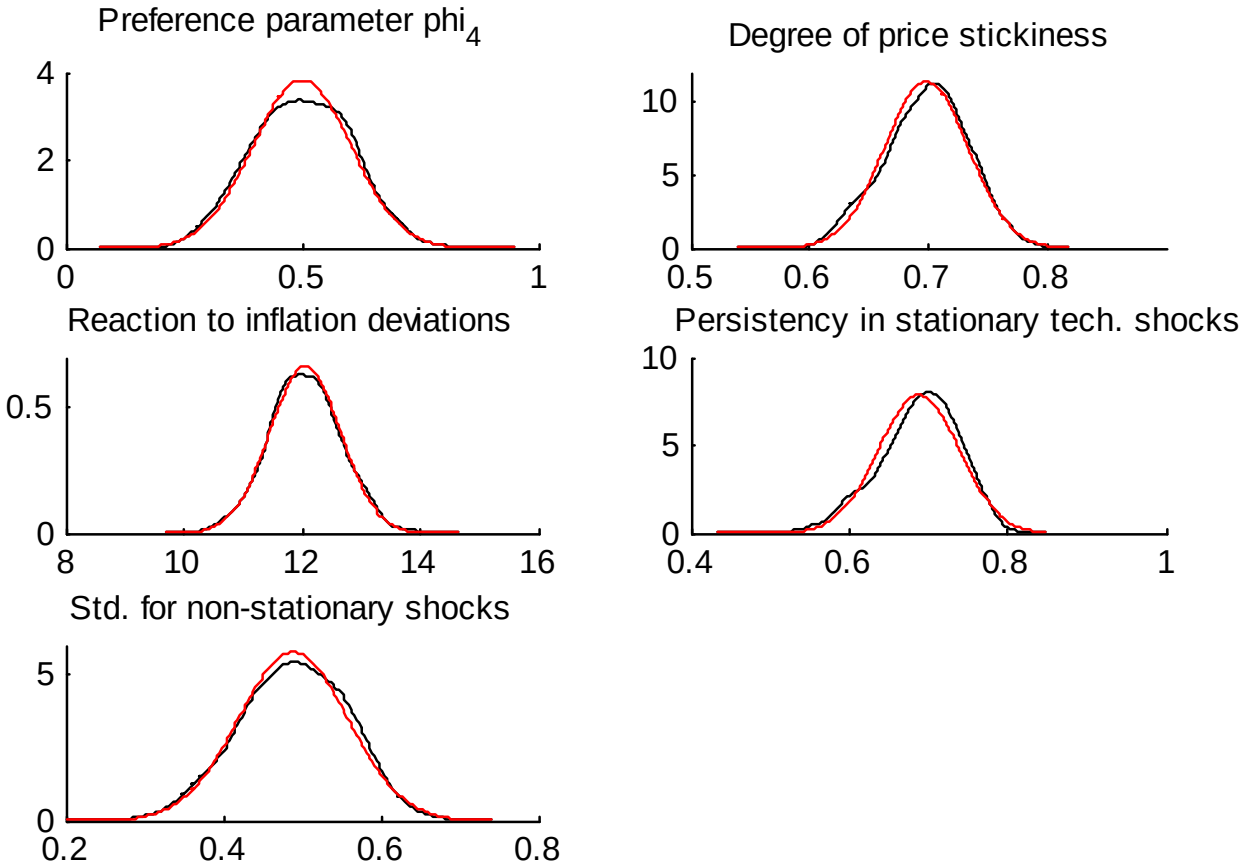
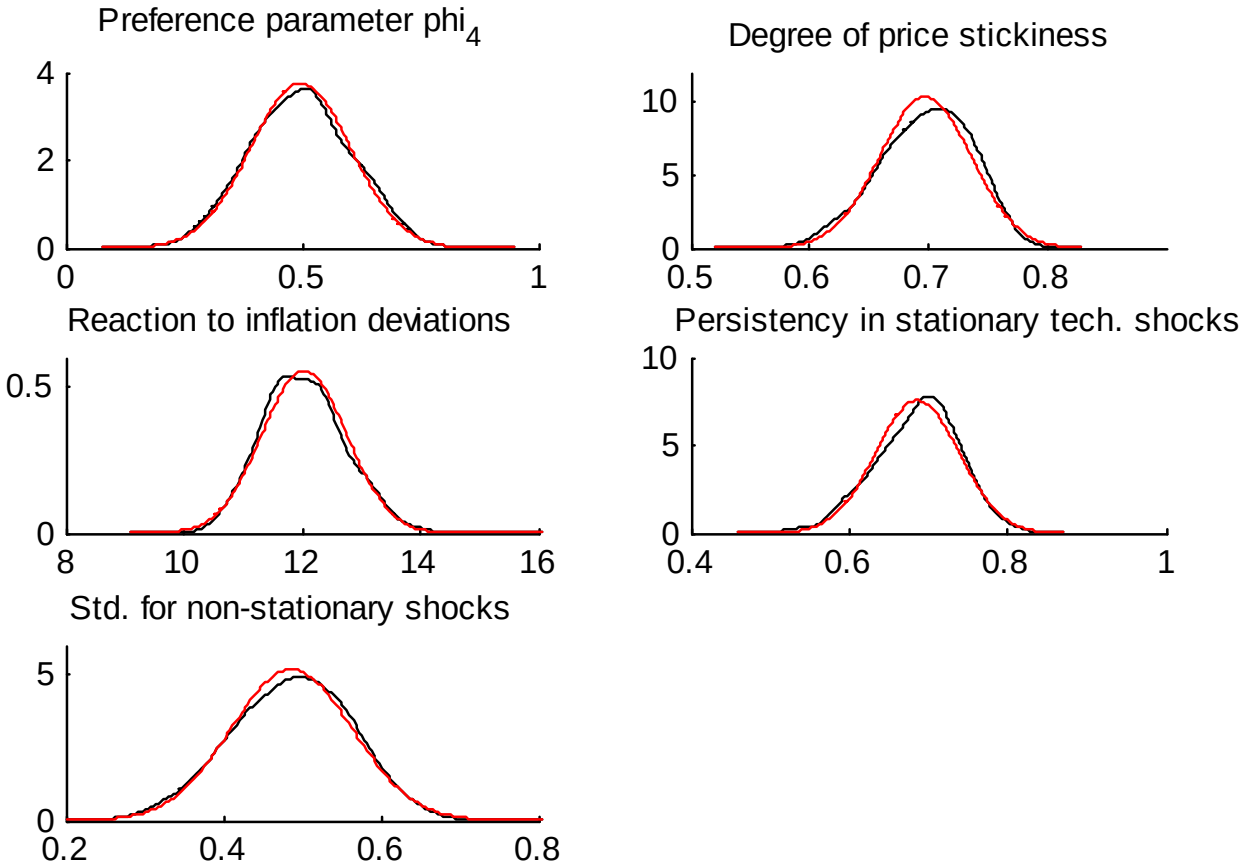


Figure 7: Distributions for the QML estimator with five Laplace distributed shocks
The results are based on 989 repetitions in the Monte Carlo study. The black line denotes the finite sample distribution for the QML estimator (estimated by kernel methods) and the red (gray) line is the normal distribution.



Research Papers 2008

2008-20: Stefan Holst Bache, Christian M. Dahl and Johannes Tang Kristensen: Determinants of Birthweight Outcomes: Quantile Regressions Based on Panel Data

2008-21: Ole E. Barndorff-Nielsen, José Manuel Corcuera, Mark Podolskij and Jeannette H.C. Woerner: Bipower variation for Gaussian processes with stationary increments

2008-22: Mark Podolskij and Daniel Ziggel: A Range-Based Test for the Parametric Form of the Volatility in Diffusion Models

2008-23: Silja Kinnebrock and Mark Podolskij: An Econometric Analysis of Modulated Realised Covariance, Regression and Correlation in Noisy Diffusion Models

2008-24: Matias D. Cattaneo, Richard K. Crump and Michael Jansson: Small Bandwidth Asymptotics for Density-Weighted Average Derivatives

2008-25: Mark Podolskij and Mathias Vetter: Bipower-type estimation in a noisy diffusion setting

2008-26: Martin Møller Andreasen: Ensuring the Validity of the Micro Foundation in DSGE Models

2008-27: Tom Engsted and Thomas Q. Pedersen: Return predictability and intertemporal asset allocation: Evidence from a bias-adjusted VAR model

2008-28: Frank S. Nielsen: Local polynomial Whittle estimation covering non-stationary fractional processes

2008-29: Per Frederiksen, Frank S. Nielsen and Morten Ørregaard Nielsen: Local polynomial Whittle estimation of perturbed fractional processes

2008-30: Mika Meitz and Pentti Saikkonen: Parameter estimation in nonlinear AR-GARCH models

2008-31: Ingmar Nolte and Valeri Voev: Estimating High-Frequency Based (Co-) Variances: A Unified Approach

2008-32: Martin Møller Andreasen: How to Maximize the Likelihood Function for a DSGE Model

2008-33: Martin Møller Andreasen: Non-linear DSGE Models, The Central Difference Kalman Filter, and The Mean Shifted Particle Filter