SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric
Analysis of Time Series

CREATES Research Paper 2010-30

Non-linear DSGE Models and The Central Difference Kalman Filter

Martin M. Andreasen

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

Non-linear DSGE Models and The Central Difference Kalman Filter*

Martin M. Andreasen†

Bank of England and CREATES

July 20, 2010

Abstract

This paper introduces a Quasi Maximum Likelihood (QML) approach based on the Central Difference Kalman Filter (CDKF) to estimate non-linear DSGE models with potentially non-Gaussian shocks. We argue that this estimator can be expected to be consistent and asymptotically normal for DSGE models solved up to third order. A Monte Carlo study shows that this QML estimator is basically unbiased and normally distributed in finite samples for DSGE models solved using a second order or a third order approximation. These results hold even when structural shocks are Gaussian, Laplace distributed, or display stochastic volatility.

Keywords: Non-linear filtering, Non-Gaussian shocks, Quasi Maximum Likelihood, Stochastic volatility, Third order perturbation.

JEL: C13, C15, E10, E32

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

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

1 Introduction

Likelihood based inference has emerged as a standard approach to estimate linearized DSGE models with unobserved state variables. The approach relies on the use of the Kalman Filter to evaluate the log-likelihood function in closed form when shocks and potential measurement errors are Gaussian. A similar closed form solution for the log-likelihood function does not exist when DSGE models are solved with non-linear terms and/or have non-Gaussian shocks. The important contribution by Fernández-Villaverde & Rubio-Ramírez (2007) addresses this problem by introducing a sequential Monte Carlo method called particle filtering as a way to estimate the log-likelihood function for non-linear DSGE modes with potentially non-Gaussian shocks. Throughout this paper we refer to the particle filter used in Fernández-Villaverde & Rubio-Ramírez (2007) as the standard Particle Filter (PF).

Although the idea of using particle filtering for likelihood inference is appealing, the method has the disadvantage of being computationally very demanding. Even for models with just two or three shocks, tens of thousands of particles are needed to get a reliable approximation to the log-likelihood function, and this makes the estimation process very time consuming - if at all feasible - for larger models. The numerical difficulties related to particle filtering imply that a relatively small number of non-linear DSGE models have so far been estimated by likelihood inference compared to the large number of estimated linearized DSGE models. Hence, more efficient filtering and estimation methods would clearly be useful in relation to estimation of non-linear DSGE models.

As an alternative to particle filtering, Norgaard, Poulsen & Ravn (2000) have developed the Central Difference Kalman Filter (CDKF) for state estimation in general non-linear and non-Gaussian state space systems. Contrary to particle filters, the updating rule for the state vector in the CDKF is restricted to have a linear functional form, and the recursive equations for the state estimator and its covariance matrix are therefore only functions of first and second moments. The CDKF approximates these moments up to at least second order accuracy by a deterministic sampling approach based on multivariate Stirling interpolations. This approximation method is computationally very fast and reasonably accurate. Like particle filtering, the CDKF frequently outperforms the Extended Kalman Filter (EKF) which for many years has been the preferred filter for non-linear and non-Gaussian systems (see for instance Jazwinski (1970)).

The contribution of this paper is to introduce a Quasi Maximum Likelihood (QML) method using the CDKF to estimate non-linear DSGE models with potentially non-Gaussian shocks. We focus on the case where the observed variables contain measurement errors, and we argue that this QML estimator can be expected to be consistent and asymptotically normal for DSGE models solved up to third order. These results hold both when Gaussian and non-Gaussian shocks are driving the economy. The main advantage of the proposed QML estimator is that it is much faster to compute than any particle filter and this greatly eases the estimation. Our QML estimator could therefore be expected to facilitate an increase in the number of estimated non-linear DSGE models.

We test the performance of the CDKF and the suggested QML estimator in a Monte Carlo study using a New Keynesian DSGE model approximated to first, second, and third order. The key results from this Monte Carlo study are as follows. Firstly, the state vector is estimated more precisely by the CDKF than the EKF and the standard PF (with 200,000 particles) when shocks are Gaussian and Laplace distributed. This shows that a linear updating rule and the multivariate Stirling interpolation used in the CDKF may lead to quite accurate approximations. However, the standard PF performs better than the CDKF and the EKF when shocks display stochastic volatility. Secondly, the quasi log-likelihood function derived from the CDKF is a better approximation to the log-likelihood function in the standard PF than the quasi log-likelihood function based on the EKF. For Gaussian shocks, the quasi log-likelihood function from the CDKF is either very close or within the 95% confidence interval for the estimated log-likelihood function in the standard PF. The same conclusion does not hold when shocks have a Laplace distribution or display stochastic volatility. Thirdly, the suggested QML estimator is found to be normally distributed and basically unbiased in finite samples for second and third order approximations to the considered DSGE model. These results hold regardless of whether shocks to the economy are Gaussian, Laplace distributed, or display stochastic volatility. Asymptotic standard errors can in all cases be computed based on the Hessian matrix and the variance of the score function. Reliable inference is here greatly facilitated by the property that only first order derivatives are needed to compute the Hessian matrix for this quasi log-likelihood function.

Based on these findings, we therefore believe that the suggested QML approach is a useful new tool for taking non-linear DSGE models to the data. Our QML estimator may be considered as an

alternative to the likelihood approach advocated by Fernández-Villaverde & Rubio-Ramírez (2007) but also as a useful supplement. For instance, the Classical researcher may use the QML estimates as good starting values for the maximization of the log-likelihood function from a particle filter as this should make the optimization considerable easier. Also the Bayesian researcher may find the QML estimator useful because i) the QML estimates can be used as good starting values for the Markov chain, and ii) the Hessian matrix of the quasi log-likelihood function can be used to specify the proposal distribution in the random walk Metropolis algorithm for the MCMC analysis. Both features should help to ensure faster convergence of the Markov chain where the log-likelihood function is estimated by a particle filter.

The rest of the paper is organized as follows. We present the state space representation of DSGE models in section 2. Some theory related to filtering is discussed in section 3 where we also present the EKF and the CDKF. We then introduce the QML approach based on the CDKF in section 4. Section 5 describes a standard DSGE model which is calibrated to account for higher order moments in the post-war US economy. The performance of the various filters and the QML estimator are examined in a Monte Carlo study in section 6. Concluding comments are provided in section 7.

2 The state space representation of DSGE models

We consider the class of DSGE models that can be represented in a dynamic state space system (see Thomas F. Cooley (1995) and Schmitt-Grohé & Uribe (2004) for illustrations). The observables in period t are denoted by the vector $\mathbf{y}_t$ with dimension $n_y \times 1$, and these observables are a function of the state vector $\mathbf{x}_t$ with dimension $n_x \times 1$. Allowing for additive measurement errors $\mathbf{v}_t$ in all observables, we then get

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

where $\mathbf{v}_t \sim \mathcal{IID}\left(\mathbf{0}, \mathbf{R}_{\mathbf{v}}\left(t\right)\right)$ denotes independent and identically distributed measurement errors. The function $\mathbf{g}\left(\cdot\right)$ is determined by the structural parameters $\boldsymbol{\theta} \in \Theta$ in the economic model and the model's equilibrium conditions. We refer to (1) as the set of measurement equations.

For the state vector, we consider a standard first order Markovian law of motion

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

where $\mathbf{w}_{t+1} \sim \mathcal{IID}\left(\mathbf{0}, \mathbf{R}_{\mathbf{w}}\left(t+1\right)\right)$. The vector $\mathbf{w}_{t+1}$ of structural innovations is assumed to be uncorrelated at all leads and lags with $\mathbf{v}_t$. We refer to (2) as the set of transition equations. The state vector is assumed to be unobserved, but observed state variables can be incorporated by letting one or more elements in $\mathbf{g}\left(\cdot\right)$ be identity mappings.

3 Filtering

The objective of filtering is to estimate the unobserved state vector $\mathbf{x}_t$ as data on $\mathbf{y}_t$ becomes available. We start in section 3.1 by deriving recursive equations for the estimation of $\mathbf{x}_t$ and its covariance matrix based on a two-step procedure of prediction and updating. The updating rule for the state vector is assumed to be linear in the observables because it leads to simple equations with only first and second moments from the state space system in (1) and (2). The next two sections show how these moments are approximated by linearization in the EKF and by multivariate Stirling interpolations in the CDKF. The final section discusses prediction and smoothing in the CDKF.

3.1 A linear updating rule for the state vector

We use the standard notation where a bar denotes a priori estimates and a hat denotes posterior estimates. For instance, $\bar{\mathbf{x}}_{t+1} \equiv E_t\left[\mathbf{x}_{t+1}\right]$ and $\hat{\mathbf{x}}_{t+1} \equiv E_{t+1}\left[\mathbf{x}_{t+1}\right]$, where E_t 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

$$\bar{\mathbf{x}}_{t+1} \equiv E_t\left[\mathbf{h}\left(\mathbf{x}_t, \mathbf{w}_{t+1}; \boldsymbol{\theta}\right)\right]. \tag{3}$$

[1] An alternative notation is $\mathbf{x}_{t+1|t} = E_t\left[\mathbf{x}_{t+1}\right]$ and $\mathbf{x}_{t+1|t+1} \equiv E_{t+1}\left[\mathbf{x}_{t+1}\right]$ 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.

The conditional error covariance matrix for this estimator is denoted by

$$\bar{\mathbf{P}}_{\mathbf{xx}}(t+1) \equiv E_t\left[(\mathbf{x}_{t+1} - \bar{\mathbf{x}}_{t+1})(\mathbf{x}_{t+1} - \bar{\mathbf{x}}_{t+1})'\right]. \tag{4}$$

The updating rule of the a priori state estimator is for tractability restricted to be linear in the observables, i.e.

$$\hat{\mathbf{x}}_{t+1} = \mathbf{b}_{t+1} + \mathbf{K}_{t+1}\mathbf{y}_{t+1}, \tag{5}$$

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

$$\mathbf{b}_{t+1} = \bar{\mathbf{x}}_{t+1} - \mathbf{K}_{t+1}\bar{\mathbf{y}}_{t+1}, \qquad \text{where } \bar{\mathbf{y}}_{t+1} \equiv E_t\left[\mathbf{g}(\mathbf{x}_{t+1}; \boldsymbol{\theta})\right]. \tag{6}$$

This gives rise to the well-known updating rule

$$\hat{\mathbf{x}}_{t+1} = \bar{\mathbf{x}}_{t+1} + \mathbf{K}_{t+1}(\mathbf{y}_{t+1} - \bar{\mathbf{y}}_{t+1}). \tag{7}$$

The value of the Kalman gain $\mathbf{K}_{t+1}$ 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))

$$\mathbf{K}_{t+1} = \mathbf{P}_{\mathbf{xy}}(t+1)\,\mathbf{P}_{\mathbf{yy}}(t+1)^{-1}, \tag{8}$$

where we have defined

$$\mathbf{P}_{\mathbf{xy}}(t+1) \equiv E_t\left[(\mathbf{x}_{t+1} - \bar{\mathbf{x}}_{t+1})(\mathbf{y}_{t+1} - \bar{\mathbf{y}}_{t+1})'\right] \tag{9}$$

$$\bar{\mathbf{P}}_{\mathbf{yy}}(t+1) \equiv E_t\left[(\mathbf{y}_{t+1} - \bar{\mathbf{y}}_{t+1})(\mathbf{y}_{t+1} - \bar{\mathbf{y}}_{t+1})'\right]. \tag{10}$$

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

$$\hat{\mathbf{P}}_{\mathbf{xx}}(t+1) = \bar{\mathbf{P}}_{\mathbf{xx}}(t+1) - \mathbf{K}_{t+1}\bar{\mathbf{P}}_{\mathbf{yy}}(t+1)\,\mathbf{K}_{t+1}'. \tag{11}$$

Thus, the filtering equations for the class of updating rules implied by (5) are given by (3), (4),

and (7) - (11).

Two remarks are in order. Firstly, if we are able to accurately evaluate the required first and second moments, then the a priori and the posterior state estimators in (3) and (7) are unbiased by construction. This result holds even though the state space system is non-linear and no distributional assumptions are imposed on $\mathbf{v}_t$ and $\mathbf{w}_t$. Secondly, the required first and second moments can be evaluated exactly when $\mathbf{g}(\cdot)$ and $\mathbf{h}(\cdot)$ are linear functions, and this leads to the Kalman Filter. Recall that the 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) imply that some approximation is needed to calculate the required moments.

3.2 The Extended Kalman Filter

One way to proceed is to linearize the state space system such that

$$\mathbf{y}_t \approx \mathbf{g}(\bar{\mathbf{x}}_t; \boldsymbol{\theta}) + \mathbf{G}_{\mathbf{x},t}(\mathbf{x}_t - \bar{\mathbf{x}}_t) + \mathbf{v}_t \tag{12}$$

$$\mathbf{x}_{t+1} \approx \mathbf{h}(\hat{\mathbf{x}}_t; \boldsymbol{\theta}) + \mathbf{H}_{\mathbf{x},t}(\mathbf{x}_t - \hat{\mathbf{x}}_t) + \mathbf{H}_{\mathbf{w},t}(\mathbf{w}_{t+1} - \bar{\mathbf{w}}_{t+1}) \tag{13}$$

where

$$\mathbf{G}_{\mathbf{x},t} \equiv \left.\frac{\partial \mathbf{g}(\mathbf{x}; \boldsymbol{\theta})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\bar{\mathbf{x}}_t} \qquad \mathbf{H}_{\mathbf{x},t} \equiv \left.\frac{\partial \mathbf{h}(\mathbf{x}; \boldsymbol{\theta})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}_t} \qquad \mathbf{H}_{\mathbf{w},t} \equiv \left.\frac{\partial \mathbf{h}(\hat{\mathbf{x}}_t, \mathbf{w}; \boldsymbol{\theta})}{\partial \mathbf{w}}\right|_{\mathbf{w}=\bar{\mathbf{w}}_{t+1}} \tag{14}$$

Using these approximations, the first and second moments in the filtering equations are given by

$$\bar{\mathbf{x}}_{t+1} = \mathbf{h}(\hat{\mathbf{x}}_t, \bar{\mathbf{w}}_{t+1}; \boldsymbol{\theta}) \qquad \bar{\mathbf{y}}_{t+1} = \mathbf{g}(\bar{\mathbf{x}}_{t+1}; \boldsymbol{\theta}) \tag{15}$$

$$\bar{\mathbf{P}}_{\mathbf{xx}}(t+1) = \mathbf{H}_{\mathbf{x},t}\hat{\mathbf{P}}_{\mathbf{xx}}(t)\mathbf{H}'_{\mathbf{x},t} + \mathbf{H}_{\mathbf{w},t}\mathbf{R}_{\mathbf{w}}(t+1)\mathbf{H}'_{\mathbf{w},t} \tag{16}$$

$$\mathbf{P}_{\mathbf{xy}}(t+1) = \bar{\mathbf{P}}_{\mathbf{xx}}(t+1)\mathbf{G}'_{\mathbf{x},t+1} \tag{17}$$

$$\bar{\mathbf{P}}_{\mathbf{yy}}(t+1) = \mathbf{G}_{\mathbf{x},t+1}\bar{\mathbf{P}}_{\mathbf{xx}}(t+1)\mathbf{G}'_{\mathbf{x},t+1} + \mathbf{R}_{\mathbf{v}}(t+1) \tag{18}$$

Applying these approximations lead to the Extended Kalman Filter (see for instance Jazwinski (1970)). The approximations in (15) are only accurate up to first order and do not account for the probability distribution of $(\hat{\mathbf{x}}_t, \mathbf{w}_{t+1})$. This is because the linearizations in (15) are done around a single point. The approximations in (16) - (18) are more precise as they are accurate up to second order. However, it is possible to improve the accuracy of these approximations with no additional computational costs as we will see in the next section.

3.3 The Central Difference Kalman Filter

The idea behind the CDKF is to approximate the non-linear expectations in (3), (4), and (7) - (11) by second order multivariate Stirling interpolations.[2] We introduce some additional notation to describe how this is done. First, let $\mathbf{S}_{\mathbf{w}}(t)$, $\mathbf{S}_{\mathbf{v}}(t)$, $\bar{\mathbf{S}}_{\mathbf{x}}(t)$, and $\hat{\mathbf{S}}_{\mathbf{x}}(t)$ be squared and upper triangular Cholesky factorizations of $\mathbf{R}_{\mathbf{w}}(t)$, $\mathbf{R}_{\mathbf{v}}(t)$, $\bar{\mathbf{P}}_{\mathbf{xx}}(t)$, and $\hat{\mathbf{P}}_{\mathbf{xx}}(t)$, respectively. That is, $\mathbf{R}_{\mathbf{w}}(t) = \mathbf{S}_{\mathbf{w}}(t)\mathbf{S}_{\mathbf{w}}(t)'$ and so on. Lower case vectors denote elements in these matrices, for instance $\mathbf{S}_{\mathbf{w}}(t+1) = \begin{bmatrix} \mathbf{s}_{\mathbf{w},1} & \mathbf{s}_{\mathbf{w},2} & ... & \mathbf{s}_{\mathbf{w},n_w} \end{bmatrix}$. We next define

$$\underset{(n_x \times n_x)}{\mathbf{S}^{(1)}_{\mathbf{xx}}}(t) = \left\{\left(h_i\left(\hat{\mathbf{x}}_t + h\hat{\mathbf{s}}_{\mathbf{x},j}, \bar{\mathbf{w}}_{t+1}; \boldsymbol{\theta}\right) - h_i\left(\hat{\mathbf{x}}_t - h\hat{\mathbf{s}}_{\mathbf{x},j}, \bar{\mathbf{w}}_{t+1}; \boldsymbol{\theta}\right)\right)/2h\right\} \tag{19}$$

$$\underset{(n_x \times n_w)}{\mathbf{S}^{(1)}_{\mathbf{xw}}}(t) = \left\{\left(h_i\left(\hat{\mathbf{x}}_t, \bar{\mathbf{w}}_{t+1} + h\mathbf{s}_{\mathbf{w},j}; \boldsymbol{\theta}\right) - h_i\left(\hat{\mathbf{x}}_t, \bar{\mathbf{w}}_{t+1} - h\mathbf{s}_{\mathbf{w},j}; \boldsymbol{\theta}\right)\right)/2h\right\} \tag{20}$$

$$\underset{(n_y \times n_x)}{\mathbf{S}^{(1)}_{\mathbf{yx}}}(t) = \left\{\left(g_i\left(\bar{\mathbf{x}}_t + h\bar{\mathbf{s}}_{\mathbf{x},j}; \boldsymbol{\theta}\right) - g_i\left(\bar{\mathbf{x}}_t - h\bar{\mathbf{s}}_{\mathbf{x},j}; \boldsymbol{\theta}\right)\right)/2h\right\} \tag{21}$$

where we use the notation $\mathbf{h}(\cdot) \equiv \begin{bmatrix} h_1(\cdot) & h_2(\cdot) & ... & h_{n_x}(\cdot) \end{bmatrix}'$ and similarly for the function $\mathbf{g}(\cdot)$. The matrices in (19) - (21) contain the first order effects of the general non-linear functions and this is denoted by the superscript (1). The corresponding matrices for the second order effects are:

$$\underset{(n_x \times n_x)}{\mathbf{S}^{(2)}_{\mathbf{xx}}}(t) = \left\{\frac{\sqrt{h^2-1}}{2h^2}\left(h_i\left(\hat{\mathbf{x}}_t + h\hat{\mathbf{s}}_{\mathbf{x},j}, \bar{\mathbf{w}}_{t+1}; \boldsymbol{\theta}\right) + h_i\left(\hat{\mathbf{x}}_t - h\hat{\mathbf{s}}_{\mathbf{x},j}, \bar{\mathbf{w}}_{t+1}; \boldsymbol{\theta}\right) - 2h_i\left(\hat{\mathbf{x}}_t, \bar{\mathbf{w}}_{t+1}; \boldsymbol{\theta}\right)\right)\right\} \tag{22}$$

$$\underset{(n_x \times n_w)}{\mathbf{S}^{(2)}_{\mathbf{xw}}}(t) = \left\{\frac{\sqrt{h^2-1}}{2h^2}\left(h_i\left(\hat{\mathbf{x}}_t, \bar{\mathbf{w}}_{t+1} + h\mathbf{s}_{\mathbf{w},j}; \boldsymbol{\theta}\right) + h_i\left(\hat{\mathbf{x}}_t, \bar{\mathbf{w}}_{t+1} - h\mathbf{s}_{\mathbf{w},j}; \boldsymbol{\theta}\right) - 2h_i\left(\hat{\mathbf{x}}_t, \bar{\mathbf{w}}_{t+1}; \boldsymbol{\theta}\right)\right)\right\} \tag{23}$$

[2] Our presentation of the CDKF is adapted to the state space system in (1) and (2).

$$\underset{(n_y \times n_x)}{\mathbf{S}^{(2)}_{\mathbf{yx}}}(t) = \left\{ \frac{\sqrt{h^2-1}}{2h^2} \left(g_i\left(\bar{\mathbf{x}}_t + h\bar{\mathbf{s}}_{\mathbf{x},j},;\boldsymbol{\theta}\right) + g_i\left(\bar{\mathbf{x}}_t - h\bar{\mathbf{s}}_{\mathbf{x},j};\boldsymbol{\theta}\right) - 2g_i\left(\bar{\mathbf{x}}_t;\boldsymbol{\theta}\right)\right)\right\} \tag{24}$$

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. It is here optimal to let h^2 equal the kurtosis of this distribution.

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

$$\begin{aligned} \bar{\mathbf{x}}_{t+1} &= \frac{h^2 - n_x - n_w}{h^2}\mathbf{h}\left(\hat{\mathbf{x}}_t, \bar{\mathbf{w}}_{t+1};\boldsymbol{\theta}\right) \\ &+ \frac{1}{2h^2}\textstyle\sum_{p=1}^{n_x}\left(\mathbf{h}\left(\hat{\mathbf{x}}_t + h\hat{\mathbf{s}}_{\mathbf{x},p}, \bar{\mathbf{w}}_{t+1};\boldsymbol{\theta}\right) + \mathbf{h}\left(\hat{\mathbf{x}}_t - h\hat{\mathbf{s}}_{\mathbf{x},p}, \bar{\mathbf{w}}_{t+1};\boldsymbol{\theta}\right)\right) \\ &+ \frac{1}{2h^2}\textstyle\sum_{p=1}^{n_w}\left(\mathbf{h}\left(\hat{\mathbf{x}}_t, \bar{\mathbf{w}}_{t+1} + h\mathbf{s}_{\mathbf{w},p};\boldsymbol{\theta}\right) + \mathbf{h}\left(\hat{\mathbf{x}}_t, \bar{\mathbf{w}}_{t+1} - h\mathbf{s}_{\mathbf{w},p};\boldsymbol{\theta}\right)\right). \end{aligned} \tag{25}$$

This approximation accounts for the distribution of the state vector $\hat{\mathbf{x}}_t$ and the distribution of the structural innovations $\mathbf{w}_{t+1}$ due to the second and third term in (25), respectively. The computation in each of these terms is very similar to computing numerical derivatives of $\mathbf{h}(\cdot)$. The important thing to notice is that the step size in (25) depends on the covariance matrix of the variable subject to the approximation, whereas the step size is arbitrary small and mutually ortogonal when numerical derivatives are computed. Norgaard et al. (2000) shows that the a priori estimator in (25) is accurate up to second order. If the state distribution and the structural shocks are normally distributed, then this estimator is even accurate up to third order.

The a priori covariance matrix of (25) is obtained by a QR decomposition of the matrix

$$\left[\begin{array}{cccc} \mathbf{S}^{(1)}_{\mathbf{xx}}(t) & \mathbf{S}^{(1)}_{\mathbf{xw}}(t) & \mathbf{S}^{(2)}_{\mathbf{xx}}(t) & \mathbf{S}^{(2)}_{\mathbf{xw}}(t) \end{array}\right]. \tag{26}$$

We follow Norgaard et al. (2000) and use the Householder transformation $\boldsymbol{\Phi}(\mathbf{A})$ to perform the QR decomposition of a rectangular matrix $\mathbf{A}$. This transformation produces a squared and upper triangular matrix $\mathbf{S} = \boldsymbol{\Phi}(\mathbf{A})$ such that $\mathbf{A}\mathbf{A}' = \mathbf{S}\mathbf{S}'$. Hence,

$$\bar{\mathbf{S}}_{\mathbf{x}}(t+1) = \Phi\left(\left[\begin{array}{cccc} \mathbf{S}^{(1)}_{\mathbf{xx}}(t) & \mathbf{S}^{(1)}_{\mathbf{xw}}(t) & \mathbf{S}^{(2)}_{\mathbf{xx}}(t) & \mathbf{S}^{(2)}_{\mathbf{xw}}(t) \end{array}\right]\right). \tag{27}$$

To see how this approximation of $\bar{\mathbf{S}}_x(t+1)$ relates to the expression in the EKF, we use the

definition of the Householder transformation to get

$$\bar{\mathbf{P}}_{\mathbf{xx}}(t+1) = \mathbf{S}_{\mathbf{xx}}^{(1)}(t)\,\mathbf{S}_{\mathbf{xx}}^{(1)}(t)' + \mathbf{S}_{\mathbf{xx}}^{(2)}(t)\,\mathbf{S}_{\mathbf{xx}}^{(2)}(t)' + \mathbf{S}_{\mathbf{xw}}^{(1)}(t)\,\mathbf{S}_{\mathbf{xw}}^{(1)}(t)' + \mathbf{S}_{\mathbf{xw}}^{(2)}(t)\,\mathbf{S}_{\mathbf{xw}}^{(2)}(t)'. \tag{28}$$

In comparison to the EKF, we thus have that $\mathbf{S}_{\mathbf{xx}}^{(1)}(t)\,\mathbf{S}_{\mathbf{xx}}^{(1)}(t)' + \mathbf{S}_{\mathbf{xx}}^{(2)}(t)\,\mathbf{S}_{\mathbf{xx}}^{(2)}(t)'$ corresponds to $\mathbf{H}_{\mathbf{x},\mathbf{t}}\hat{\mathbf{P}}_{\mathbf{xx}}(t)\,\mathbf{H}_{\mathbf{x},\mathbf{t}}'$, and $\mathbf{S}_{\mathbf{xw}}^{(1)}(t)\,\mathbf{S}_{\mathbf{xw}}^{(1)}(t)' + \mathbf{S}_{\mathbf{xw}}^{(2)}(t)\,\mathbf{S}_{\mathbf{xw}}^{(2)}(t)'$ corresponds to $\mathbf{H}_{\mathbf{w},t}\mathbf{R}_{\mathbf{w}}(t+1)\,\mathbf{H}_{\mathbf{w},t}'$. However, the approximations in (28) are in general more accurate than those in the EKF (Norgaard et al. (2000)).

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

$$\bar{\mathbf{y}}_{t+1} = \frac{h^2 - n_x}{h^2}\mathbf{g}(\bar{\mathbf{x}}_{t+1};\boldsymbol{\theta}) + \frac{1}{2h^2}\sum_{p=1}^{n_x}\left(\mathbf{g}(\bar{\mathbf{x}}_{t+1} + h\bar{\mathbf{s}}_{\mathbf{x},p};\boldsymbol{\theta}) + \mathbf{g}(\bar{\mathbf{x}}_{t+1} - h\bar{\mathbf{s}}_{\mathbf{x},p};\boldsymbol{\theta})\right) \tag{29}$$

This estimator has the same structure and properties as the a priori state estimator in (25). The covariance matrix of $\bar{\mathbf{y}}_{t+1}$ is calculated based on

$$\bar{\mathbf{S}}_{\mathbf{y}}(t+1) = \Phi\left(\begin{bmatrix} \mathbf{S}_{\mathbf{yx}}^{(1)}(t+1) & \mathbf{S}_{\mathbf{v}}(t+1) & \mathbf{S}_{\mathbf{yx}}^{(2)}(t+1) \end{bmatrix}\right), \tag{30}$$

and the Kalman gain is given by

$$\mathbf{K}_{t+1} = \bar{\mathbf{S}}_{\mathbf{x}}(t+1)\,\mathbf{S}_{\mathbf{yx}}^{(1)}(t+1)'\left[\bar{\mathbf{S}}_{\mathbf{y}}(t+1)\,\bar{\mathbf{S}}_{\mathbf{y}}(t+1)'\right]^{-1}. \tag{31}$$

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

$$\hat{\mathbf{S}}_{\mathbf{x}}(t+1) = \Phi\left(\begin{bmatrix} \bar{\mathbf{S}}_{\mathbf{x}}(t+1) - \mathbf{K}_{t+1}\mathbf{S}_{\mathbf{yx}}^{(1)}(t+1) & \mathbf{K}_{t+1}\mathbf{S}_{\mathbf{v}}(t+1) & \mathbf{K}_{t+1}\mathbf{S}_{\mathbf{yx}}^{(2)}(t+1) \end{bmatrix}\right) \tag{32}$$

An overview of the CDKF is given in appendix A.

We emphasize two properties in relation to the CDKF. Firstly, if the functions $\mathbf{g}(\cdot)$ and $\mathbf{h}(\cdot)$ are linear in $\mathbf{x}_t$ and $\mathbf{w}_{t+1}$, then the CDKF reduces to the Kalman Filter. Secondly, the number of function evaluations of $\mathbf{g}(\cdot)$ and $\mathbf{h}(\cdot)$ in the CDKF are identical to the number of function

evaluations in the EKF when derivatives are approximated by double differences, i.e. by

$$\left.\frac{\partial f(x)}{\partial x}\right|_{x=\bar{x}} = \frac{f(\bar{x}+\epsilon) - f(\bar{x}-\epsilon)}{2} \tag{33}$$

where $\epsilon > 0$ is a small number. Hence, the improved accuracy of the CDKF compared to the EKF comes at no additional computational costs.

3.4 Prediction and smoothing in the CDKF

Prediction ofthe observables in the CDKF is obtained by iterating (25), (27), and (29) 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.[3] 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.

4 Quasi maximum likelihood estimation

This section presents our quasi log-likelihood estimator for the structural coefficients $\boldsymbol{\theta}$ in non-linear DSGE models with potentially non-Gaussian shocks. We derive this estimator in section 4.1, and its asymptotic properties are discussed in section 4.2. Section 4.3 shows how to estimate the asymptotic distribution of our suggested estimator.

4.1 The quasi log-likelihood function

The EKF and the CDKF estimate the mean and the covariance matrix of the posterior state distribution. These moments are in general insufficient to derive the likelihood function except in the well-known case where all shocks are Gaussian and the functions $\mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta})$ and $\mathbf{h}(\mathbf{x}_t,\mathbf{w}_{t+1};\boldsymbol{\theta})$

[3] 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 section Norgaard et al. (2000) show that the CDKF has marginally higher theoretical accuracy than the UKF for normally distributed variables.

are linear in $\mathbf{x}_t$ and $(\mathbf{x}_t, \mathbf{w}_{t+1})$, respectively. Hence, it is in general not possible to use the EKF or the CDKF to estimate the structural parameters $\boldsymbol{\theta}$ by maximum likelihood or Bayesian methods.

A commonly used assumption for non-linear state space systems is to approximate the conditional density $\mathbf{y}_{t+1}|\,\mathbf{y}_{1:t}$ by a normal distribution, i.e.

$$\mathbf{y}_{t+1}|\,\mathbf{y}_{1:t} \overset{a}{\sim} \mathcal{N}\left(\bar{\mathbf{y}}_{t+1}\left(\boldsymbol{\theta}\right), \bar{\mathbf{P}}_{\mathbf{yy}}\left(\boldsymbol{\theta};t+1\right)\right) \tag{34}$$

for $t = 1, ..., T$. Given the higher accuracy of the CDKF compared to the EKF, it is natural to use the CDKF to compute the moments in this distribution. If the initial state vector is uncorrelated with $\mathbf{v}_t$ and $\mathbf{w}_t$ for all values of t, then a quasi log-likelihood function based on the CDKF and (34) is given by

$$\begin{aligned} L^{CDKF}\left(\boldsymbol{\theta}, \mathbf{y}_{1:T}\right) &= \frac{-n_y T}{2}\log\left(2\pi\right) - \frac{1}{2}\sum_{t=1}^{T}\log\left(\left|\bar{\mathbf{P}}_{\mathbf{yy}}^{CDKF}\left(\boldsymbol{\theta};t\right)\right|\right) \\ &\quad -\frac{1}{2}\sum_{t=1}^{T}\left(\mathbf{y}_t - \bar{\mathbf{y}}_t^{CDKF}\left(\boldsymbol{\theta}\right)\right)'\left(\bar{\mathbf{P}}_{\mathbf{yy}}^{CDKF}\left(\boldsymbol{\theta};t\right)\right)^{-1}\left(\mathbf{y}_t - \bar{\mathbf{y}}_t^{CDKF}\left(\boldsymbol{\theta}\right)\right). \end{aligned} \tag{35}$$

Here, $\bar{\mathbf{y}}_t^{CDKF}\left(\boldsymbol{\theta}\right)$ denotes the a priori estimate of the observables from the CDKF and $\bar{\mathbf{P}}_{\mathbf{yy}}^{CDKF}\left(\boldsymbol{\theta};t\right)$ is the related covariance matrix. The quasi log-likelihood function based on the EKF, denoted $L^{EKF}\left(\boldsymbol{\theta}, \mathbf{y}_{1:T}\right)$, is derived in a similar manner using the first and second moments from the EKF.

We then suggest to estimate the structural parameters $\boldsymbol{\theta}$ by maximizing the quasi log-likelihood function $L^{CDKF}\left(\boldsymbol{\theta}, \mathbf{y}_{1:T}\right)$, that is

$$\hat{\boldsymbol{\theta}}^{CDKF} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\max}\;\; L^{CDKF}\left(\boldsymbol{\theta}, \mathbf{y}_{1:T}\right). \tag{36}$$

We emphasize two convenient numerical properties in relation to this estimator. Firstly, the function $L^{CDKF}\left(\boldsymbol{\theta}, \mathbf{y}_{1:T}\right)$ is smooth in $\boldsymbol{\theta}$, and this makes the optimization relative easy. For small models, local optimization routines such as the Newton-Raphson method and its various extensions may be used with different starting values. For larger models, global optimization routines such as Simulated Annealing and evolutionary algorithms may be more effective (see Hansen, Müller & Koumoutsakos (2003) and Andreasen (2010*b*)). In comparison, the estimated log-likelihood function in particle filters do not in general display smoothness in $\boldsymbol{\theta}$ due to the resampling step and this

makes the optimization very challenging.[4]

Secondly, the numerical requirement for evaluating the quasi log-likelihood function in (35) is minimal compared to any particle filter. To realize this, note first that the CDKF uses $n_x \times (1 + 2(n_x + n_w))$ function evaluations to compute $\bar{\mathbf{x}}_{t+1}$ and the matrices for its covariance matrix. Additional $n_y(1 + 2n_x)$ function evaluations are used to compute $\bar{\mathbf{y}}_{t+1}$ and the matrices for its covariance matrix.[5] In comparison, the standard PF used in Fernández-Villaverde & Rubio-Ramírez (2007) requires $N(n_x + n_y)$ function evaluations where N is the number of particles.[6] As an illustration, consider the case with five observables ($n_y = 5$), five shocks ($n_w = 5$), and ten state variables ($n_x = 10$). This implies 415 function evaluations in the CDKF in each time period. The particle filter requires $15N$ function evaluations in each time period where N typically is between $20,000$ and $60,000$. It is therefore obvious that the CDKF is many times faster to compute than the standard PF.

4.2 Asymptotic properties

The starting point for our asymptotic analysis is the work by Bollerslev & Wooldrigde (1992) for dynamic models that jointly specify the conditional mean and the conditional covariance matrix. They examine the asymptotic properties of estimating such models by maximizing a quasi log-likelihood function which is derived from a Gaussian assumption although this distributional assumption may be violated. Hence, their setup is similar to ours for $\hat{\boldsymbol{\theta}}^{CDKF}$ in (35) and (36). The score function $\mathbf{s}(\boldsymbol{\theta})_t$ implied by (35) is:[7]

$$\begin{aligned} \underset{(1\times n_\theta)}{\mathbf{s}(\boldsymbol{\theta})'_{t+1}} &= \nabla_{\boldsymbol{\theta}} \bar{\mathbf{y}}^{CDKF}(\boldsymbol{\theta})'_{t+1} \mathbf{F}_{t+1}(\boldsymbol{\theta})^{-1} \mathbf{u}(\boldsymbol{\theta})_{t+1} \\ &\quad + \frac{1}{2} \nabla_{\boldsymbol{\theta}} \mathbf{F}_{t+1}(\boldsymbol{\theta})' \left[\mathbf{F}_{t+1}(\boldsymbol{\theta})^{-1} \otimes \mathbf{F}_{t+1}(\boldsymbol{\theta})^{-1} \right] vec\left[\mathbf{u}_{t+1}(\boldsymbol{\theta}) \mathbf{u}_{t+1}(\boldsymbol{\theta})' - \mathbf{F}_{t+1}(\boldsymbol{\theta}) \right] \end{aligned} \tag{37}$$

where

$$\mathbf{u}_t(\boldsymbol{\theta}) \equiv \mathbf{y}_t - \bar{\mathbf{y}}_t^{CDKF}(\boldsymbol{\theta}) \quad \text{and} \quad \mathbf{F}_t(\boldsymbol{\theta}) \equiv \bar{\mathbf{P}}_{\mathbf{yy}}^{CDKF}(\boldsymbol{\theta}; t) \tag{38}$$

[4] The particle filters by Pitt (2002) and Flury & Shephard (2009) are important exceptions where the estimated log-likelihood function is smooth in $\boldsymbol{\theta}$.

[5] We abstract from the computational costs related to the QR decomposition in the CDKF and the simple matrix multiplications used in the filter.

[6] The particle filter also uses a resampling step which we ignore for simplicity in our comparison of the two filters.

[7] The derivative of a matrix is defined as in Bollerslev & Wooldrigde (1992).

Here, $\mathbf{s}(\boldsymbol{\theta})_t'$ has dimension $1 \times n_\theta$ where n_θ is the number of elements in $\boldsymbol{\theta}$. If the conditional mean is correctly specified *and* the expectation $\bar{\mathbf{y}}_t^{CDKF}(\boldsymbol{\theta}) \equiv E_t[\mathbf{g}(\mathbf{x}_{t+1};\boldsymbol{\theta})]$ can be evaluated exactly, then

$$E_t[\mathbf{u}_{t+1}(\boldsymbol{\theta}_0)] = \mathbf{0}, \tag{39}$$

where $\boldsymbol{\theta}_0$ denotes the true value of $\boldsymbol{\theta}$. If the conditional covariance matrix is also correctly specified at $\boldsymbol{\theta}_0$ *and* the expectation for this covariance matrix can be evaluated exactly, then

$$E_t\left[\mathbf{u}_{t+1}(\boldsymbol{\theta}_0)\,\mathbf{u}_{t+1}(\boldsymbol{\theta}_0)'\right] = \mathbf{F}_{t+1}(\boldsymbol{\theta}_0). \tag{40}$$

As a result,

$$E_t\left[\mathbf{s}(\boldsymbol{\theta}_0)_{t+1}'\right] = \mathbf{0}. \tag{41}$$

Bollerslev & Wooldrigde (1992) show that this implies consistency and normality of $\hat{\boldsymbol{\theta}}^{CDKF}$ given standard regularity conditions. Furthermore,

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}^{CDKF} - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{A}_0^{-1}\mathbf{B}_0\mathbf{A}_0^{-1}\right), \tag{42}$$

where

$$\mathbf{B}_0 \equiv E\left[\mathbf{s}(\boldsymbol{\theta}_0, \mathbf{y}_{1:T})\,\mathbf{s}(\boldsymbol{\theta}_0, \mathbf{y}_{1:T})'\right] \tag{43}$$

$$\mathbf{A}_0 \equiv E\left[\left.\frac{\partial^2 L(\boldsymbol{\theta}, \mathbf{y}_{1:T})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right] \tag{44}$$

Accordingly, the asymptotic properties of $\hat{\boldsymbol{\theta}}^{CDKF}$ for a correctly specified DSGE model depends on the precision by which we are able to evaluate the first and second moments in $L^{CDKF}(\boldsymbol{\theta}, \mathbf{y}_{1:T})$. This insight provides a strong theoretical argument in favour of a QML estimator based on the CDKF compared to a QML estimator using the EKF, because the CDKF delivers a higher level of precision for first and second moments and should therefore have better asymptotic properties. The possibly misspecified distribution for $\mathbf{y}_{t+1}|\,\mathbf{y}_{1:t}$ in (36) is not important for consistency and normality of $\hat{\boldsymbol{\theta}}^{CDKF}$, although $\hat{\boldsymbol{\theta}}^{CDKF}$ will be more efficient the closer the true distribution of $\mathbf{y}_{t+1}|\,\mathbf{y}_{1:t}$ is to the normal distribution. This observation also implies that the stated results for $\hat{\boldsymbol{\theta}}^{CDKF}$ hold even if shocks to the DSGE models are non-Gaussian. Hence, we only need to evaluate

whether the precision delivered by the CDKF for the first and second moments is sufficient to ensure consistency and normality of $\hat{\boldsymbol{\theta}}^{CDKF}$.

The answer to this question depends on the chosen approximation order for the DSGE model. We introduce our way of reasoning by starting with a linearized DSGE model. Here, first and second moments are 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 linearized DSGE model (see for instance Hamilton (1994)).[8]

When the DSGE model is solved up to second order, then first and second moments in the model are accurate up to second and third order, respectively.[9] This implies that the precision in the CDKF is sufficient for the first moment, but there are approximation errors in the second moments because the CDKF does not match all third order terms. However, these approximation errors are likely to be insignificant as the second moments of $\mathbf{y}_{t+1}|\,\mathbf{y}_{1:t}$ are small in most cases. Alternatively, if the state vector is (approximately) Gaussian, then all third order terms are zero and the second moments in the CDKF are therefore accurate up to third order (Norgaard et al. (2000)). Thus, when a DSGE model is solved up to second order, the precision delivered by the CDKF should in all realistic settings be sufficient and our suggested QML estimator can be expected to be consistent and asymptotically normal.

For a DSGE model solved up to third order, it holds that first and second moments in the model are accurate up to third and fourth order, respectively. Hence, the estimate of the first moment in the CDKF induces approximation errors in the third order terms of this moment, unless the state vector is (approximately) Gaussian and these third order terms are zero. Approximation errors are also present in the second moments, where the CDKF does not match all third and fourth order terms. For most DSGE models, these unmatched third and fourth order terms in the first and second moments are likely to be small and our suggested QML estimator can therefore be expected to be consistent and asymptotically normal.

Finally, when the DSGE model is solved beyond third order, it becomes harder to ensure

[8] Inference is here used in the sense that the approximated DSGE model is the true data generating process. The paper by Fernandez-Villaverde, Rubio-Ramirez & Santos (2006) studies the case where the exact solution to the DSGE model is considered as the true data generating process.

[9] This result is illustrated in the appendix.

consistency and normality of the QML estimator. This is because the CDKF does not match any of the extra terms induced by solving the model beyond third order, and reliable performance of our QML estimator therefore requires that these higher order terms are insignificant.

4.3 The estimated asymptotic distribution

The variance of the score function $\mathbf{B}_0$ can be estimated in a standard fashion by first order numerical derivatives of the quasi log-likelihood function evaluated at $\hat{\boldsymbol{\theta}}^{CDKF}$. As shown by Bollerslev & Wooldrigde (1992), it is also possible to estimate the Hessian matrix $\mathbf{A}_0$ based on first order derivatives. This is a convenient property of our estimator because it is often difficult to compute reliable estimates of the Hessian matrix from double numerical derivatives. Using the notation in Harvey (1989), the Hessian matrix may be estimated by

$$\hat{A}_{ij} = \frac{1}{2T}\sum_{t=1}^{T} trace\left(\hat{\mathbf{F}}_t^{-1}\frac{\partial \hat{\mathbf{F}}_t}{\partial \hat{\theta}_i^{CDKF}}\hat{\mathbf{F}}_t^{-1}\frac{\partial \hat{\mathbf{F}}_t}{\partial \hat{\theta}_j^{CDKF}}\right) + \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial \hat{\mathbf{u}}_t}{\partial \hat{\theta}_i^{CDKF}}\right)'\hat{\mathbf{F}}_t^{-1}\frac{\partial \hat{\mathbf{u}}_t}{\partial \hat{\theta}_j^{CDKF}} \qquad (45)$$

for $i, j = 1, 2...., n_\theta$ where all first order derivatives in (45) are computed numerically.

5 A New Keynesian DSGE model

This section presents a standard New Keynesian DSGE model following the work of Christiano, Eichenbaum & Evans (2005) and Smets & Wouters (2007). We describe the model in section 5.1 and discuss how to approximate the model solution up to third order in section 5.2. The model is then calibrated in section 5.3 to account for higher order moments in the post-war US economy.

5.1 The model

We emphasize two features in relation to our model. Firstly, stochastic and deterministic trends are included to make the subsequent Monte Carlo study as realistic as possible because trends allow us to work with growth rates. Secondly, we consider the case where potentially non-Gaussian shocks drive the economy.

The households: A representative household considers

$$U_t = E_t \sum_{l=0}^{\infty} \beta^l \frac{\left(\left(\frac{c_{t+l} - b x_{t-1+l}}{z^*_{t+l}} \right)^{1-\phi_2} (1 - h_{t+l})^{\phi_2} \right)^{1-\phi_1} - 1}{1-\phi_1}, \tag{46}$$

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

$$c_t = \left[\int_0^1 c_{i,t}^{\frac{\eta-1}{\eta}} di \right]^{\frac{\eta}{\eta-1}}. \tag{47}$$

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

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

$$k_{t+1} = (1 - \delta) k_t + i_t \left(1 - \frac{\kappa}{2} (\frac{i_t}{i_{t-1}} - \mu_i)^2 \right). \tag{48}$$

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

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

$$E_t D_{t,t+1} x^h_{t+1} + c_t + (e_t \Upsilon_t)^{-1} i_t = \frac{x^h_t}{\pi_t} + r^k_t k_t + w_t h_t + \phi_t. \tag{49}$$

The left hand side of (49) is the household's total expenditures in period t which consists of i) state-contingent claims $E_t D_{t,t+1} x^h_{t+1}$, ii) consumption c_t, and iii) investment $(e_t \Upsilon_t)^{-1} i_t$. Changes in $e_t \Upsilon_t$ are investment specific shocks, which follow an exogenous AR(1) process along a deterministic trend, i.e. $\ln \Upsilon_{t+1} = \ln \Upsilon_t + \ln \mu_{\Upsilon,ss}$ and

$$\ln e_{t+1} = \rho_e \ln e_t + \epsilon_{e,t+1}. \tag{50}$$

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

The right hand side of (49) is the household's total wealth in period t. It consists of: i) pay-off

from state-contingent assets purchased in the previous period x_t^h/π_t, ii) income from selling capital services to the firms $r_t^k k_t$, iii) labor income $w_t h_t$, and iv) dividends received from firms ϕ_t. Note that π_t is the gross inflation rate.

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

$$y_{i,t} = \begin{cases} k_{i,t}^{\theta} \left(a_t z_t h_{i,t}\right)^{1-\theta} - \psi z_t^* & \text{if } k_{i,t}^{\theta} \left(a_t z_t h_{i,t}\right)^{1-\theta} - \psi z_t^* > 0 \\ 0 & \text{else} \end{cases} \tag{51}$$

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

$$\ln a_{t+1} = \rho_a \ln a_t + \sigma_a \epsilon_{a,t+1}, \tag{52}$$

where $\epsilon_{a,t+1} \sim \mathcal{IID}\left(0, Var\left(\epsilon_{a,t+1}\right)\right)$. The variable z_t in (51) denotes non-stationary technology shocks. We let $\mu_{z,t} \equiv z_t/z_{t-1}$ and assume

$$\ln \mu_{z,t+1} = \ln \mu_{z,ss} + \epsilon_{z,t+1}, \tag{53}$$

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

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

$$d_{i,t} \equiv E_t \sum_{l=0}^{\infty} D_{t,t+l} P_{t+l} \phi_{i,t+l}. \tag{54}$$

Here, $D_{t,t+l}$ is the nominal stochastic discount factor and the expression for real dividend payments from the i'th firm $\phi_{i,t}$ is given below in (56). 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 quantity of good

i is allocated to consumption and investment, which implies

$$y_t = c_t + (e_t \Upsilon_t)^{-1} i_t. \tag{55}$$

The second constraint is the budget restriction:

$$\phi_{i,t} = (P_{i,t}/P_t)\, y_{i,t} - r_t^k k_{i,t} - w_t h_{i,t}. \tag{56}$$

The first term in (56) denotes the real revenue from sales of the i'th good. The next two terms in (56) are the firm's expenditures on capital services $r_t^k k_{i,t}$ and payments to workers $w_t h_{i,t}$.

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

The central bank: The central bank determines the gross one-period nominal interest rate R_t according to the rule

$$\ln\left(\frac{R_t}{R_{ss}}\right) = \rho_R \ln\left(\frac{R_{t-1}}{R_{ss}}\right) + \beta_\pi \ln\left(\frac{\pi_t}{\pi_{ss}}\right) + \beta_y \ln\left(\frac{y_t}{y_{ss} z_t^*}\right) + \epsilon_{R,t+1}, \tag{57}$$

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

Shock distributions: Three specifications for the shocks are considered. As a benchmark case, we first let all shocks be Gaussian. In our second specification, we let shocks be generated from the Laplace distribution which has thicker tails than the Gaussian distribution. The presence of Laplace distributed shocks is interesting as it implies that large shocks occur more often than with Gaussian shocks, and this allow us to test whether the CDKF is robust to state outliers.

Our final specification, considers another way to model non-Gaussian shocks by introducing stochastic volatility into the model. We focus on the case where only the process for stationary technology shocks displays stochastic volatility. Applying the specification in Justiniano & Primiceri

(2008), the law of motion for a_t in (52) is replaced by

$$\ln a_{t+1} = \rho_a \ln a_t + \sigma_{a,t+1}\epsilon_{a,t+1}, \tag{58}$$

where the process for stochastic volatility $\sigma_{a,t+1}$ is given by

$$\ln \sigma_{a,t+1} = \rho_{\sigma_a} \ln \sigma_{a,t} + \epsilon_{\sigma_a,t+1}. \tag{59}$$

Here, $\epsilon_{\sigma_a,t+1} \sim \mathcal{NID}\left(0, Var\left(\epsilon_{\sigma_a,t+1}\right)\right)$.[10] Non-stationary technology shocks z_t and investment specific shocks e_t are assumed to be Gaussian in this final specification, and we do not consider monetary policy shocks, i.e. $Var\left(\epsilon_{R,t+1}\right) = 0$. The latter restriction is imposed for numerical convenience as the presence of stochastic volatility in a_t introduces $\sigma_{a,t}$ as an additional state variable. Hence, by omitting monetary policy shocks, we are left with the same number of state variables as in the two shock specifications without stochastic volatility. An overview of the three specifications for the structural shocks is provided in Table 1.

< Table 1 about here >

5.2 Solving the model

We now present the state space representation of our DSGE model. As in section 2, let $\mathbf{y}_t$ contain the non-predetermined variables and let $\mathbf{x}_t$ contain the predetermined state vector. Both variables are expressed in deviation from the deterministic steady state. Denoting the structural innovations by $\boldsymbol{\epsilon}_{t+1}$, the exact solution to our model is then given by

$$\mathbf{y}_t = \mathbf{g}\left(\mathbf{x}_t, \sigma\right) \tag{60}$$

$$\mathbf{x}_{t+1} = \mathbf{h}\left(\mathbf{x}_t, \sigma\right) + \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+1} \tag{61}$$

[10] Other specifications of stochastic volatility in DSGE models are discussed in Andreasen (2010*d*).

where σ is the perturbation parameter. This is also the exact solution when stochastic volatility is included in the model because the processes in (58)-(59) can be represented as

$$\ln a_t = \sigma_{a,t} \ln v_t \tag{62}$$

$$\ln v_{t+1} = \rho \frac{\sigma_{a,t}}{\sigma_{a,t+1}} \ln v_t + \epsilon_{a,t+1} \tag{63}$$

$$\ln \sigma_{a,t+1} = \rho_{\sigma_a} \ln \sigma_{a,t} + \epsilon_{\sigma_a,t+1} \tag{64}$$

The proof of this result is given in Andreasen (2010*a*). Hence, our specification of stochastic volatility can be expressed as a combination of two AR(1) processes where the local persistency coefficient in the auxiliary process $\ln v_t$ is modified by the term $\sigma_{a,t}/\sigma_{a,t+1}$. The representation in (62)-(64) is convenient because it shows that we only need to perturb the variables $(a_t, v_t, \sigma_{a,t})$ and it implies only two state variables, $(v_t, \sigma_{a,t})$. In comparison, the representation in (58)-(59) requires perturbing $(a_t, \sigma_{a,t}, \epsilon_{a,t+1}, \epsilon_{\sigma,t+1})$ and has as a minimum three state variables, $(a_t, \sigma_{a,t}, \epsilon_{a,t+1})$. In particular, the smaller number of state variables induced by the representation in (62)-(64) is convenient as it reduces the computing time when DSGE models are solved numerically.

The functions $\mathbf{g}(\cdot)$ and $\mathbf{h}(\cdot)$ are unknown, and we therefore approximate them up to third order. We apply the codes by Schmitt-Grohé & Uribe (2004) to compute all the first and second order terms. The third order terms are computed using the codes derived in Andreasen (2010*a*) which extends the results in Schmitt-Grohé & Uribe (2004) to third order. We finally apply the pruning scheme to the approximated solution (see Kim, Kim, Schaumburg & Sims (2008) and Andreasen (2010*c*) for further details).

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

$$\mathbf{y}_t^{obs} = \mathbf{M}_1 \mathbf{g}(\mathbf{x}_t, \sigma, \boldsymbol{\theta}) - \mathbf{M}_2 \mathbf{g}(\mathbf{x}_{t-1}, \sigma, \boldsymbol{\theta}) + \mathbf{v}_t \tag{65}$$

$$\mathbf{x}_{t+1} = \mathbf{h}\left(\mathbf{x}_t, \sigma\right) + \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+1} \tag{66}$$

where $\mathbf{M}_1$ and $\mathbf{M}_2$ are selection matrices with appropriate dimensions. The presence of $\mathbf{x}_{t-1}$ in (65) is due to the three growth rates in $\mathbf{y}_t^{obs}$.

5.3 Model calibration

The model is calibrated to US data from 1956Q4 to 2009Q2. We focus on matching i) mean values, ii) standard deviations, iii) skewness, and iv) kurtosis for the five series in $\mathbf{y}_t^{obs}$.[11] The calibrated coefficients are fairly standard and summarized in Table 2. We only note that the conditional standard deviation in the volatility process $\sqrt{Var\left(\epsilon_{\sigma_a,t+1}\right)}$ has a relative large value of 0.3 which is needed to get notable effects of stochastic volatility in the model.

< Table 2 about here >

The empirical and simulated moments are reported in Table 3 for a second order approximation. Our model is successful at matching all mean values and standard deviations when shocks are Gaussian. The model also generates sizeable deviations from normality in the nominal interest rate and the inflation rate. This is evident from the values of skewness and kurtosis for these series which differ from 0 and 3, respectively. Accordingly, our model has significant non-linearities and should therefore be challenging for non-linear filters. The main difference when going from Gaussian to Laplace distributed shocks is that the latter shocks increase the value of kurtosis. The presence of stochastic volatility has the same effect, although the increase in kurtosis tends to be smaller.

< Table 3 about here >

The simulated moments using a third order approximation are reported in Table 4. Moving from a second order to a third order approximation is seen to increase the values of skewness and kurtosis for all variables. For the nominal interest rate and the inflation rate, the values of these moments are even seen to be larger than the corresponding empirical moments. Another indicator

[11] We use data from the Federal Reserve Bank of St. Louis. The quarterly interest rate is measured by the rate in the secondary market (TB3MS). The quarterly inflation rate is for consumer prices. The growth rate in consumption is calculated from real consumption expenditures (PCECC96). The series for real private fixed investment (FPIC96) is used to calculate the growth rate in investment. The growth rate in output is calculated from real GDP (GDPC96). All growth rates are expressed in quarterly terms and in per capita based on the total population in the US.

of the strong non-linearities in the model is to note that the standard deviation for the nominal interest rate and the inflation rate are somewhat higher at third order than at second order.

< Table 4 about here >

6 A Monte Carlo study

This section conducts a Monte Carlo study to evaluate the performance of the CDKF in the context of DSGE models. We start in section 6.1 by examining how accurately the CDKF estimates the unobserved state variables compared with the EKF and the standard PF. In section 6.2, the values of the quasi log-likelihood function using the CDKF and the EKF are compared to the estimated value of the log-likelihood function in the standard PF. Section 6.3 explores the finite sample properties of our suggested QML estimator based on the CDKF and compares its performance with a QML estimator that uses the EKF.

6.1 State estimation

This section studies the ability of the CDKF to estimate the unknown state variables. We simulate 20 sample paths of length $T = 200$ for a given approximation order to the DSGE model in order to generate 20 test economies. This is done for each of the three shock specifications. The accuracyof all filters is measured by the root mean squared errors (RMSE)

$$RMSE_i = \sum_{j=1}^{n_x} \sqrt{\frac{\sum_{t=1}^{T} \left(x_{j,t} - \hat{x}_{j,t}\right)^2}{T}} \quad \text{for } i = 1, 2, ..., 20, \tag{67}$$

where $x_{j,t}$ is the true value of the j'th state variable in period t and $\hat{x}_{j,t}$ is the estimated value. The initial state $\mathbf{x}_0$ is taken to be known and we let $\mathbf{P}_{\mathbf{xx}}(t=0) = 10^{-6}$. The step size h in the CDKF is always set equal to the optimal value for the Gaussian distribution $\left(h = \sqrt{3}\right)$ even though the distribution subject to the multivariate Stirling interpolation may be non-Gaussian.[12] We compare the performance of the CDKF with the EKF and the standard PF using 200,000 particles. The latter is included to represent a close approximation to the optimal state estimator.

[12] These assumptions are maintained throughout the Monte Carlo study.

We start by considering the case where all shocks are Gaussian in Figure 1. The first chart in this figure shows the RMSE for a first order approximation to the DSGE model, and the CDKF (market with stars) therefore reduces to the Kalman Filter which is the optimal state estimator. The performance of the standard PF is reported as a 95% confidence interval which we compute from 100 repetitions of the filter on the same test economy. The standard PF is seen to give a satisfying approximation to the optimal estimator as the confidence intervals either contain or are very close to the optimal state estimates from the Kalman Filter.

For a second order approximation, we first note that the CDKF gives more precise state estimates than the EKF (market with squares). More surprisingly, the CDKF is doing just as well as the standard PF and clearly outperforms this filter for some test economies. This finding indicates that a linear updating rule and the multivariate Stirling interpolation used in the CDKF are reasonable approximations for this class of test economies.

For a third order approximation, the CDKF still outperforms the EKF, and the CDKF also does marginally better than the standard PF in this case. Given the strong non-linearities in our DSGE model, we consider the latter as a very surprising result.

< Figure 1 about here >

For Laplace distributed shocks in Figure 2, we find more or less the same results as with Gaussian shocks. That is, the CDKF clearly outperforms the EKF, and the CDKF is marginally better than the standard PF. Hence, the good performance of the CDKF is found to be robust to strong non-linearties and Laplace distributed shocks. Recall that these shocks are interesting to consider because they generate state outliers more frequently than Gaussian shocks.

< Figure 2 about here >

The results for shocks displaying stochastic volatility are shown in Figure 3. Here, we do not consider a first order approximation because it does not capture the presence of stochastic volatility as emphasized by Fernández-Villaverde & Rubio-Ramírez (2007). For a second and third order approximation, the standard PF is seen to outperform the CDKF and the EKF as the 95% confidence intervals from the particle filter are lower than the RMSE for the two other filters. An inspection of the state distributions in the standard PF shows that many of these distributions

are highly skewed and non-Gaussian, and this may explain the benefit of tracking the entire state distribution in the particle filter instead of just focusing on the first two moments as in the CDKF and the EKF. These state distributions are on the other hand typically symmetric and bell-shaped with Gaussian and Laplace distributed shocks although the DSGE model displays strong non-linearities. Note finally, that the level of the RMSE is rather large with stochastic volatility in the model in comparision to Gaussian and Laplace distributed shocks in Figure 1 and 2. This is because the process for $\sigma_{\alpha,t}$ is very volatile and hence dominates the expression for the RMSE.

< Figure 3 about here >

The accuracy results for the three filters are summarized in Table 5 where we compute the average RMSE across the 20 test economies for the various shock specifications and the different approximations to the model. In line with our previous findings, we see that the CDKF outperforms the EKF and the standard PF when shocks are Gaussian or Laplace distributed. This is indicated by bold figures for the CDKF in Table 5. Hence, more than 200,000 particles are needed in the standard PF if this filter is to outperform the CDKF for the considered test economies. The benefit of particle filtering is however evident with stochastic volatility where the RMSE is substantially lower for the standard PF than for the CDKF and the EKF.

< Table 5 about here >

The performance of a given filter should also be evaluated in relation to the time it takes to compute the filter. Table 6 therefore shows the average computing time for each of three filters.[13] The CDKF is seen to be very fast to compute. It only takes about 0.03 seconds for one evaluation of this filter with a first order approximation, and this number only increases to 0.10 seconds and 0.50 seconds for a second order and a third order approximation, respectively. The corresponding figures for the standard PF using 200,000 particles are 20.25 seconds for a first order approximation, 68.42 seconds for a second order approximation, and 299,41 seconds for a third order approximation. Hence, the CDKF becomes a more attractive alternative to particle filtering when the approximation order to the DSGE model is increased.

[13] Our presentation and implementation of the CDKF propogates the square root of covariance matrices through time to ensure that these matrices remain positive semi-definite. We therefore also use a square root implementation of the EKF when comparing the execution time of the two filters. However, this implementation of the EKF is marginally slower to compute compared with the version of the EKF given in section 3.2.

< Table 6 about here >

6.2 The quasi log-likelihood function

Before we turn to the performance of the suggested QML estimator, it is interesting to see how close the quasi log-likelihood function L^{CDKF} is to the estimated log-likelihood function in the standard PF, L^{PF}. Also the difference between the quasi log-likelihood function based on the CDKF and the EKF, i.e. L^{CDKF} versus L^{EKF}, is interesting to examine because it shows the quantitative effect of using different methods to estimate first and second moments in the two filters. We therefore study the relationship between these three functions in the current section.

A natural metric when comparing these functions, is to express the quasi log-likelihood functions in percentage deviation from the value of the log-likelihood function. The latter is in general not available and we therefore use the estimated log-likelihood function from the standard PF. Hence, for the j'th test economy we compute

$$\Psi_j^i = 100\frac{\left(L_j^i - \bar{L}_j^{PF}\right)}{\bar{L}_j^{PF}} \tag{68}$$

for $i = \{\text{CDKF,EKF}\}$ and $j = 1, 2, ..., 20$. Here, $\bar{L}_j^{PF} = 1/100 \sum_{k=1}^{100} L_{k,j}^{PF}$ denotes the mean value of the log-likelihood function in the standard PF from 100 evaluation of this filter on the j'th test economy. The normalizing constant in (68) is stochastic, and we therefore find it useful to also report uncertainty bounds. This is done by computing the 95% confidence interval $\pm 1.96 \times 100\frac{std\left(L_j^{PF}\right)}{\bar{L}_j^{PF}}$, where $std\left(L_j^{PF}\right)$ denotes the standard deviation of the estimated log-likelihood function. Accordingly, this confidence interval shows the variation in the estimated log-likelihood function (when normalized by $\bar{L}_j^{PF}$) which is due to Monte Carlo variation.[14]

The values of Ψ_j^i with Gaussian shocks are shown in Figure 4. For a first order approximation, the CDKF (and the EKF) reduces to the Kalman Filter which reports the exact value of the log-likelihood function. We thus see from the first chart in Figure 4 that the particle filter gives a good

[14] We prefer this normalized metric in terms of $\bar{L}_j^{PF}$ instead of simply reporting the level of L^{CDKF}, L^{EKF}, and L^{PF} because the latter would make the differences between the three functions hard to notice in the subsequent charts.

approximation to the log-likelihood function because the exact value is either within or very close to the 95% confidence interval for the standard PF.

We draw two conclusions from the middle chart of Figure 4 which is for a second order approximation. Firstly, the values of Ψ_j^{CDKF} are closer to zero than Ψ_j^{EKF} for most of the 20 test economies. This means that the quasi log-likelihood function using the CDKF is a better approximation to the log-likelihood function than the quasi log-likelihood function based on the EKF. We therefore expect QML estimates from the CDKF to be more efficient than QML estimates based on the EKF because the Maximum Likelihood (ML) estimator is asymptotically efficient. Secondly, the quasi log-likelihood functions using the CDKF and the EKF are both very close to the log-likelihood function from the standard PF. For the CDKF, we even note that in 13 out of the 20 test economies, the value of L^{CDKF} lies within the 95% confidence interval. This implies that the efficiency loss of using our suggested QML estimator instead of a fully efficient ML estimator is likely to be small for these test economies.

For a third order approximation in the bottom chart of Figure 4, we again find that L^{CDKF} is closer to the log-likelihood function than L^{EKF}, and both quasi log-likelihood functions are fairly close to L^{PF}. Given the strong non-linearities in our DSGE model, we consider the latter result as a quite surprising.

< Figure 4 about here >

Results based on Laplace distributed shocks are shown in Figure 5. Regardless of the approximation order to the DSGE model, we find that L^{CDKF} and L^{EKF} display fairly large deviations from the estimated log-likelihood function. This means that our suggested QML estimator in this case is likely to be somewhat less efficient than the ML estimator. For a second and third order approximation, we also note that L^{CDKF} in general is closer to the log-likelihood function than L^{EKF}. The same two conclusions also hold when shocks display stochastic volatility as shown by Figure 6.

< Figure 5 about here >

< Figure 6 about here >

The results from Figure 4 - 6 are summarized in Table 7. The first two columns in this table display the average distance between L^{PF} and the quasi log-likelihood functions across the 20 test economies. We see that the value of the quasi log-likelihood function using the CDKF on average is closer to L^{PF} than the quasi log-likelihood function derived based on the EKF. This is a very robust result as it holds for a second and a third order approximation and for the three considered shock specifications. For Gaussian shocks, the average distance between L^{CDKF} and L^{PF} is only 2.29 and is therefore very close to the Monte Carlo variation of 2.01 in the standard PF (column three). The difference between L^{CDKF} and L^{PF} is also seen to be small compared to the average value of L^{PF} which is slightly larger than 4000 (column four). In other words, the CDKF provides a very good approximation to the log-likelihood function for these test economies when shocks are Gaussian.

The same conclusion does not hold with Laplace distributed shocks where the quasi log-likelihood functions differ substantially from the log-likelihood in the standard PF. Even larger differences appear between the quasi log-likelihood functions and the estimated log-likelihood functions when shocks display stochastic volatility. In particular the latter finding is interesting because the state space representation of the DSGE model with stochastic volatility has the same structure as when all shocks are Gaussian. The key difference between the two specifications stems from the fact that the system with stochastic volatility has much larger second and third order terms than the system without this feature.

< Table 7 about here >

6.3 QML estimates

This section examines the finite sample properties of the QML estimator for the structural parameters in our DSGE model. We only consider five of the parameters in the model to be unknown as this makes the simulation study numerically feasible. The five unknown parameters are: i) the preference parameter (ϕ_1), ii) the degree of price stickiness (α), iii) the central bank's reaction to deviations from the inflation rate target (β_π), iv) the degree of persistency in stationary technology shocks (ρ_a), and v) the standard deviation for non-stationary technology shocks $\left(\sqrt{Var\left(\epsilon_{z,t+1}\right)}\right)$. The values of these and all the other parameters are given in Table 1.

We start by examining the properties of the suggested QML estimator when shocks are Gaussian in Table 8. For a second order approximation, the QML estimator based on the CDKF is basically unbiased. We also note that the estimates of ϕ_1, α, and β_π are closer to their desired level for the QML estimator based on the CDKF than for the QML estimator using the EKF. The opposite is marginally the case for ρ_a whereas both QML estimators give unbiased estimates of $\sqrt{Var\left(\epsilon_{z,t+1}\right)}$. The true standard errors are smallest when using the CDKF, which is in line with the results from the previous section where L^{CDKF} was found to be closer to L^{PF} than L^{EKF}. The asymptotic standard errors provide more or less unbiased estimates of these standard errors for both QML estimators. Finally, the Type I errors at a 5% significance level are in most cases closer to 5% for the CDKF than for the EKF. This shows that the asymptotic normal distribution for $\boldsymbol{\theta}^{CDKF}$ is a reasonable approximation to its distribution for a finite sample.

Using a third order approximation, we find that the biases in the level of the structural parameters and their standard errors are quite small for both QML estimators. We also note that the true standard errors for ϕ_1, α, and ρ_a are smaller for a third order approximation than for a second order approximation, and the true standard errors in a second order approximation are smaller than for a first order approximation. Hence, a more accurate approximation to the model implies more information about the parameters in the quasi log-likelihood function and therefore more efficient estimates. An & Schorfheide (2007) document a similar result using Bayesian estimation methods when going from a first order to a second order approximation.

< Table 8 about here >

The QML estimates for Laplace distributed shocks are shown in Table 9. For a second order and third order approximation, we find that the QML estimates using the CDKF only have negligible biases, and these biases are slightly smaller than those related to the EKF. The QML estimates based on the CDKF are futher seen to have slightly smaller standard errors than the EKF. The latter result is consistent with the finding in the previous section where L^{CDKF} also with Laplace distributed shocks was closer to L^{PF} than L^{EKF}. The estimates of the standard errors are basically unbiased for both QML estimators. However, the two estimators tend to produce too high Type I errors, but so does the QML estimator for a first order approximation to the model. This suggests that the high Type I errors are unrelated to the presence of non-linearities in the model and instead

relates to the Laplace distributed shocks, which imply that a longer sample is needed for convergence to the asymptotic normal distribution. Note also that for these shocks, a third order approximation to the model in general give more efficient estimates when compared to the linearized model.

< Table 9 about here >

The results for shocks with stochastic volatility are reported in Table 10. Again, our suggested QML estimator based on the CDKF results in negligible biases for a second and a third order approximation, and the same conclusion holds for the estimated standard errors. We also observe a tendency for the CDKF to do better than the EKF along these dimensions, although both filters perform quite well. Note also that the standard errors for the QML estimates with the CDKF in general are smaller than the standard errors for the EKF. As for the previous two shock specifications, most estimates become more efficient when we increase the approximation order.

The good performance of our QML estimator with stochastic volatility is interesting because section 6.1 found relatively large RMSE for the state estimates in the CDKF when compared to the standard PF. Hence, the omitted higher order terms in the CDKF (and also in the EKF) seem to be of less importance for the performance of the considered QML estimator even in a situation with very strong non-linearities. This finding therefore supports the conjecture in section 4.2 that the unmatched terms in the CDKF are unlikely to be significant for the performance of a QML estimator based on the CDKF when DSGE models are approximated up to third order.

< Table 10 about here >

7 Conclusion

This paper suggests a QML estimator based on the CDKF to estimate non-linear DSGE models with potentially non-Gaussian shocks. Focus is devoted to the case where measurement errors are present in the observables, and we argue that this QML estimator can be expected to be consistent and asymptotically normal for DSGE models solved up to third order. These results hold when Gaussian and potentially non-Gaussian shocks are driving the economy. The main advantage of this estimator is that it is much faster to implement compared with an estimator that relies on the use of a particle filter.

The performance of the CDKF and the suggested QML estimator is examined on a standard New Keynesian DSGE model solved by first, second, and third order approximations. We find that the CDKF gives more precise state estimates than the EKF and the standard PF when shocks are Gaussian and Laplace distributed. However, the standard PF performs better than the CDKF and the EKF when shocks display stochastic volatility. We also show that the quasi log-likelihood function derived from the CDKF is a better approximation to the log-likelihood function in the standard PF than the quasi log-likelihood function based on the EKF. The distribution of the suggested QML estimator is further found to be well approximated by its asymptotic distribution, and the estimator is more or less unbiased in finite samples for a second and a third order approximation to the considered DSGE model. We found a tendency for the QML estimator based on the CDKF to be more efficient than a QML estimator based on the EKF which is consistent with L^{CDKF} being closer to L^{PF} than L^{EKF}. We also found that a third order approximation to the DSGE model in general gives more efficient QML estimates when compared to a first order approximation.

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:

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

  Updating step:

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

  Quasi log-likelihood function

- - Let $\mathbf{u}_{t+1} \equiv \mathbf{y}_{t+1} - \bar{\mathbf{y}}_{t+1}$ and $\mathbf{F}_{t+1} = \bar{\mathbf{S}}_{\mathbf{y}}(t+1)\bar{\mathbf{S}}_{\mathbf{y}}(t+1)'$
  - $L_{t+1} = L_t - \frac{n_y}{2}\log(2\pi) - \frac{1}{2}\log(|\mathbf{F}_{t+1}|) - \frac{1}{2}\mathbf{u}'_{t+1}\mathbf{F}^{-1}_{t+1}\mathbf{u}_{t+1}$

B The smoother for the CDKF

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

$$\mathbf{x}^{sm}_t = \hat{\mathbf{x}}_t + \mathbf{K}_{t+1}\left[\mathbf{x}^{sm}_{t+1} - \bar{\mathbf{x}}_{t+1}\right]$$
$$\mathbf{P}^{sm}_{\mathbf{xx}}(t) = \hat{\mathbf{P}}_{\mathbf{xx}}(t) + \mathbf{K}_{t+1}\left(\mathbf{P}^{sm}_{\mathbf{xx}}(t+1) - \bar{\mathbf{P}}_{\mathbf{xx}}(t+1)\right)\mathbf{K}'_{t+1}$$

where the smoothing gain is given by

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

and

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

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

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

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

$$\mathbf{S}_{\mathbf{x}}^{s}(t)=\Phi\left(\left[\begin{array}{cccc}\hat{\mathbf{S}}_{\mathbf{x}}(t)-\mathbf{K}_{t+1}\mathbf{S}_{\mathbf{xx}}^{(1)}(t) & \mathbf{K}_{t+1}\mathbf{S}_{\mathbf{x}}^{sm}(t+1) & \mathbf{K}_{t+1}\mathbf{S}_{\mathbf{xw}}^{(1)}(t) & \mathbf{K}_{t+1}\mathbf{S}_{\mathbf{xx}}^{(2)}(t) & \mathbf{K}_{t+1}\mathbf{S}_{\mathbf{xw}}^{(2)}(t)\end{array}\right]\right).$$

This smoothing recursion is started at the last time step, because it holds that $\mathbf{x}_{T}^{sm}=\hat{\mathbf{x}}_{T}$ and $\mathbf{P}_{\mathbf{xx}}^{sm}(T)=\hat{\mathbf{P}}_{\mathbf{xx}}(T)$. Thus, the procedure 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 Accuracy of first and second moments in approximated DSGE models

For first moments it follows trivially that the accuracy of the approximation is given by the approximation order. The situation is different for the second order moments. We illustrate this by considering the scalar case, i.e. x_t and y_t are scalars. Both variables are expressed in deviation from the deterministic steady state. A sixth order approximation to $y_t=g(x_t)$ is given by

$$y_t=g_x x_t+\frac{1}{2!}g_{2x}x_t^2+\frac{1}{3!}g_{3x}x_t^3+\frac{1}{4!}g_{4x}x_t^4+\frac{1}{5!}g_{5x}x_t^5+\frac{1}{6!}g_{6x}x_t^6$$

We denote moments of x_t by $m^i=E\left[x^i\right]$ for $i=1,2,....$ This implies

$$\begin{aligned}Var(y_t)=&\,g_x^2m^2+\left(\tfrac{1}{2!}g_{2x}\right)^2m^4+\left(\tfrac{1}{3!}g_{3x}\right)^3m^6+\left(\tfrac{1}{4!}g_{4x}\right)^4m^8+\left(\tfrac{1}{5!}g_{5x}\right)^5m^{10}+\left(\tfrac{1}{6!}g_{6x}\right)^{12}m^{12}\\&+2\left(\tfrac{1}{2!}g_xg_{2x}m^3+\tfrac{1}{3!}g_xg_{3x}m^4+\tfrac{1}{4!}g_xg_{4x}m^5+\tfrac{1}{5!}g_xg_{5x}m^6+\tfrac{1}{6!}g_xg_{6x}m^7\right)\\&+2\left(\tfrac{1}{2!}\tfrac{1}{3!}g_{2x}g_{3x}m^5+\tfrac{1}{2!}\tfrac{1}{4!}g_{2x}g_{4x}m^6+\tfrac{1}{2!}\tfrac{1}{5!}g_{2x}g_{5x}m^7+\tfrac{1}{2!}\tfrac{1}{6!}g_{2x}g_{6x}m^8\right)\\&+2\left(\tfrac{1}{3!}\tfrac{1}{4!}g_{3x}g_{4x}m^7+\tfrac{1}{3!}\tfrac{1}{5!}g_{3x}g_{5x}m^8+\tfrac{1}{3!}\tfrac{1}{6!}g_{3x}g_{6x}m^9\right)\\&+2\left(\tfrac{1}{4!}\tfrac{1}{5!}g_{4x}g_{5x}m^9+\tfrac{1}{4!}\tfrac{1}{6!}g_{4x}g_{6x}m^{10}\right)\\&+2\left(\tfrac{1}{5!}\tfrac{1}{6!}g_{5x}g_{6x}m^{11}\right)\end{aligned}$$

We therefore have the following:
A second order approximation of $Var(y_t)$ is given by:
$Var(y_t)^{2nd}=g_x^2m^2$

A third order approximation of $Var(y_t)$ is given by:
$Var(y_t)^{3th}=g_x^2m^2+g_xg_{2x}m^3$

A fourth order approximation of $Var(y_t)$ is given by:
$Var(y_t)^{4th}=g_x^2m^2+g_xg_{2x}m^3+\left(\tfrac{1}{2!}g_{2x}\right)^2m^4+\tfrac{2}{3!}g_xg_{3x}m^4$

A fifth order approximation of $Var(y_t)$ is given by:
$Var(y_t)^{5th}=g_x^2m^2+g_xg_{2x}m^3+\left(\tfrac{1}{2!}g_{2x}\right)^2m^4+\tfrac{2}{3!}g_xg_{3x}m^4+\tfrac{2}{4!}g_xg_{4x}m^5+\tfrac{2}{2!}\tfrac{1}{3!}g_{2x}g_{3x}m^5$

Computing the variance in a DSGE model approximated up to first order:
$y_t=g_xx_t$
$\Downarrow$
$Var(y_t)^{app}=g_x^2m^2$
which is accurate up to second order

Computing the variance in a DSGE model approximated up to second order:

$y_t = g_x x_t + \frac{1}{2!} g_{2x} x_t^2$

$\Downarrow$

$Var\left(y_t\right)^{app} = g_x^2 m^2 + \left(\frac{1}{2!} g_{2x}\right)^2 m^4 + g_x g_{2x} m^3$

which is accurate up to third order and not fourth order because we are missing the term $\frac{2}{3!} g_x g_{3x} m^4$

Computing the variance in a DSGE model approximated up to third order:

$y_t = g_x x_t + \frac{1}{2!} g_{2x} x_t^2 + \frac{1}{3!} g_{3x} x_t^3$

$\Downarrow$

$Var\left(y_t\right)^{app} = g_x^2 m^2 + \left(\frac{1}{2!} g_{2x}\right)^2 m^4 + \left(\frac{1}{3!} g_{3x}\right)^2 m^6 + g_x g_{2x} m^3 + \frac{2}{3!} g_x g_{3x} m^4 + \frac{1}{3!} g_{2x} g_{3x} m^5$

which is accurate up to fourth order. Note that we are missing the term $\frac{2}{4!} g_x g_{4x} m^5$ to have precision of fifth order.

Table 1: Shock specifications

Specification	Label	$\epsilon_{e,t+1}$	$\epsilon_{a,t+1}$	$\epsilon_{z,t+1}$	$\epsilon_{R,t+1}$
1	Gaussian shocks	Gaussian	Gaussian	Gaussian	Gaussian
2	Laplace shocks	Laplace	Laplace	Laplace	Laplace
3	Stochastic volatility	Gaussian	Stochastic volatility	Gaussian	None

Table 2: Calibration for the DSGE model

Label	Parameter	Value
Discount factor	β	0.9995
Habit degree	b	0.80
Habit persistence	ρ_x	0.95
Preference	ϕ_1	6
Preference	ϕ_2	0.88
Adj costs for investments	κ	2.0
Depreciation rate	δ	0.025
Cobb-Douglas parameter	θ	0.36
Price elasticity	η	6
Degree of price stickiness	α	0.85
Reaction to lagged interest rate	ρ_R	0.99
Reaction to inflation	β_π	1.65
Reaction to output	β_y	0.15
Inflation rate in steady state	π_{ss}	1.0070
Growth rate in technology shocks	$\mu_{z,ss}$	1.0044
Growth rate in investment shocks	$\mu_{\Upsilon,ss}$	1.0007
Persistency in stationary technology shocks	ρ_a	0.9
Persistency in investment shocks	ρ_e	0.9
Persistency in the volatility process	ρ_{σ_a}	0.9
std. of nonstationary technology shocks	$\sqrt{Var\left(\epsilon_{z,t+1}\right)}$	0.008
std. of stationary technology shocks	$\sqrt{Var\left(\epsilon_{a,t+1}\right)}$	0.012
std. of investment shocks	$\sqrt{Var\left(\epsilon_{e,t+1}\right)}$	0.030
std. of shocks to interest rate rule	$\sqrt{Var\left(\epsilon_{R,t+1}\right)}$	0.001
std. of shocks to volatility	$\sqrt{Var\left(\epsilon_{\sigma_a,t+1}\right)}$	0.300
std. of errors in the interest rate	$\sqrt{Var\left(v_{R,t}\right)}$	0.001
std. of errors in inflation	$\sqrt{Var\left(v_{\pi,t}\right)}$	0.001
std. of errors in the growth rate for consumption	$\sqrt{Var\left(v_{\Delta c,t}\right)}$	0.002
std. of errors in the growth rate for investments	$\sqrt{Var\left(v_{\Delta i,t}\right)}$	0.002
std. of errors in growth rate for output	$\sqrt{Var\left(v_{\Delta y,t}\right)}$	0.002

Table 3: Empirical and simulated moments: Second order approximation

All model moments are calculated based on a simulated time series of 1,000,000 observations.

	R_t	π_t	Δc_t	Δi_t	Δy_t
Empirical moments					
Mean	0.0131	0.0088	0.0055	0.0056	0.0048
Standard deviation	0.0070	0.0063	0.0071	0.0254	0.0092
Skewness	1.0787	0.9700	-0.6719	-1.1384	-0.4525
Kurtosis	4.8934	4.7352	4.9795	6.8528	4.5435
Model moments: Gaussian shocks					
Mean	0.0134	0.0083	0.0048	0.0055	0.0048
Standard deviation	0.0088	0.0063	0.0084	0.0270	0.0118
Skewness	0.5664	1.1100	0.0337	-0.1646	0.0495
Kurtosis	3.9029	5.0441	3.0840	3.0413	3.0406
Model moments: Laplace shocks					
Mean	0.0134	0.0083	0.0048	0.0055	0.0048
Standard deviation	0.0088	0.0062	0.0084	0.0269	0.0118
Skewness	0.5969	1.1653	0.0217	-0.2488	0.0167
Kurtosis	4.1262	5.4804	4.4121	4.1642	3.7135
Model moments: Stochastic volatility					
Mean	0.0134	0.0083	0.0048	0.0055	0.0048
Standard deviation	0.0101	0.0071	0.0086	0.0289	0.0124
Skewness	0.4748	0.9541	0.0311	-0.1359	0.0385
Kurtosis	4.2746	5.273	3.1045	3.1986	3.1465

Table 4: Empirical and simulated moments: Third order approximation

All model moments are calculated based on a simulated time series of 1,000,000 observations.

	R_t	π_t	Δc_t	Δi_t	Δy_t
Empirical moments					
Mean	0.0131	0.0088	0.0055	0.0056	0.0048
Standard deviation	0.0070	0.0063	0.0071	0.0254	0.0092
Skewness	1.0787	0.9700	-0.6719	-1.1384	-0.4525
Kurtosis	4.8934	4.7352	4.9795	6.8528	4.5435
Model moments: Gaussian shocks					
Mean	0.0134	0.0083	0.0048	0.0055	0.0048
Standard deviation	0.0130	0.0107	0.0079	0.0268	0.0116
Skewness	1.3251	1.6429	0.0286	-0.1688	0.0525
Kurtosis	9.4417	11.5575	3.0623	3.0731	3.0228
Model moments: Laplace shocks					
Mean	0.0134	0.0083	0.0048	0.0055	0.0048
Standard deviation	0.0131	0.0108	0.0079	0.0267	0.0115
Skewness	1.4602	1.7684	0.0149	-0.2590	0.0254
Kurtosis	11.7384	14.3213	4.4023	4.3128	3.6517
Model moments: Stochastic volatility					
Mean	0.0134	0.0083	0.0048	0.0055	0.0048
Standard deviation	0.0173	0.0142	0.0084	0.0314	0.0135
Skewness	1.1484	1.4440	0.0270	-0.1780	0.0083
Kurtosis	9.8258	11.7128	3.2898	3.9907	3.7408

Table 5: Average RMSE of state estimates across the 20 test economies

Figures in bold indicate that the filter outperforms the other filters.

	Standard PF (200.000)	EKF	CDKF
Gaussian shocks			
First order	0.0466	**0.0444**	**0.0444**
Second order	0.0507	0.0566	**0.0472**
Third order	0.0517	0.0579	**0.0491**
Laplace shocks			
First order	0.0490	**0.0452**	**0.0452**
Second order	0.0527	0.0554	**0.0488**
Third order	0.0551	0.0577	**0.0513**
Stochastic volatility			
Second order	**0.6854**	1.0251	0.9439
Third order	**0.6287**	1.0049	0.9273

Table 6: Average number of seconds

The results for the EKF and the CDKF are for a square root implementation of these filters on the 20 test economies. All computations are done in Fortran 90 on Dell SC1435 compute-nodes, each with 2 dualcore Opteron 2.6 GHz, 8 GB memory, and 250 GB disk.

	First order	Second order	Third order
Standard PF (200.000)	20.25	68,42	299,41
EKF	0.02	0.10	0.47
CDKF	0.03	0.10	0.50

Table 7: The level of the quasi log-likelihood functions compared to L^{PF}

The average distance is measured by $\sqrt{\frac{1}{20}\sum_{j=1}^{20}\left(L_j^i - \bar{L}_j^{PF}\right)^2}$ for $i = \{\text{EKF,CDKF}\}$. $\bar{L}_j^{PF}$ and $std\left(L_j^{PF}\right)$ denote the mean and standard deviation, respectively, of the log-likelihood function for the j'th test economy across 100 evaluations of the standard PF using 200,000 particles.

	Average distance from L^{PF} for:			
	EKF	CDKF	$\frac{1}{20}\sum_{j=1}^{20}\bar{L}_j^{PF}$	$\frac{1}{20}\sum_{j=1}^{20} std\left(L_j^{PF}\right)$
Gaussian shocks				
First order	2.02	2.02	4055.3	1.96
Second order	5.11	2.29	4054.0	2.01
Third order	5.66	2.39	4063.0	2.01
Laplace shocks				
First order	30.77	30.77	4069.5	11.50
Second order	37.46	35.29	4078.3	10.76
Third order	35.86	33.82	4089.6	10.69
Stochastic volatility				
Second order	60.87	38.73	4043.0	2.93
Third order	95.27	72.53	4026.5	3.74

Table 8: The QML estimator: Gaussian shocks

The results are based on 1000 repetitions in the Monte Carlo study with a sample size of T=200. The true standard errors are calculated as the standard deviation for these estimates. The Type I error is calculated at a 5 percentage significance level. Using a first order approximation, both the EKF and the CDKF reduce to the standard Kalman Filter. Figures in bold indicate that the filter outperforms the other filter. The true values are: $\phi_1 = 6$, $\alpha = 0.85$, $\beta_\pi = 1.65$, $\rho_a = 0.90$, and $\sqrt{Var\left(\epsilon_{z,t}\right)} = 0.0080$.

	EKF				CDKF			
	Bias in level	True SE	Bias in SE	Type I	Bias in level	True SE	Bias in SE	Type I: 5%
First order								
ϕ_1					0.0027	0.4184	-0.0029	0.0500
α					-0.0001	0.0071	-0.0002	0.0580
β_π		same as for CDKF			0.0126	0.0901	-0.0008	0.0500
ρ_a					-0.0008	0.0141	-0.0002	0.0490
$\sqrt{Var\left(\epsilon_{z,t}\right)}$					-0.0000	0.0004	-0.0000	0.0540
Second order								
ϕ_1	0.0476	0.4049	**-0.0005**	0.0460	**-0.0078**	**0.3970**	0.0034	**0.0500**
α	-0.0006	0.0064	-0.0003	0.0700	**-0.0000**	**0.0063**	-0.0003	**0.0690**
β_π	0.0172	0.0945	-0.0045	0.0590	**0.0118**	**0.0910**	**-0.0020**	**0.0460**
ρ_a	**0.0009**	0.0143	-0.0007	0.0700	-0.0010	**0.0142**	**-0.0004**	**0.0600**
$\sqrt{Var\left(\epsilon_{z,t}\right)}$	-0.0000	0.0004	-0.0000	**0.0560**	-0.0000	0.0004	-0.0000	0.0580
Third order								
ϕ_1	**-0.0115**	0.3802	**0.0010**	0.0460	-0.0190	0.3802	-0.0021	0.0540
α	**0.0001**	0.0060	-0.0005	0.0730	-0.0002	**0.0059**	-0.0005	**0.0700**
β_π	0.0183	0.0963	-0.0033	0.0570	**0.0118**	**0.0933**	**-0.0025**	**0.0490**
ρ_a	**0.0000**	0.0127	-0.0004	**0.0610**	-0.0005	0.0127	-0.0004	0.0630
$\sqrt{Var\left(\epsilon_{z,t}\right)}$	-0.0000	0.0004	-0.0000	**0.0550**	-0.0000	0.0004	-0.0000	0.0570

Table 9: The QML estimator: Laplace distributed shocks

The results are based on 1000 repetitions in the Monte Carlo study with a sample size of T=200. The true standard errors are calculated as the standard deviation for these estimates. The Type I error is calculated at a 5 percentage significance level. Using a first order approximation, both the EKF and the CDKF reduce to the standard Kalman Filter. Figures in bold indicate that the filter outperforms the other filter. The true values are: $\phi_1 = 6$, $\alpha = 0.85$, $\beta_\pi = 1.65$, $\rho_a = 0.90$, and $\sqrt{Var\left(\epsilon_{z,t}\right)} = 0.0080$.

	EKF				CDKF			
	Bias in level	True SE	Bias in SE	Type I	Bias in level	True SE	Bias in SE	Type I: 5%
First order								
ϕ_1					0.0318	0.4239	-0.0076	0.0610
α					0.0003	0.0077	-0.0006	0.0870
β_π		same as for CDKF			0.0139	0.0955	-0.0083	0.0740
ρ_a					-0.0003	0.0157	-0.0016	0.0780
$\sqrt{Var\left(\epsilon_{z,t}\right)}$					0.0006	0.0006	-0.0001	0.1830
Second order								
ϕ_1	0.0981	0.4448	-0.0300	0.0750	**0.0382**	**0.4406**	**-0.0289**	**0.0730**
α	**0.0004**	0.0072	-0.0007	0.0990	0.0008	**0.0070**	-0.0007	**0.0970**
β_π	0.0199	0.1023	-0.0110	0.0800	**0.0165**	**0.0993**	**-0.0090**	0.0800
ρ_a	0.0027	0.0151	-0.0018	0.0990	**0.0010**	**0.0150**	**-0.0016**	**0.0840**
$\sqrt{Var\left(\epsilon_{z,t}\right)}$	0.0006	0.0006	-0.0000	0.1810	0.0006	0.0006	-0.0000	**0.1770**
Third order								
ϕ_1	**0.0068**	**0.4085**	**-0.0257**	**0.0690**	-0.0150	0.4094	-0.0283	0.0730
α	0.0013	0.0063	-0.0006	0.1020	**0.0008**	**0.0061**	**-0.0005**	**0.0880**
β_π	0.0237	0.1031	-0.0081	**0.0700**	**0.0167**	**0.1004**	**-0.0078**	0.0720
ρ_a	0.0004	**0.0132**	**-0.0013**	**0.0800**	**0.0000**	0.0133	-0.0014	0.0840
$\sqrt{Var\left(\epsilon_{z,t}\right)}$	0.0007	0.0006	-0.0001	0.1820	**0.0006**	0.0006	-0.0001	**0.1770**

Table 10: The QML estimator: Stochastic volatility

The results are based on 1000 repetitions in the Monte Carlo study with a sample size of T=200. The true standard errors are calculated as the standard deviation for these estimates. The Type I error is calculated at a 5 percentage significance level. Using a first order approximation, both the EKF and the CDKF reduce to the standard Kalman Filter. Figures in bold indicate that the filter outperforms the other filter. The true values are: $\phi_1 = 6$, $\alpha = 0.85$, $\beta_\pi = 1.65$, $\rho_a = 0.90$, and $\sqrt{Var\left(\epsilon_{z,t}\right)} = 0.0080$.

	EKF				CDKF			
	Bias in level	True SE	Bias in SE	Type I	Bias in level	True SE	Bias in SE	Type I: 5%
Second order								
ϕ_1	0.0437	0.3862	-0.0024	**0.0500**	**-0.0067**	**0.3811**	**-0.0005**	0.0550
α	-0.0006	0.0063	-0.0004	**0.0700**	**-0.0001**	**0.0062**	-0.0004	0.0730
β_π	0.0167	0.0887	-0.0034	0.0600	**0.0121**	**0.0853**	**-0.0009**	**0.0580**
ρ_a	0.0010	0.0133	-0.0009	0.0780	**-0.0008**	**0.0132**	**-0.0007**	**0.0720**
$\sqrt{Var\left(\epsilon_{z,t}\right)}$	0.0000	0.0005	-0.0000	0.0660	0.0000	0.0005	-0.0000	**0.0620**
Third order								
ϕ_1	**-0.0151**	0.3601	**-0.0139**	**0.0580**	-0.0224	**0.3597**	-0.0168	0.0710
α	0.0002	0.0058	-0.0004	0.0790	**-0.0001**	**0.0056**	**-0.0003**	**0.0630**
β_π	0.0194	0.0904	-0.0018	**0.0620**	**0.0124**	**0.0877**	**-0.0013**	0.0700
ρ_a	**-0.0001**	0.0118	-0.0006	**0.0670**	-0.0007	0.0118	-0.0006	0.0680
$\sqrt{Var\left(\epsilon_{z,t}\right)}$	0.0000	0.0005	-0.0000	0.0690	0.0000	0.0005	-0.0000	**0.0640**

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

Andreasen, M. M. (2010*a*), 'How non-gaussian shocks affect risk premia in non-linear DSGE models', *Working Paper* .

Andreasen, M. M. (2010*b*), 'How to maximize the likelihood function for a DSGE model', *Computational Economics* **35**, 127–154.

Andreasen, M. M. (2010*c*), 'Properties of the pruning scheme for non-linear approximated DSGE models', *Working Paper* .

Andreasen, M. M. (2010*d*), 'Stochastic volatility and DSGE models', *Economics Letters* **108**, 7–9.

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.

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

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

Fernandez-Villaverde, J., Rubio-Ramirez, J. F. & Santos, M. S. (2006), 'Convergence properties of the likelihood of computed dynamic models', *Econometrica* **74**(1), 93–119.

Flury, T. & Shephard, N. (2009), 'Learning and filtering via simulation: Smoothly jittered particle filters', *Working Paper* .

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

Hansen, N., Müller, S. D. & Koumoutsakos, P. (2003), 'Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES)', *Evolutionary Computation* **11**(1), 1–18.

Harvey, A. C. (1989), 'Forecasting, structural time series models and the kalman filter', *Cambridge University Press* .

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* **98**(3), 604–641.

Kim, J., Kim, S., Schaumburg, E. & Sims, C. A. (2008), 'Calculating and using second-order accurate solutions of discrete time dynamic equilibrium models', *Journal of Economic Dynamics and Control* **32**, 3397–3414.

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

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

Pitt, M. K. (2002), 'Smooth particle filters for likelihood evaluation and maximisation', *Warwick Economic Research Papers, no. 651* .

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.

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

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: RMSE for the state vector with Gaussian shocks

200,000 particles are used in the standard PF. The 95 pct. confidence intercal (CI) for the standard PF is computed from 100 repetitions of the filter with different random numbers on the same test economy.

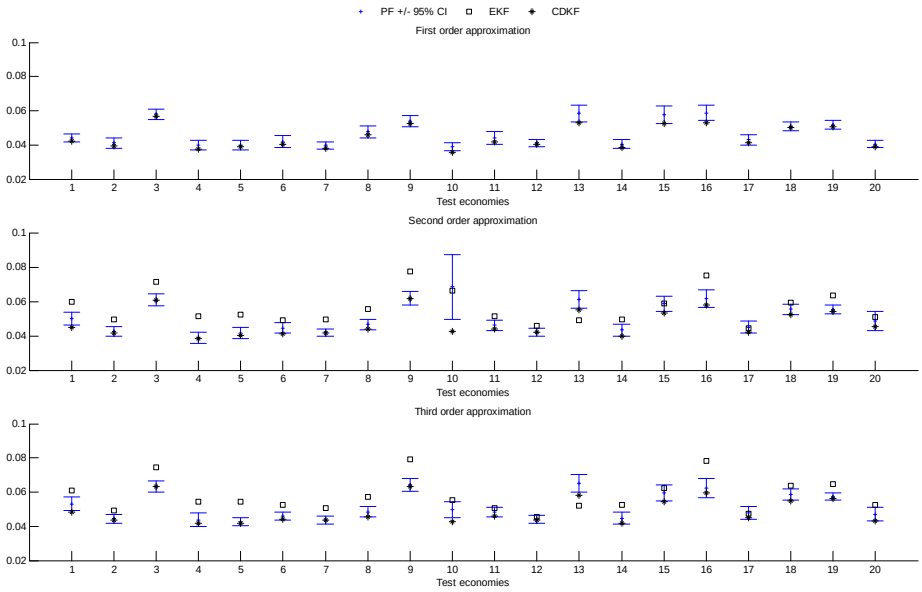


Figure 2: RMSE for the state vector with Laplace distributed shocks

200,000 particles are used in the standard PF. The 95 pct. confidence intercal (CI) for the standard PF is computed from 100 repetitions of the filter with different random numbers on the same test economy.

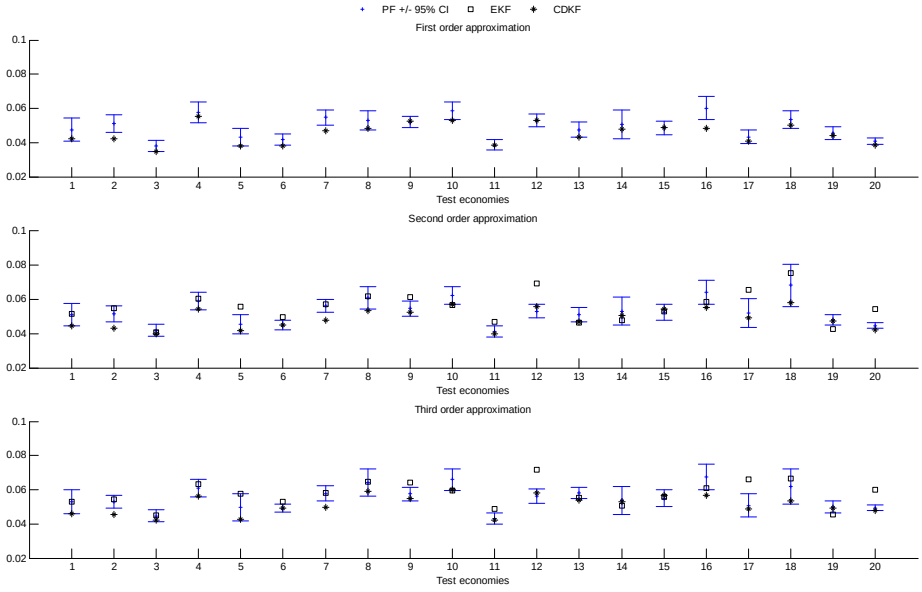


Figure 3: RMSE for the state vector with stochastic volatility

200,000 particles are used in the standard PF. The 95 pct. confidence intercal (CI) for the standard PF is computed from 100 repetitions of the filter with different random numbers on the same test economy.

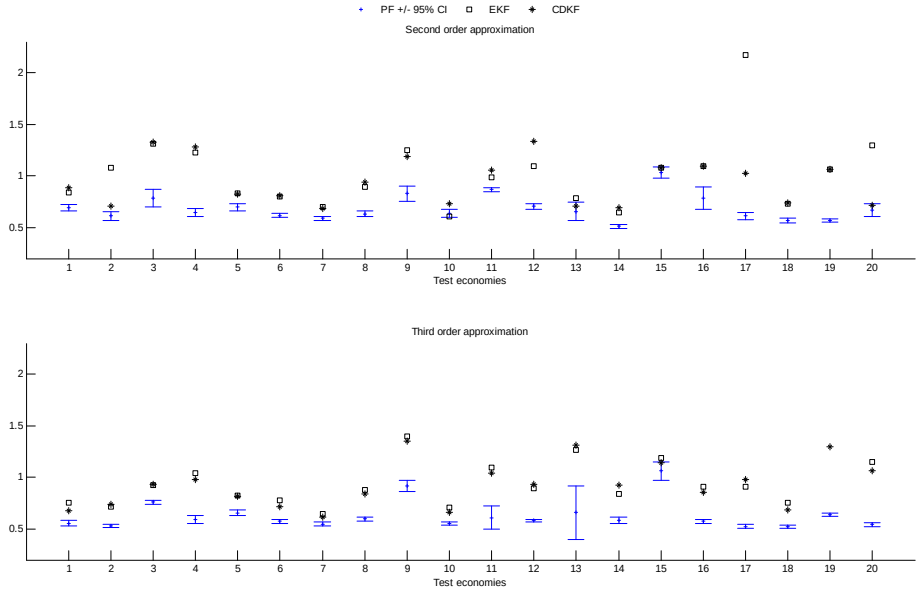


Figure 4: Quasi log-likelihood functions: Gaussian shocks

The value of the quasi log-likelihood function is expressed in percentage deviation from the estimated mean value of the log-likelihood function in the standard PF using 200,000 particles. 100 repetitions of the standard PF with different random numbers are used to compute the reported 95 percentage confidence interval for each test economy.

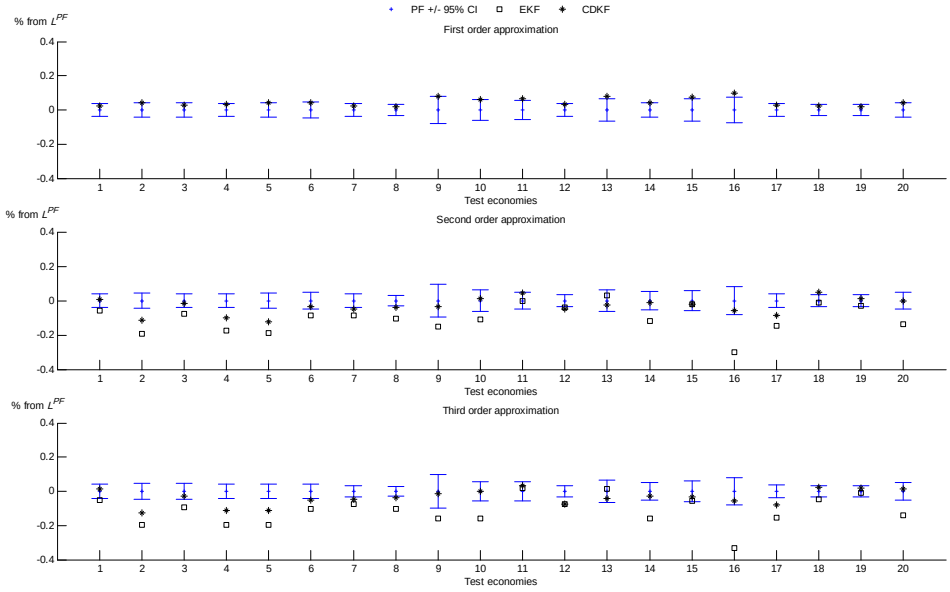


Figure 5: Quasi log-likelihood functions: Laplace distributed shocks

The value of the quasi log-likelihood function is expressed in percentage deviation from the estimated mean value of the log-likelihood function in the standard PF using 200,000 particles. 100 repetitions of the standard PF with different random numbers are used to compute the reported 95 percentage confidence interval for each test economy.

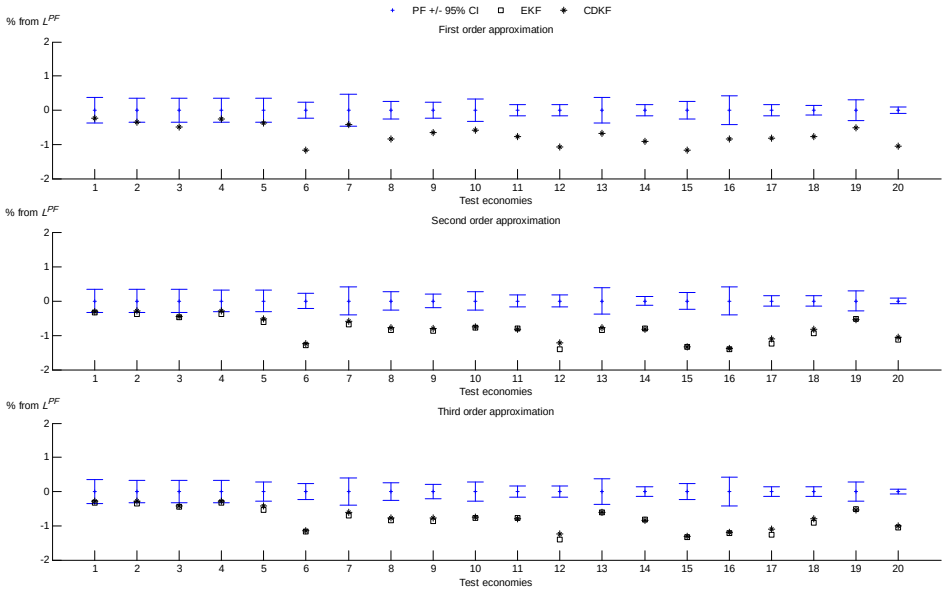
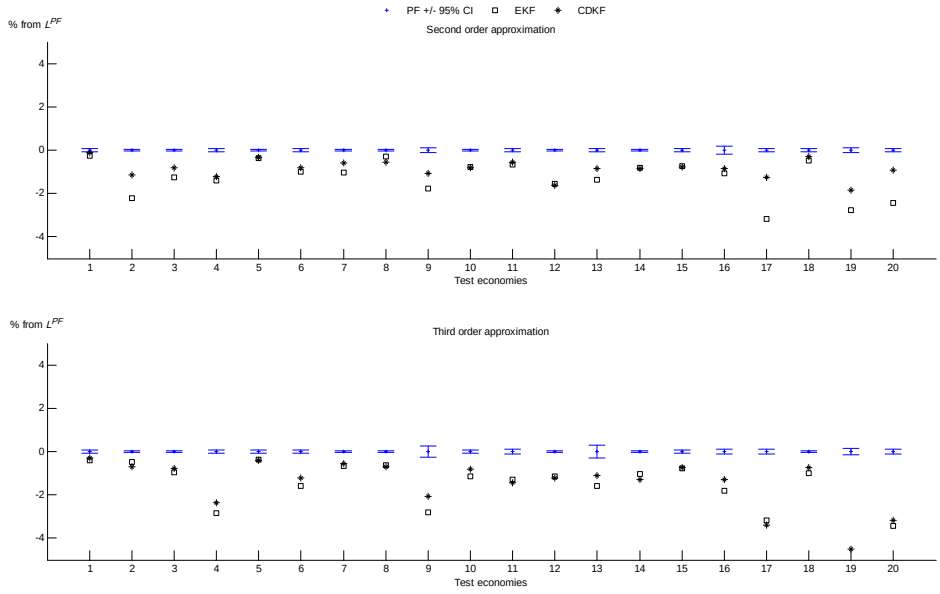


Figure 6: Quasi log-likelihood functions: Stochastic volatility

The value of the quasi log-likelihood function is expressed in percentage deviation from the estimated mean value of the log-likelihood function in the standard PF using 200,000 particles. 100 repetitions of the standard PF with different random numbers are used to compute the reported 95 percentage confidence interval for each test economy.



Research Papers 2010

2010-16: Tim Bollerslev and Viktor Todorov: Estimation of Jump Tails

2010-17: Ole E. Barndorff–Nielsen, Fred Espen Benth and Almut E. D. Veraart: Ambit processes and stochastic partial differential equations

2010-18: Ole E. Barndorff–Nielsen, Fred Espen Benth and Almut E. D. Veraart: Modelling energy spot prices by Lévy semistationary processes

2010-19: Jeroen V.K. Rombouts and Lars Stentoft: Multivariate Option Pricing with Time Varying Volatility and Correlations

2010-20: Charlotte Christiansen: Intertemporal Risk-Return Trade-off in

2010-21: Marco Aiolfi, Carlos Capistrán and Allan Timmermann : Forecast Combinations

2010-22: Ivan Nourdin, Giovanni Peccati and Mark Podolskij: Quantitative Breuer-Major Theorems

2010-23: Matias D. Cattaneo, Richard K. Crump and Michael Jansson: Bootstrapping Density-Weighted Average Derivatives

2010-24: Søren Johansen and Morten Ørregaard Nielsen: Likelihood inference for a fractionally cointegrated vector autoregressive model

2010-25: Tom Engsted and Bent Nielsen: Testing for rational bubbles in a co-explosive vector autoregression

2010-26: Robinson Kruse: On European monetary integration and the persistence of real effective exchange rates

2010-27: Sanne Hiller and Robinson Kruse: Milestones of European Integration: Which matters most for Export Openness?

2010-28: Robinson Kruse: Forecasting autoregressive time series under changing persistence

2010-29: Nikolaus Hautsch and Mark Podolskij: Pre-Averaging Based Estimation of Quadratic Variation in the Presence of Noise and Jumps: Theory, Implementation, and Empirical Evidence

2010-30: Martin M. Andreasen: Non-linear DSGE Models and The Central Difference Kalman Filter