Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood*

Jesús Fernández-Villaverde — University of Pennsylvania

Juan F. Rubio-Ramírez — Federal Reserve Bank of Atlanta

June 10, 2004

Abstract

This paper compares two methods for undertaking likelihood-based inference in dynamic equilibrium economies: a Sequential Monte Carlo filter and the Kalman filter. The Sequential Monte Carlo filter exploits the nonlinear structure of the economy and evaluates the likelihood function of the model by simulation methods. The Kalman filter estimates a linearization of the economy around the steady state. We report two main results. First, both for simulated and for real data, the Sequential Monte Carlo filter delivers a substantially better fit of the model to the data as measured by the marginal likelihood. This is true even for a nearly linear case. Second, the differences in terms of point estimates, although relatively small in absolute values, have important effects on the moments of the model. We conclude that the nonlinear filter is a superior procedure for taking models to the data.

*Corresponding author: Jesús Fernández-Villaverde, Department of Economics, 160 McNeil Building, 3718 Locust Walk, University of Pennsylvania, Philadelphia PA 19104. E-mail: jesusfv@econ.upenn.edu. We thank Sally Burke, Will Roberds, Chris Sims, Tao Zha, participants at several seminars, the editor, and two anonymous referees for useful comments. Valuable assistance was provided by the University of Minnesota Supercomputer Institute. Beyond the usual disclaimer, we must note that any views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta or of the Federal Reserve System.

1. Introduction

Recently, a growing literature has focused on the formulation and estimation of dynamic equilibrium models using a likelihood-based approach. Examples include the seminal paper of Sargent (1989), and, more recently, Bouakez, Cardia, and Ruge-Murcia (2002), DeJong, Ingram, and Whiteman (2000), Dib (2001), Fernández-Villaverde and Rubio-Ramírez (2003), Hall (1996), Ireland (2002), Kim (2000), Landon-Lane (1999), Lubik and Schorfheide (2003), McGrattan, Rogerson, and Wright (1997), Moran and Dolar (2002), Otrok (2001), Rabanal and Rubio-Ramírez (2003), Schorfheide (2000), and Smets and Wouters (2003), to name just a few. Most of these papers have used the Kalman filter to estimate a linear approximation to the original model.

This paper studies the effects of estimating the nonlinear representation of a dynamic equilibrium model instead of working with its linearized version. We document how the estimation of the nonlinear solution of the economy substantially improves the empirical fitting of the model. The marginal likelihood of the economy, i.e., the probability that the model assigns to the data, increases by two orders of magnitude. This is true even for our application, the neoclassical growth model, which is nearly linear. We also report that, although the effect of linearization on point estimates is small, the impact on the moments of the model is of first-order importance. This finding is key for applied economist because quantitative models are widely judged by their ability to match data moments.

Dynamic equilibrium models have become a standard tool in quantitative economics (see Cooley, 1995, for a summary of applications). These models can be described as a likelihood function for observables, given the model's structural parameters- those characterizing preferences and technology. The advantage of thinking about models as a likelihood function is that, once we can evaluate the likelihood, inference is a direct exercise. In a classical environment we maximize this likelihood function to get point estimates and standard errors. A Bayesian researcher can use the likelihood and her priors about the parameters to find the posterior. The advent of Markov chain Monte Carlo algorithms has facilitated this task. In addition, we can compare models by likelihood ratios (Vuong, 1989) or Bayes factors (Geweke, 1998) even if the models are misspecified and nonnested.

The previous discussion points out the need to evaluate the likelihood function. The task is conceptually simple, but its implementation is more cumbersome. Dynamic equilibrium economies do not have a "paper and pencil" solution. This means that we can study only an approximation to them, usually generated by a computer. The lack of a closed form for the solution of the model complicates the process of finding the likelihood.

The literature shows how to write this likelihood analytically only in a few cases (see Rust, 1994, for a survey). Outside those, Sargent (1989) proposed an approach that has

become popular. Sargent noticed that a standard procedure for solving dynamic models is to linearize them. This can be done either directly in the equilibrium conditions or by generating a quadratic approximation to the utility function of the agents. Both approaches imply that the optimal decision rules are linear in the states of the economy. The resulting linear system of difference equations can be solved with standard methods (see Anderson *et al.*, 1996, and Uhlig, 1999, for a detailed explanation).

For estimation purposes, Sargent emphasized that the resulting system has a linear representation in a state-space form. If, in addition, we assume that the shocks exogenously hitting the economy are normal, we can use the Kalman filter to evaluate the likelihood. It has been argued (for example, Kim *et al.*, 2003) that this linear solution is likely to be accurate enough for fitting the model to the data.

However, exploiting the linear approximation to the economy can be misleading. For instance, linearization may be an inaccurate approximation if the nonlinearities of the model are important or if we are traveling far away from the steady state of the model. Also, accuracy in terms of the policy function of the model does not necessarily imply accuracy in terms of the likelihood function. Finally, the assumption of normal innovations may be a poor representation of the dynamics of the shocks in the data.

An alternative to linearization is to work instead with the nonlinear representation of the model and to apply a nonlinear filter to evaluate the likelihood. This is possible thanks to the recent development of Sequential Monte Carlo methods (see the seminal paper of Gordon, Salmond, and Smith, 1993, and the review of the literature by Doucet, de Freitas and Gordon, 2000, for extensive references). Fernández-Villaverde and Rubio-Ramírez (2004) build on this literature to show how a Sequential Monte Carlo filter delivers a consistent and efficient evaluation of the likelihood function of a nonlinear and/or nonnormal dynamic equilibrium model.

The presence of the two alternative filters begets the following question: how different are the answers provided by each of them? We study this question with the canonical stochastic neoclassical growth model with leisure choice. We estimate the model using both simulated and real data and compare the results obtained with the Sequential Monte Carlo filter and the Kalman filter.

Why do we choose the neoclassical growth model for our comparison? First, this model is the workhorse of modern macroeconomics. Since any lesson learned in this paper is conditional on our particular model, we want to select an economy that is the foundation of numerous applications. Second, even if the model is nearly linear for the standard calibration, the answers provided by each of the filters are nevertheless quite different. In this way, we make our point that linearization has a nontrivial impact on estimation in the simplest possible environment.

Our main finding is that, while linearization may have a second-order effect on the accuracy of the policy function given some parameter values, it has a first-order impact on the model's likelihood function. Both for simulated and for real data, the Sequential Monte Carlo filter generates an overwhelmingly better fit of the model as measured by the marginal likelihood. This is true even if most differences in the point estimates of the parameters generated by each of the two filters are small.

Why is the marginal likelihood so much higher for the Sequential Monte Carlo? Because this filter delivers point estimates for the parameters that imply model's moments which are closer to the moments of the data. This result is crucial in applied work because models are often judged by their ability to match empirical moments.

Our finding is not the first in the literature that suggests that accounting for nonlinearities substantially improves the measures of fit of a model. For example, Sims and Zha (2002) report that the ability of a structural VAR to account for the dynamics of the output and monetary policy increases by several orders of magnitude when they allow the structural equation variances to change over time. A similar finding is emphasized by the literature on regime switching (Kim and Nelson, 1999) and by the literature on the asymmetries of the business cycle (Kim and Piger, 2002).

The rest of the paper is organized as follows. In section 2 we discuss the two alternatives to evaluate the likelihood of a dynamic equilibrium economy. Section 3 presents our application. Section 4 describes the estimation algorithm, and section 5 reports our main findings. Section 6 concludes.

2. Two Frameworks to Evaluate the Likelihood

In this section we describe the nonlinear and the linear filters used to evaluate the likelihood function of a dynamic equilibrium economy. First, we present the state-space representation of a dynamic equilibrium model solved by nonlinear and linear methods. Second, we show how to use a Sequential Monte Carlo filter to evaluate the likelihood of the nonlinear state-space representation of the economy. Finally, we do the same with the Kalman filter.

2.1. The State-Space Representation

Assume that we observe $y^T = \{y_t\}_{t=1}^T$, a realization of the $n-$dimensional random variable $Y^T = \{Y_t\}_{t=1}^T \in R^{nT}$. The researcher is interested in evaluating the likelihood function of the observable y^T implied by a dynamic equilibrium economy at any given γ, $L\left(y^T;\gamma\right) = p\left(y^T;\gamma\right)$, where $\gamma \in \Upsilon$ is the vector collecting the parameters that characterize preferences, information, and technology in the model.

Unfortunately, in general, it is not possible to compute this function. Part of the reason is that most dynamic models do not have a closed-form solution. Consequently, just to solve the model before any estimation, we need to approximate the equilibrium path using numerical techniques. This approximation is going to affect the characterization of the likelihood.

There are two main routes to evaluate the likelihood function. If we opt to solve the model nonlinearly, we can use a Sequential Monte Carlo. If we linearize the model, we can apply the Kalman filter. We now describe both methodologies.

2.1.1. The Nonlinear Solution of the Model

Dynamic equilibrium economies solved using nonlinear methods have the following state-space representation. The vector of state variables, S_t, evolves over time according to the transition equation:

$$S_t = f\left(S_{t-1}, W_t; \gamma\right) \tag{1}$$

where $\{W_t\}$ is a sequence of exogenous random variables. The observable y_t is governed by the measurement equation:

$$Y_t = g\left(S_t, V_t; \gamma\right) \tag{2}$$

where $\{V_t\}$ is a sequence of exogenous independent random variables. The sequences $\{W_t\}$ and $\{V_t\}$ are independent of each other. Assuming independence of $\{W_t\}$ and $\{V_t\}$ is only for notational convenience. Generalization to more involved structures is achieved by increasing the dimension of the state space. Along some dimension, the function g can be the identity mapping if a state is observed without noise.

The functions f and g depend on the equations that describe the equilibrium of the model - policy functions, laws of motion for variables, resource constraints - and on the nonlinear solution method used to approximate the policy functions.

To ensure that the model is not stochastically singular, we assume that $\dim\left(W_t\right) + \dim\left(V_t\right) \geq \dim\left(Y_t\right)$. We do not impose any restrictions on how those degrees of stochasticity are achieved.

2.1.2. The Linear Solution of the Model

If we pick a linear method to solve the same model, the state-space representation has the following linear form:

$$S_t = E\left(\gamma\right) + A\left(\gamma\right) S_{t-1} + B\left(\gamma\right) W_t \tag{3}$$

$$Y_t = F\left(\gamma\right) + C\left(\gamma\right) S_t + D\left(\gamma\right) V_t \tag{4}$$

where $A(\gamma)$, $B(\gamma)$, $C(\gamma)$, $D(\gamma)$, $E(\gamma)$, and $F(\gamma)$ are matrices with the required dimension that depend on the parameters of the model. Notice how these equations are nothing more than a particular case of (1) and (2). Also, we make the same assumptions regarding stochastic singularity as above.

We have presented two representations of the same economy. Section 2.2 introduces a Sequential Monte Carlo filter to evaluate the likelihood function implied by (1) and (2). Section 2.3 exploits the Kalman filter to calculate the likelihood entailed by (3) and (4).

2.2. The Nonlinear Approach: A Sequential Monte Carlo Filter

Fernández-Villaverde and Rubio-Ramírez (2004) propose the following Sequential Monte Carlo to evaluate the likelihood function of y^T induced by (1) and (2).

We assume that we can partition $\{W_t\}$ into two separate sequences $\{W_{1,t}\}$ and $\{W_{2,t}\}$, such that $W_t = (W_{1,t}, W_{2,t})$ and $\dim(W_{2,t}) + \dim(V_t) = \dim(Y_t)$. Let $W_i^t = \{W_{i,m}\}_{m=1}^t$, for $i = 1, 2$, $V^t = \{V_m\}_{m=1}^t$, and $S^t = \{S_m\}_{m=0}^t$ for $\forall t$. We also define $W_i^0 = \{\emptyset\}$ and $y^0 = \{\emptyset\}$. We could work under weaker assumptions, paying the cost of heavier notation.

Given our assumptions, we factor the likelihood function as follows:

$$p\left(y^T;\gamma\right) = \prod_{t=1}^{T} p\left(y_t|y^{t-1};\gamma\right)$$

$$= \prod_{t=1}^{T} \int\int p\left(y_t|W_1^t, S_0, y^{t-1};\gamma\right) p\left(W_1^t, S_0|y^{t-1};\gamma\right) dW_1^t dS_0. \tag{5}$$

Conditional on having N draws of $\left\{\left\{s_0^{t|t-1,i}, w_1^{t|t-1,i}\right\}_{i=1}^N\right\}_{t=1}^T$ from the sequence of densities $\{p(W_1^t, S_0|y^{t-1};\gamma)\}_{t=1}^T$, this likelihood can be approximated by:

$$p\left(y^T;\gamma\right) \simeq p_{SMC}\left(y^T;\gamma\right) = \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} p\left(y_t|w_1^{t|t-1,i}, s_0^{t|t-1,i}, y^{t-1};\gamma\right) \tag{6}$$

using a law of large numbers. Thus, the problem of evaluating $p_{SMC}\left(y^T;\gamma\right)$ is equivalent to the problem of drawing from $\{p(W_1^t, S_0|y^{t-1};\gamma)\}_{t=1}^T$. The Sequential Monte Carlo filter accomplishes this objective.

Let us fix some additional notation. Let $\left\{s_0^{t,i}, w_1^{t,i}\right\}_{i=1}^N$ be a sequence of N i.i.d. draws from $p(W_1^t, S_0|y^t;\gamma)$ and $\left\{s_0^{t|t-1,i}, w_1^{t|t-1,i}\right\}_{i=1}^N$ be a sequence of N i.i.d. draws from $p(W_1^t, S_0|y^{t-1};\gamma)$. We call each draw $\left(s_0^{t,i}, w_1^{t,i}\right)$ a particle and the whole sequence $\left\{s_0^{t,i}, w_1^{t,i}\right\}_{i=1}^N$ a swarm of particles.

Fernández-Villaverde and Rubio-Ramírez (2004) prove the following result that shows

how to use $p\left(W_1^t, S_0|y^{t-1};\gamma\right)$ as an important sampling density to draw from $p\left(W_1^t, S_0|y^t;\gamma\right)$.

Proposition 1. *Let* $\left\{s_0^{t|t-1,i}, w_1^{t|t-1,i}\right\}_{i=1}^N$ *be a draw from* $p\left(W_1^t, S_0|y^{t-1};\gamma\right)$. *Let the sequence* $\{\widetilde{s}_0^i, \widetilde{w}_1^i\}_{i=1}^N$ *be a draw with replacement from* $\left\{s_0^{t|t-1,i}, w_1^{t|t-1,i}\right\}_{i=1}^N$ *where* q_t^i, *defined as*

$$q_t^i = \frac{p\left(y_t|w_1^{t|t-1,i}, s_0^{t|t-1,i}, y^{t-1};\gamma\right)}{\sum_{i=1}^N p\left(y_t|w_1^{t|t-1,i}, s_0^{t|t-1,i}, y^{t-1};\gamma\right)},$$

is the probability of $\left(s_0^{t|t-1,i}, w_1^{t|t-1,i}\right)$ *being drawn* $\forall i$. *Then* $\{\widetilde{s}_0^i, \widetilde{w}_1^i\}_{i=1}^N$ *is a draw from* $p\left(W_1^t, S_0|y^t;\gamma\right)$.

The proposition 1 shows how a draw $\left\{s_0^{t|t-1,i}, w_1^{t|t-1,i}\right\}_{i=1}^N$ from $p\left(W_1^t, S_0|y^{t-1};\gamma\right)$ can be used to get a draw $\left\{s_0^{t,i}, w_1^{t,i}\right\}_{i=1}^N$ from $p\left(W_1^t, S_0|y^t;\gamma\right)$. This result is crucial. Given particles at $t-1$, $\left\{s_0^{t-1,i}, w_1^{t-1,i}\right\}_{i=1}^N$ distributed according to $p\left(W_1^{t-1}, S_0|y^{t-1};\gamma\right)$, we can use $p\left(W_{1,t};\gamma\right)$ to generate proposal draws $\left\{s_0^{t|t-1,i}, w_1^{t|t-1,i}\right\}_{i=1}^N$ from $p\left(W_1^t, S_0|y^{t-1};\gamma\right)$. Then, we can exploit of proposition 1 and resample from $\left\{s_0^{t|t-1,i}, w_1^{t|t-1,i}\right\}_{i=1}^N$ a new swarm of particles, $\left\{s_0^{t,i}, w_1^{t,i}\right\}_{i=1}^N$distributed according to $p\left(W_1^t, S_0|y^t;\gamma\right)$. The output of the algorithm $\left\{\left\{s_0^{t|t-1,i}, w_1^{t|t-1,i}\right\}_{i=1}^N\right\}_{t=1}^T$ is used to compute the likelihood (6).

This is the key step in the filter. A naive extension of Monte Carlo techniques diverges as T grows because only one particle will eventually accumulate all the information. To avoid this problem, we do not carry over all the simulations to the next period. We keep those with higher probability of explaining the data.

The following pseudocode summarizes the description of the algorithm:

Step 0, Initialization: `Set` $t \rightsquigarrow 1$. `Initialize` $p\left(W_1^{t-1}, S_0|y^{t-1};\gamma\right) = p\left(S_0;\gamma\right)$.

Step 1, Prediction: `Sample` N `values` $\left\{s_0^{t|t-1,i}, w_1^{t|t-1,i}\right\}_{i=1}^N$ `from the conditional density` $p\left(W_1^t, S_0|y^{t-1};\gamma\right) = p\left(W_{1,t};\gamma\right) p\left(W_1^{t-1}, S_0|y^{t-1};\gamma\right)$.

Step 2, Filtering: `Assign to each draw` $\left(s_0^{t|t-1,i}, w_1^{t|t-1,i}\right)$ `the weight` q_t^i `as defined in proposition 1.`

Step 3, Sampling: `Sample` N `times with replacement from` $\left\{s_0^{t|t-1,i}, w_1^{t|t-1,i}\right\}_{i=1}^N$ `with probabilities` $\{q_t^i\}_{i=1}^N$. `Call each draw` $\left(s_0^{t,i}, w_1^{t,i}\right)$. `If` $t < T$ `set` $t \rightsquigarrow t+1$ `and go to step 1. Otherwise stop.`

An important point is that the algorithm can be implemented on a good desktop computer. All programs needed for the computation of the model and the Sequential Monte Carlo were coded in Fortran 95 and compiled in Intel Fortran 8.0 to run on Windows-based machines. On a Pentium 4 at 3.00 GHz, each draw from the posterior using the Sequential Monte Carlo with 60,000 particles takes around 6.1 seconds. That implies a total of about 88 hours for each simulation of 50,000 draws. This relatively low computational cost opens the nonlinear estimation of dynamic equilibrium models to practitioners, and should be seen as an important selling point of our procedure.[1]

It is important to note that we are presenting here only a basic Sequential Monte Carlo filter and that the literature has presented several refinements to improve efficiency (see, for example, Kitigawa, 1996, and Pitts and Shephard, 1999). The interested reader can find further details, comparison with alternative schemes, and a discussion of convergence in Fernández-Villaverde and Rubio-Ramírez (2004).

2.3. The Linear Approach: The Kalman Filter

Now we describe how to evaluate the likelihood function implied by (3) and (4) with the Kalman filter.

To apply this filter, we need to assume that $\{W_t\}$ and $\{V_t\}$ are both normally distributed. Therefore, we can define $\widetilde{W}_t = B(\gamma) W_t$ and $\widetilde{V}_t = D(\gamma) V_t$ to be normal with distributions $\widetilde{W}_t \sim N(0, Q(\gamma))$ and $\widetilde{V}_t \sim N(0, R(\gamma))$.

The formulae of the Kalman filter allow us to recursively build a linear forecast of y_t, called $y_{t|t-1}$, based on previous observations. Also, it implies that the difference between y_t and its prediction $y_{t|t-1}$ is normally distributed with zero mean and variance $\Sigma_{t|t-1}$. In the interest of space we refer the reader to Harvey (1989) for details regarding how to compute $y_{t|t-1}$ and $\Sigma_{t|t-1}$. Suffice it to say that we need to keep track of the linear prediction of s_t (called $s_{t|t-1}$) and its variance $P_{t|t-1}$. The output of the Kalman filter can be used to calculate the likelihood function as follows:

$$p\left(y^T;\gamma\right) \simeq p_{KF}\left(y^T;\gamma\right) = \prod_{t=1}^{T} \frac{1}{\sqrt{(2\pi)^n \left|\Sigma_{t|t-1}\right|}} \exp\left(-\frac{\left(y_t - y_{t|t-1}\right)' \Sigma_{t|t-1}^{-1} \left(y_t - y_{t|t-1}\right)}{2}\right) \quad (7)$$

We can also check that, if the transition and measurement equations of the model are linear and the innovations normally distributed, the Sequential Monte Carlo and the Kalman filter deliver *precisely* the same value for the likelihood function (up to a numerical error).

[1] All of the code is available upon request from the corresponding author.

3. An Application

This section presents an example of how to implement the two alternatives described above to evaluate the likelihood function of a dynamic equilibrium model. We select the neoclassical growth model as our application. The reasons are twofold. First, this environment is the workhorse of quantitative macroeconomics. In this way, we perform our comparison in an application that is "representative" of a large number of papers. Since any lesson learned is conditional on our particular model, we want to deal with a case that can be partially extrapolated to other setups. Second, the application of the two procedures delivers answers that are substantially different even if the model is nearly linear. The neoclassical growth model is a simple environment where we can make our main point. For a more nonlinear model, the disparities are more striking.

The rest of the section is organized as follows. First, we introduce the neoclassical growth model. Second, we discuss our linear and nonlinear approaches to solution methods. Third, we compute $p_{SMC}(y^T;\gamma)$ and $p_{KF}(y^T;\gamma)$.

3.1. The Neoclassical Growth Model

We work with the stochastic neoclassical growth model with leisure. Since this model is well known (see Cooley and Prescott, 1995), we present it only to fix notation.

There is a representative agent in the economy, whose preferences about consumption c_t and leisure l_t are represented by the utility function

$$U = E_0 \sum_{t=1}^{\infty} \beta^{t-1} \frac{\left(c_t^{\theta} (1-l_t)^{1-\theta}\right)^{1-\tau}}{1-\tau}$$

where $\beta \in (0,1)$ is the discount factor, τ controls the elasticity of intertemporal substitution, θ pins down labor supply, and E_0 is the conditional expectation operator.

The only good of this economy is produced according to the production function $e^{z_t} k_t^{\alpha} l_t^{1-\alpha}$ where k_t is the aggregate capital stock, l_t is the aggregate labor input, and z_t is the technology level. z_t follows an AR(1) $z_t = \rho z_{t-1} + \epsilon_t$ with $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon)$. We consider the stationary case (i.e., $|\rho| < 1$). The law of motion for capital is $k_{t+1} = i_t + (1-\delta)k_t$ where i_t is investment. Finally, the economy satisfies the resource constraint $c_t + i_t = e^{z_t} k_t^{\alpha} l_t^{1-\alpha}$.

A competitive equilibrium can be defined in a standard way. Since both welfare theorems hold, we can solve the equivalent and simpler social planner's problem. We can think about this problem as finding policy functions for consumption $c(\cdot,\cdot)$, labor $l(\cdot,\cdot)$, and next period's capital $k'(\cdot,\cdot)$ that deliver the optimal choices as functions of the two state variables, capital and the technology level.

3.2. The Solution Methods

The Sequential Monte Carlo filter is independent of the particular nonlinear solution method employed. Aruoba, Fernández-Villaverde and Rubio-Ramírez (2003) document that the finite elements method delivers an accurate, fast, and stable solution for a wide range of parameter values in a model exactly like the one considered here. Therefore, we choose this method for our nonlinear approach. Details of how to implement the finite elements method are also provided by Aruoba, Fernández-Villaverde, and Rubio-Ramírez (2003). For the linearized approach, the situation is easier, since all the methods existing in the literature (conditional on applicability) deliver *exactly* the same solution. Out of pure convenience, we use the undetermined coefficients procedure described by Uhlig (1999).

3.3. Evaluating $p_{SMC}\left(y^T;\gamma\right)$

Let $\gamma^1 \equiv (\theta, \rho, \tau, \alpha, \delta, \beta, \sigma_\epsilon) \in \Upsilon^1 \subset R^7$ be the parameters of the neoclassical growth model. Since the finite element method requires the shocks be bounded between -1 and 1, we transform the productivity shock as $\lambda_t = \tanh(z_t)$. Let $S_t = (k_t, \lambda_t)$ be the states of the model and set $W_t = \epsilon_t$. Define $V_t \sim \mathcal{N}(0, \Sigma)$ as the vector of measurement errors. To economize on parameters we assume that Σ is diagonal with entries σ_1^2, σ_2^2 and σ_3^2. Define $\gamma^2 = (\sigma_1^2, \sigma_2^2, \sigma_3^2) \in \Upsilon^2 \subset R_+^3$ and $\gamma = (\gamma^1, \gamma^2) \in \Upsilon$. Finally, call the approximated labor policy function $l_{fem}(\cdot,\cdot;\gamma)$ and capital accumulation $k_{fem}(\cdot,\cdot;\gamma)$, where we make the dependence on the structural parameter values explicit.

The transition equation for this model is:

$$\begin{aligned} k_{t+1} &= f_1(S_{t-1}, W_t; \gamma) = k_{fem}\left(k_t, \tanh^{-1}(\lambda_t); \gamma\right) \\ \lambda_t &= f_2(S_{t-1}, W_t; \gamma) = \tanh(\rho \tanh^{-1}(\lambda_{t-1}) + \epsilon_t). \end{aligned}$$

If we assume that the observed time series, y_t, has three components: output, gdp_t, hours worked, $hours_t$, and gross investment, inv_t, the measurement equation is:

$$\begin{aligned} gdp_t &= g_1(S_t, V_t; \gamma) = x_1(S_t; \gamma) + V_{1,t} = e^{\tanh^{-1}(\lambda_t)} k_t^\alpha l_{fem}\left(k_t, \tanh^{-1}(\lambda_t); \gamma\right)^{1-\alpha} + V_{1,t} \\ hours_t &= g_2(S_t, V_t; \gamma) = x_2(S_t; \gamma) + V_{2,t} = l_{fem}\left(k_t, \tanh^{-1}(\lambda_t); \gamma\right) + V_{2,t} \\ inv_t &= g_3(S_t, V_t; \gamma) = x_3(S_t; \gamma) + V_{3,t} = k_{fem}\left(k_t, \tanh^{-1}(\lambda_t); \gamma\right) - (1-\delta)\, k_t + V_{3,t} \end{aligned}$$

We comment on two assumptions made for convenience: the observables and the presence of measurement error. First, the selection of observables keeps the dimensionality of the problem low while capturing some of the most important dynamics of the data. Three dimensions will be enough to document the differences between the two filters. Second, we

add measurement errors to avoid stochastic singularity. Nothing in our procedure critically depends on the presence of measurement errors. For example, we could instead work with a version of the model with shocks to technology, preferences, and depreciation. This alternative environment might be more empirically interesting but it would make the solution of the model much more complicated. Since our goal here is to evaluate the impact of linearization on estimation, we follow the simple route.

Given that we have four sources of uncertainty, we set $\dim(W_{2,t}) = 0$ and $W_{1,t} = W_t = \epsilon_t$. Drawing from $p(W_{1,t};\gamma)$ is then equivalent to drawing from a normal with mean zero and standard deviation σ_ϵ. Since $S_t = g(S_{t-1}, W_t;\gamma)$ the reader can note, first, that $p(y_t|W_1^t, S_0, y^{t-1};\gamma) = p(y_t|S_t;\gamma)$, and second, that drawing from $p(W_1^t, S_0|y^{t-1};\gamma)$ is equivalent to sampling from $p(S_t|y^{t-1};\gamma)$. This allows us to write the likelihood function as:

$$p\left(y^T;\gamma\right) = \prod_{t=1}^{T} \int p(y_t|S_t;\gamma)\, p\left(S_t|y^{t-1};\gamma\right) dS_t. \tag{8}$$

But since our measurement equation implies that $p(y_t|S_t;\gamma) = (2\pi)^{-\frac{3}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{\omega(S_t;\gamma)}{2}}$ where $\omega(S_t;\gamma) = (y_t - x(S_t;\gamma)))' \Sigma^{-1} (y_t - x(S_t;\gamma))\ \forall t$, we can rewrite (8) as

$$p\left(y^T;\gamma\right) = (2\pi)^{-\frac{3T}{2}} |\Sigma|^{\frac{-T}{2}} \prod_{t=1}^{T} \int e^{-\frac{\omega(S_t;\gamma)}{2}} p\left(S_t|y^{t-1};\gamma\right) dS_t.$$

The last expression is simple to handle. With the particles $\left\{\left\{s_0^{t|t-1,i}, w_1^{t|t-1,i}\right\}_{i=1}^{N}\right\}_{t=1}^{T}$ coming from our filter, we build the states $\left\{\{s_t^i\}_{i=1}^{N}\right\}_{t=1}^{T}$ and the prediction error $\left\{\{\omega(s_t^i;\gamma)\}_{i=1}^{N}\right\}_{t=1}^{T}$. We set $s_{0|0}^i = S_{ss}$, $\forall i$. Therefore, the likelihood is approximated by:

$$p\left(y^T;\gamma\right) \simeq p_{SMC}\left(y^T;\gamma\right) = (2\pi)^{-\frac{3T}{2}} |\Sigma|^{\frac{-T}{2}} \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} e^{-\frac{\omega(s_t^i;\gamma)}{2}} \tag{9}$$

3.4. Evaluating $p_{KF}\left(y^T;\gamma\right)$

Let γ, W_t, and V_t be defined as in section 3.3. The undetermined coefficients method delivers linear functions of output, investment, and labor, depending on current capital and the technology level. Then we build the transition equation for the model and, as in the previous case, the measurement equations $gdp_t = y_t + V_{1,t}$, $hours_t = l_t + V_{2,t}$ and $inv_t = y_t - c_t + V_{3,t}$. Mapping these equations into the notation of equations (3) and (4), we apply the Kalman filter and evaluate $p_{KF}\left(y^T;\gamma\right)$ as described in section 2.3. We initialize the filter to $s_{0|0} = S_{ss}$ and $P_{0|0} = 0$.

4. The Estimation Algorithm

Now we explain how to incorporate the likelihood functions in a Bayesian estimation algorithm. In the Bayesian approach, the main inference tool is the parameters' posterior distribution given the data, $\pi\left(\gamma|y^T\right)$. The posterior density is proportional to the likelihood times the prior. Therefore, we need to specify priors on the parameters, $\pi(\gamma)$, and to evaluate the likelihood function.

We specify our priors in section 5.1, and the likelihood function is evaluated either by (6) or by (7), depending on how we solve the model. Since none of these posteriors have a closed-form, we use a Metropolis-Hasting algorithm to draw from them. We call $\pi_{SMC}\left(\gamma|y^T\right)$ to the posterior implied by the Sequential Monte Carlo filter and $\pi_{KF}\left(\gamma|y^T\right)$ to the posterior derived from the Kalman filter. To simplify the notation, we let $f_{SMC}(\cdot,\cdot;\gamma_i)$ and $g_{SMC}(\cdot,\cdot;\gamma_i)$ be defined by (1) and (2), and $f_{KF}(\cdot,\cdot;\gamma_i)$ and $g_{KS}(\cdot,\cdot;\gamma_i)$ by (3) and (4).

The algorithm to draw a chain $\{\gamma_i\}_{i=1}^M$ from $\pi_j\left(\gamma|y^T\right)$, $\forall j \in \{SMC, KF\}$ is as follows:

Step 0, Initialization: Set $i \rightsquigarrow 0$ and initial γ_i. Compute functions $f_j(\cdot,\cdot;\gamma_i)$ and $g_j(\cdot,\cdot;\gamma_i)$. Evaluate $\pi(\gamma_i)$ and $p_j\left(y^T;\gamma_i\right)$ using (6) or (7). Set $i \rightsquigarrow i+1$.

Step 1, Proposal draw: Get a proposal draw $\gamma_i^p = \gamma_{i-1} + \upsilon_i$, where $\upsilon_i \sim N(0,\Psi)$.

Step 2, Solving the model: Solve the model for γ_i^p and compute $f_j(\cdot,\cdot;\gamma_i^p)$ and $g_j(\cdot,\cdot;\gamma_i^p)$.

Step 3, Evaluating the proposal: Evaluate $\pi(\gamma_i^p)$ and $p_j\left(y^T|\gamma_i^p\right)$ using either (6) or (7).

Step 4, Accept/Reject: Draw $\chi_i \sim U(0,1)$. If $\chi_i \leq \frac{\pi\left(\gamma_i^p\right)p_j\left(y^T|\gamma_i^p\right)}{\pi\left(\gamma_{i-1}\right)p_j\left(y^T|\gamma_{i-1}\right)}$ set $\gamma_i = \gamma_i^p$, otherwise $\gamma_i = \gamma_{i-1}$. If $i < M$, set $i \rightsquigarrow i+1$ and go to step 1. Otherwise stop.

We used standard methods to check the convergence of the chain generated by the Metropolis-Hasting algorithm (see Mengersen, Robert, and Guihenneuc-Jouyaux, 1999). Also, we selected the variance of the innovations in the proposals for the parameters to achieve an acceptance rate of proposals of around 40 percent.

We concentrate in this paper on Bayesian inference. However, we could also perform classical inference, maximizing the likelihood function obtained in the previous section, and building an asymptotic variance-covariance matrix using standard numerical methods. Also, the value of the likelihood function at its maximum would be useful to compute likelihood ratios for model comparison purposes.

5. Findings

We undertake two main exercises. In our first one, we simulate "artificial" data using the nonlinear solution of the model for a particular choice of values of γ^*. Then, we define some priors over γ, and we draw from its posterior distribution implied by both $p_{SMC}\left(y^T;\gamma\right)$ and $p_{KF}\left(y^T;\gamma\right)$. Finally, we compute the marginal likelihood of the "artificial" data implied by each likelihood approximation. This exercise answers the following two questions: (1) How accurate is the estimation of the "true" parameter values, γ^*, implied by each filter? and (2) How big is the improvement delivered by the Sequential Monte Carlo filter over the Kalman filter? From the posterior means, we address the first question. From the marginal likelihoods, we respond to the second.

Since the difference between the policy functions implied by the finite elements and the linear method depends greatly on γ^*, we perform the described exercise for two different values of γ^*, one with low risk aversion and low variance, γ_l^*, when both policies are close, and another with high risk aversion and high variance, γ_h^*, when the policies are farther away.

Our second exercise uses U.S. data to estimate the model with the Sequential Monte Carlo and the Kalman filters. This exercise answers the following question: Is the Sequential Monte Carlo providing a better explanation of the real data?

We divide our exposition in three parts. First, we specify the priors for the parameters. Second, we present results from the "artificial" data experiment. Finally, we present the results with U.S. data.

5.1. The Priors

We postulate flat priors for all 10 parameters, subject to some boundary constraints to make the priors proper. This choice is motivated by two considerations. First, since we are going to estimate our model using "artificial" data generated at some value γ^*, we do not want to bias the results in favor of any alternative by our choice of priors. Second, with a flat prior, the posterior is proportional to the likelihood function (except for the very small issue of the bounded support of the priors). As a consequence, our experiment can be interpreted as a classical exercise in which the mode of the likelihood function is the maximum likelihood estimate. A researcher who prefers more informative priors can always reweight the likelihood to accommodate her priors (see Geweke, 1998).

The parameter governing labor supply, θ, follows a uniform distribution between 0 and 1. That constraint imposes only a positive marginal utility of leisure. The persistence of the technology shock, ρ, also follows a uniform distribution between 0 and 1. The parameter τ follows a uniform distribution between 0 and 100. That choice rules out only risk loving behavior and risk aversions that will predict differences in interest rates several orders of

magnitude higher than the observed ones. The prior for the technology parameter, α, is uniform between 0 and 1. The prior on the depreciation rate ranges between 0 and 0.05, covering all national accounts estimates of quarterly depreciation. The discount factor, β, is allowed to vary between 0.75 and 1, implying steady state annual interest rates between 0 and 316 percent. The standard deviation of the innovation of productivity, σ_ϵ, follows a uniform distribution between 0 and 0.1, a bound 15 times higher than the usual estimates. We also pick this prior for the three standard deviations of the measurement errors. Table 5.1 summarizes our discussion.

Table 5.1: Priors for the Parameters of the Model

Parameters	Distribution	Hyperparameters
θ	Uniform	0,1
ρ	Uniform	0,1
τ	Uniform	0,100
α	Uniform	0,1
δ	Uniform	0,0.05
β	Uniform	0.75,1
σ_ϵ	Uniform	0,0.1
σ_1	Uniform	0,0.1
σ_2	Uniform	0,0.1
σ_3	Uniform	0,0.1

5.2. Results with "Artificial" Data

We simulate observations from the model and use them as data for the estimation. We generate data from two different calibrations.

First, to make our experiment as realistic as possible, we present a benchmark calibration of the model. The discount factor $\beta = 0.9896$ matches an annual interest rate of 4.27 percent. The risk aversion $\tau = 2$ is a common choice in the literature. $\theta = 0.357$ matches the micro evidence of labor supply. We reproduce the labor share of national income with $\alpha = 0.4$. The depreciation rate $\delta = 0.02$ fixes the investment/output ratio and $\rho = 0.95$ and $\sigma = 0.007$ match the historical properties of the Solow residual of the U.S. economy. With respect to the standard deviations of the measurement errors we set them equal to 0.01 percent of the steady state value of output, 0.35 percent of the steady state value of hours and 0.2 percent of the steady state value of investment based on our priors regarding the relative importance of measurement errors in the National Income and Product Accounts. We summarize the chosen values in Table 5.2.

Table 5.2: Calibrated Parameters

Parameter	θ	ρ	τ	α	δ	β	σ_ϵ	σ_1	σ_2	σ_3
Value	0.357	0.95	2.0	0.4	0.02	0.99	0.007	$1.58*10^{-4}$	0.0011	$8.66*10^{-4}$

The second calibration, which we will call extreme, maintains the same parameters except that it increases τ to 50 (implying a relative risk aversion of 24.5) and σ_ϵ to 0.035. This high risk aversion and variance introduce a strong nonlinearity to the economy. This particular choice of parameters allows us to check the differences between the Sequential Monte Carlo filter and the Kalman filter in a highly nonlinear world while maintaining a familiar framework. We justify our choice, thus, not basing it on empirical considerations but on its usefulness as a "test" case.

After generating a sample of size 100 for each of the two calibrations,[2] we apply our priors and our likelihood evaluation algorithms. For the Sequential Monte Carlo filter, we use 60,000 particles to get 50,000 draws from the posterior distribution. Since we do not suffer from an attrition problem, we do not replenish the swarm. See Fernández-Villaverde and Rubio-Ramírez (2004) for further details of this issue and of convergence. For the Kalman filter, we also get 50,000 draws. In both cases, we have a long burn-in period.

In Figure 5.1 we plot the likelihood function in logs of the model, given our simulated data for the Sequential Monte Carlo filter (continuous line) and the Kalman filter (discontinuous line).We draw in each panel the likelihood function for an interval around the calibrated value of the structural parameter, keeping all the other parameters fixed at their calibrated values. We can think of each panel then as a transversal cut of the likelihood function. To facilitate the comparison, we show the "true" value for the parameter corresponding to the direction being plotted with a vertical line, and we do not draw values lower than -20,000.

Figure 5.1 reveals two points. First, both likelihoods have the same shape and are centered on the "true" value of the parameter, although the Kalman filter delivers a slight bias for four parameters (α, δ, β and θ). Note that, since we are assuming flat priors, none of the curvature of the likelihoods is coming from the prior. Second, there is a difference in level between the likelihood generated by the Sequential Monte Carlo filter and the one delivered by the Kalman filter. This is a first proof that the nonlinear model fits the data better even for this nearly linear economy.

Table 5.3 conveys similar information: the point estimates are approximately equal regardless of the filter. However the Sequential Monte Carlo delivers estimates that are better in the sense of being closer to their pseudotrue value.[3]

[2] The results were robust when we used different simulated data from the same model. We omit details because of space considerations.

[3] The whole posteriors are available upon request from the authors. We also checked that the numerical errors of the estimates were negligible.

Table 5.3: Nonlinear versus Linear Posterior Distributions Benchmark Case

Nonlinear (SMC filter)			Linear (Kalman filter)		
Parameters	Mean	s.d	Parameters	Mean	s.d
θ	0.357	0.10×10^{-3}	θ	0.374	0.06×10^{-3}
ρ	0.950	0.29×10^{-3}	ρ	0.914	0.24×10^{-3}
τ	2.000	0.92×10^{-3}	τ	3.536	1.17×10^{-3}
α	0.400	0.10×10^{-3}	α	0.443	0.08×10^{-3}
δ	0.020	0.02×10^{-3}	δ	0.030	0.02×10^{-3}
β	0.990	0.02×10^{-3}	β	0.978	0.03×10^{-3}
σ_ϵ	0.007	0.04×10^{-4}	σ_ϵ	0.011	0.36×10^{-4}
σ_1	1.58×10^{-4}	0.15×10^{-6}	σ_1	1.86×10^{-4}	4.64×10^{-7}
σ_2	1.12×10^{-3}	0.68×10^{-6}	σ_2	5.55×10^{-4}	2.01×10^{-6}
σ_3	5.64×10^{-4}	0.81×10^{-6}	σ_3	2.42×10^{-3}	1.75×10^{-6}

Table 5.4 reports the logmarginal likelihood differences between the nonlinear and the linear case. We compute the marginal likelihood with Geweke's (1998) harmonic mean proposal. Consequently, we need to specify a bound on the support of the weight density. To ensure robustness, we report the distances for a range of values of the truncation value p from 0.1 to 0.9. All the values convey the same message: the nonlinear solution method fits the data two orders of magnitude better than the linear approximation. A good way to read this number is to use Jeffreys' (1961) rule: if one hypothesis is more than 100 times more likely than the other, the evidence is decisive in its favor. This translates into differences in logmarginal likelihoods of 4.6 or higher. Our value of 73.6 is, then, well beyond decisiveness in favor of nonlinear filtering.

Table 5.4: Logmarginal Likelihood Difference Benchmark Case

p	Nonlinear vs. Linear
0.1	73.631
0.5	73.627
0.9	73.603

Fernández-Villaverde, Rubio-Ramírez, and Santos (2004) provide a theoretical explanation for this finding. They show how the bound on the error in the likelihood induced by the linear approximation of the policy function gets compounded with the size of the sample. The intuition is as follows. Small errors in the policy function accumulate at the same rate as the sample size grows. This means that, as the sample size goes to infinity, a linear approximation will deliver an approximation of the likelihood that will fail to converge.

We now move to study the results for the extreme calibration. Figure 5.2 is equivalent to Figure 5.1 for the extreme case. First, note how the likelihood generated by the Sequential Monte Carlo filter is again centered on the "true" value of the parameter. In comparison, the likelihood generated by the Kalman filter is not. These differences will have an important impact on the marginal likelihood. Table 5.5 recasts the same information in terms of means and standard deviations of the posteriors. As in the benchmark case, the Sequential Monte Carlo delivers better estimates of the parameters of the model.

Table 5.5: Nonlinear versus Linear Posterior Distributions Extreme Case

Nonlinear (SMC filter)			Linear (Kalman filter)		
Parameters	Mean	s.d	Parameters	Mean	s.d
θ	0.357	0.08×10^{-3}	θ	0.337	0.06×10^{-3}
ρ	0.950	0.17×10^{-3}	ρ	0.894	0.19×10^{-3}
τ	50.000	0.24×10^{-1}	τ	67.70	0.10×10^{-1}
α	0.400	0.05×10^{-3}	α	0.346	0.04×10^{-3}
δ	0.020	0.05×10^{-4}	δ	0.010	0.02×10^{-4}
β	0.990	0.08×10^{-4}	β	0.996	0.08×10^{-4}
σ_ϵ	3.50×10^{-2}	0.03×10^{-4}	σ_ϵ	3.61×10^{-2}	0.09×10^{-4}
σ_1	1.58×10^{-4}	0.06×10^{-6}	σ_1	1.72×10^{-4}	0.05×10^{-6}
σ_2	1.12×10^{-3}	0.05×10^{-5}	σ_2	9.08×10^{-4}	0.03×10^{-5}
σ_3	8.66×10^{-4}	0.02×10^{-5}	σ_3	2.64×10^{-3}	0.04×10^{-5}

Table 5.6 reports the logmarginal likelihood differences between the nonlinear and the linear case for the extreme calibration for different p's. Again, we can see how the evidence in favor of the nonlinear filter is overwhelming.

Table 5.6: Logmarginal Likelihood Difference Extreme Case

p	Nonlinear vs. Linear
0.1	117.608
0.5	117.592
0.9	117.564

As a conclusion, our exercise shows how even for a nearly linear case such as the neoclassical growth model, an estimation that respects the nonlinear structure of the economy improves substantially the ability of the model to fit the data. This may indicate that we greatly handicap dynamic equilibrium economies when we linearize them before taking them to the data and that some empirical rejections of these models may be due to the biases introduced by linearization.

Our results do not imply, however, that we should completely abandon linear methods. We have also shown that their accuracy for point estimates is acceptable. For some exercises where only point estimates are required, the extra computational cost of the Sequential Monte Carlo filter may not compensate for the reduction in bias. Practitioners should weight the advantages and disadvantages of each procedure in their particular application.

5.3. Results with Real Data

Now we estimate the neoclassical growth model with U.S. quarterly data. We use real output per capita, average hours worked, and real gross fixed investment per capita from 1964:Q1 to 2003:Q1. We first remove a trend from the data using an H-P filter. In this way, we do not need to model explicitly the presence of a trend and its possible changes.

Table 5.7: Nonlinear versus Linear Posterior Distributions Real Data

Nonlinear (SMC filter)			Linear (Kalman filter)		
Parameters	Mean	s.d	Parameters	Mean	s.d
θ	0.390	0.11×10^{-2}	θ	0.423	0.19×10^{-3}
ρ	0.978	0.52×10^{-2}	ρ	0.941	0.27×10^{-3}
τ	1.717	0.12×10^{-1}	τ	1.965	0.12×10^{-2}
α	0.324	0.71×10^{-3}	α	0.412	0.37×10^{-3}
δ	0.006	0.36×10^{-4}	δ	0.019	0.79×10^{-5}
β	0.997	0.92×10^{-4}	β	0.973	0.43×10^{-5}
σ_ϵ	0.020	0.11×10^{-3}	σ_ϵ	0.009	0.43×10^{-5}
σ_1	0.45×10^{-1}	0.42×10^{-3}	σ_1	0.11×10^{-3}	0.18×10^{-6}
σ_2	0.15×10^{-1}	0.25×10^{-3}	σ_2	0.83×10^{-2}	0.34×10^{-5}
σ_3	0.38×10^{-1}	0.39×10^{-3}	σ_3	0.26×10^{-1}	0.18×10^{-4}

Table 5.7 presents the results from the posterior distributions from 100,000 draws for each filter, again after a long burn-in period. The discount factor, β, is estimated to be 0.997 with the nonlinear filter and 0.973 with the Kalman filter. This is an important difference when using quarterly data. The linear model compensates for the lack of curvature induced by its certainty equivalence with more impatience. The parameter controlling the elasticity of substitution, τ, is estimated by the nonlinear filter to be 1.717 and by the Kalman filter to be 1.965. The parameter α is close to the canonical value of one-third in the case of the Sequential Monte Carlo, and it is higher (0.412) in the case of the Kalman filter. Finally, we note how the standard deviation of the parameters is estimated to be much higher when we use the nonlinear filter than when we employ the Kalman filter, indicating that the nonlinear likelihood is more dispersed.

It is difficult to assess whether the differences in point estimates documented in Table 5.7 are big or small. A possible answer is based on the impact of the different estimates on the moments generated by the model. Macroeconomists often use these moments to evaluate the model's ability to account for the data. Table 5.8 presents the moments of the real data and reports the moments that the stochastic neoclassical growth model generates by simulation when we calibrated it at the mean of the posterior distribution of the parameters given by each of the two filters.

Table 5.8: Nonlinear versus Linear Moments Real Data

	Real Data		Nonlinear (SMC filter)		Linear (Kalman filter)	
	Mean	s.d	Mean	s.d	Mean	s.d
output	1.95	0.073	1.91	0.129	1.61	0.068
hours	0.36	0.014	0.36	0.023	0.34	0.004
inv	0.42	0.066	0.44	0.073	0.28	0.044

We highlight two observations from Table 5.8. First, the nonlinear model matches the data much better than the linearized one. This difference is significant because the moments are nearly identical if we simulate the model using the linear or the nonlinear solution method with the same set of parameter values. The differences come thus from the point estimates delivered by each procedure. The nonlinear estimation nails down the mean of each of the three observables and does a fairly good job with the standard deviations. Second, the estimation by the nonlinear filter implies a higher output, investment, and hours worked than the estimation by the linear filter.

The main reason for these two differences is the higher β estimated by the Sequential Monte Carlo. The lower discount factor induces a higher accumulation of capital and, consequently, a higher output, investment, and hours worked. The differences for the standard deviation of the economy are also important. The nonlinear economy is also more volatile than the linearized model in terms of the standard deviation of output and hours.

Table 5.9 reports the logmarginal likelihood differences between the nonlinear and the linear case. As in the previous cases, the real data strongly support the nonlinear version of the economy with differences in log terms of around 93. The differences in moments discussed above are one of the main driving forces behind the finding. A second force is that the likelihood function generated by the Sequential Monte Carlo is less concentrated than the one coming from the Kalman filter.[4]

[4] Another way to think about the marginal likelihood is as a measure of the ability of the model to forecast within sample. The much higher marginal likelihood of the nonlinear model indeed translates into a better forecasting record within the sample (this is also true for simulated data). We omit details regarding this superior forecasting power of the nonlinear procedure because of space constraints.

Table 5.9: Logmarginal Likelihood Difference Real Data

p	Nonlinear vs. Linear
0.1	93.65
0.5	93.55
0.9	93.55

6. Conclusions

We have compared the effects of estimating dynamic equilibrium models using a Sequential Monte Carlo filter proposed by Fernández-Villaverde and Rubio-Ramírez (2004) and a Kalman filter. The Sequential Monte Carlo filter exploits the nonlinear structure of the economy and evaluates the likelihood function of the model by simulation methods. The Kalman filter estimates a linearization of the economy around the deterministic steady state. The advantage of the Kalman filter is its simplicity and speed. We compare both methodologies using the neoclassical growth model. We report two main results. First, both for simulated and for real data, the Sequential Monte Carlo filter delivers a substantially better fit of the model to the data. This difference exists even for a nearly linear case. Second, the differences in terms of point estimates, even if small in absolute terms, have quite important effects on the moments of the model. From these two results we conclude that the nonlinear filter is superior as a procedure for taking models to the data.

An additional advantage of the Sequential Monte Carlo filter is that it allows the estimation of nonnormal economies. Nonnormalities or the presence of stochastic volatility may be important to account for the dynamics of macro data. Future research will address how much accuracy is gained with the use of a Sequential Monte Carlo filter when estimating this class of models.

References

[1] Anderson, E.W., Hansen, L.P., McGrattan, E.R., and Sargent, T.J. (1996). "On the Mechanics of Forming and Estimating Dynamic Linear Economies", in H.M. Amman *et al.* (eds.) *Handbook of Computational Economics*, Elsevier.

[2] Aruoba, S.B., J. Fernández-Villaverde, and J. Rubio-Ramírez (2003). "Comparing Solution Methods for Dynamic Equilibrium Economies". *Federal Reserve Bank of Atlanta Working Paper 2003-27.*

[3] Boukazez, H., E. Cardia, and F.J. Ruge-Murcia (2002). "Habit Formation and the Persistence of Monetary Shocks". *Bank of Canada Working Paper* 2002-27.

[4] Cooley, T.F. *Frontiers of Business Cycle Research.* Princeton University Press.

[5] Cooley, T.F. and E.C. Prescott (1995). "Economic Growth and Business Cycles". In T.F. Cooley (ed.), *Frontiers of Business Cycle Research.* Princeton University Press.

[6] DeJong, D.N., B.F. Ingram, and C.H. Whiteman (2000). "A Bayesian Approach to Dynamic Macroeconomics". *Journal of Econometrics* 98, 203-223.

[7] Dib, A. (2001). "An Estimated Canadian DSGE Model with Nominal and Real Rigidities". *Bank of Canada Working Paper* 2001-26.

[8] Fernández-Villaverde, J. and J. Rubio-Ramírez (2003). "Comparing Dynamic Equilibrium Models to Data: A Bayesian Approach". *Journal of Econometrics,* forthcoming.

[9] Fernández-Villaverde, J. and J. Rubio-Ramírez (2004). "Estimating Nonlinear Dynamic Equilibrium Economies: a Likelihood Approach". *Mimeo*, University of Pennsylvania.

[10] Fernández-Villaverde, J., J. Rubio-Ramírez, and M. Santos (2004). "Convergence Properties of the Likelihood of Computed Dynamic Equilibrium Models". *Mimeo*, University of Pennsylvania.

[11] Geweke, J. (1998). "Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication". *Federal Reserve Bank of Minneapolis Staff Report 249.*

[12] Hall, G.J. (1996). "Overtime, Effort and the Propagation of Business Cycle Shocks". *Journal of Monetary Economics* 38, 139-160.

[13] Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press.

[14] Ireland, P. (2002). "Technology Shocks in the New Keynesian Model". *Mimeo,* Boston College.

[15] Jeffreys, H. (1961). *Theory of Probability* (3rd ed.). Oxford University Press.

[16] Kim, C. and C.R. Nelson (1999). *State-Space Models with Regime Switching.* MIT Press.

[17] Kim, C. and J. Piger (2002). "Nonlinearity and the Permanent Effects of Recessions". *Federal Reserve Bank of St. Louis Working Paper 2002-014E.*

[18] Kim, J. (2000). "Constructing and Estimating a Realistic Optimizing Model of Monetary Policy," *Journal of Monetary Economics*, 45, 329-59.

[19] Kim, J., S. Kim, E. Schaumburg, and C. Sims (2003). "Calculating and Using Second Order Accurate Solutions of Discrete Time Dynamic Equilibrium Models". *Mimeo*, Princeton University.

[20] Kitagawa, G. (1996). "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models". *Journal of Computational and Graphical Statistics* 5, 1-25.

[21] Landon-Lane, J. (1999). "Bayesian Comparison of Dynamic Macroeconomic Models". Ph. D. Thesis, University of Minnesota.

[22] Lubik,T. and F. Schorfheide (2003). "Do Central Banks Respond to Exchange Rates? A Structural Investigation". *Mimeo*, University of Pennsylvania.

[23] McGrattan, E., R. Rogerson, and R. Wright (1997). "An Equilibrium Model of the Business Cycle with Household Production and Fiscal Policy". *International Economic Review* 33, 573-601.

[24] Mengersen, K.L., Robert, C.P., and Guihenneuc-Jouyaux, C. (1999), "MCMC Convergence Diagnostics: a 'reviewww"'. In J. Berger, J. Bernardo, A.P. Dawid and A.F.M. Smith (eds.) *Bayesian Statistics 6*, Oxford Sciences Publications.

[25] Moran, K. and V. Dolar (2002). "Estimated DGE Models and Forecasting Accuracy: A Preliminary Investigation with Canadian Data". *Bank of Canada Working Paper* 2002-18.

[26] Otrok, C. (2001). "On Measuring the Welfare Cost of Business Cycles". *Journal of Monetary Economics* 47, 61-92.

[27] Rabanal, P. and J. Rubio-Ramírez (2003). "Comparing New Keynesian Models of the Business Cycle: a Bayesian Approach". *Federal Reserve Bank of Atlanta Working Paper 2003-30.*

[28] Rust, J. (1994). "Structural Estimation of Markov Decision Processes". In Engle and McFadden (eds), *Handbook of Econometrics*, volume 4. North Holland.

[29] Sargent, T.J. (1989). "Two Models of Measurements and the Investment Accelerator". *Journal of Political Economy* 97, 251-287.

[30] Schorfheide, F. (2000). "Loss Function-Based Evaluation of DGSE Models". *Journal of Applied Econometrics* 15, 645-670.

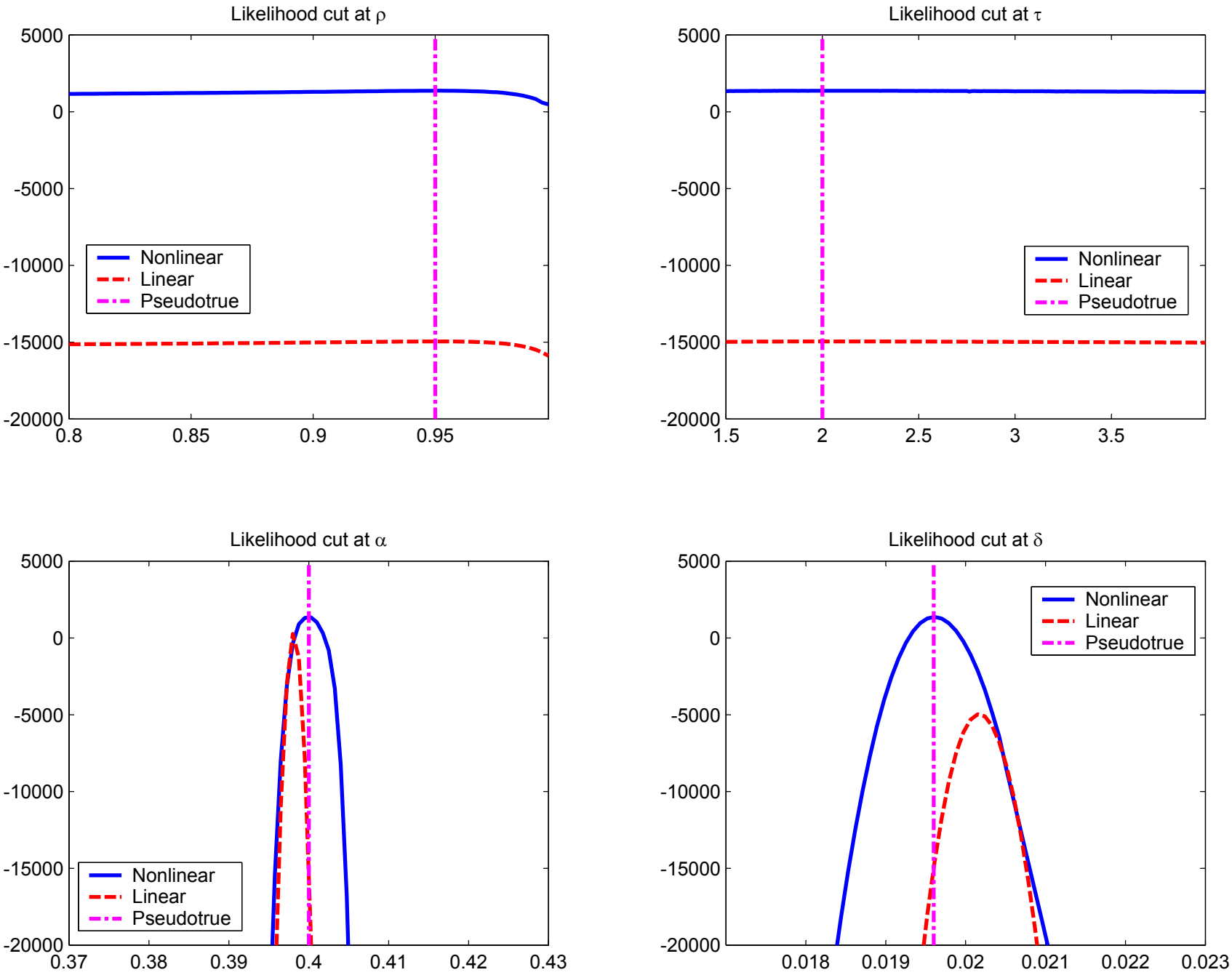
[31] Sims, C.A. and T. Zha (2002). "Macroeconomic Switching". *Mimeo*, Princeton University.

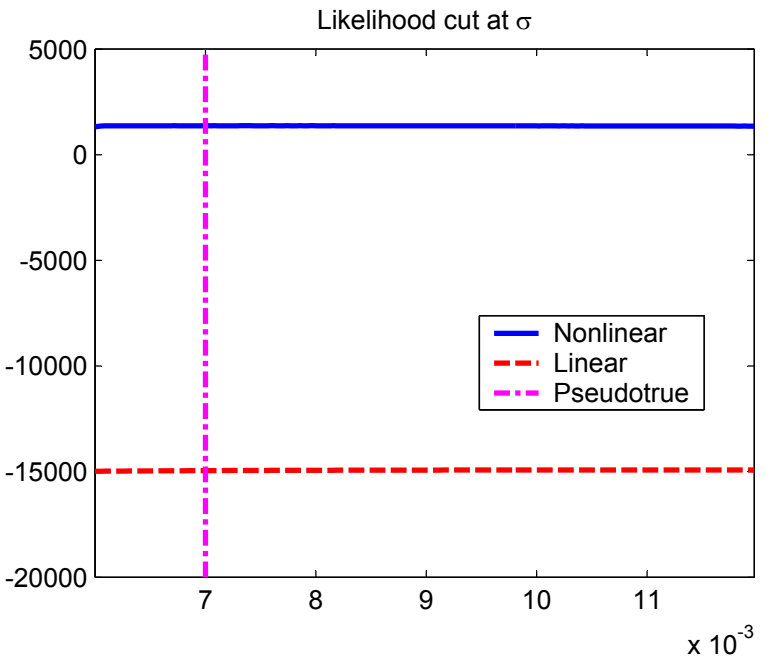
[32] Smets, F. and R. Wouters (2003). "Shocks and Frictions in US Business Cycle Fluctuations: A Bayesian DSGE Approach". *Mimeo,* European Central Bank.

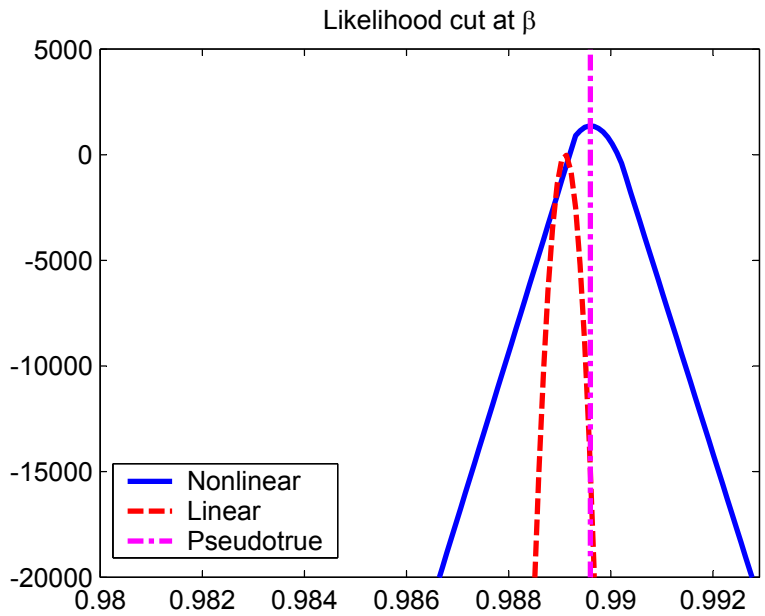
[33] Uhlig, H. (1999), "A Toolkit for Analyzing Nonlinear Dynamic Stochastic Models Easily" in R. Marimón and A. Scott (eds) *Computational Methods for the Study of Dynamic Economies.* Oxford University Press.

[34] Vuong, Q.H. (1989). "Likelihood Ratio Test for Model Selection and Non-Nested Hypothesis". *Econometrica* 57, 307-333.

Figure 5.1: Likelihood Function, Benchmark Calibration




Likelihood cut at σ
5000
0
-5000
-10000
-15000
-20000
7
8
9
10
11
x 10^-3
Nonlinear
Linear
Pseudotrue


Likelihood cut at β
5000
0
-5000
-10000
-15000
-20000
0.98
0.982
0.984
0.986
0.988
0.99
0.992
Nonlinear
Linear
Pseudotrue

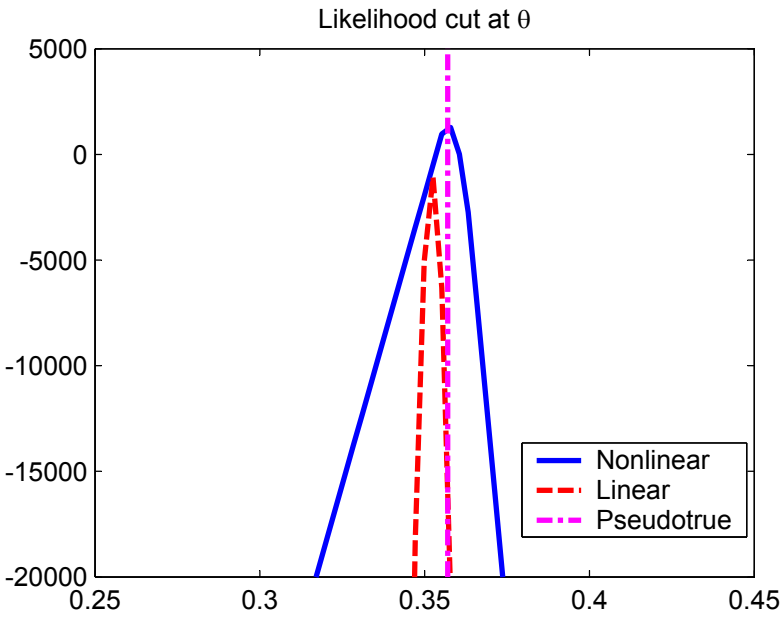
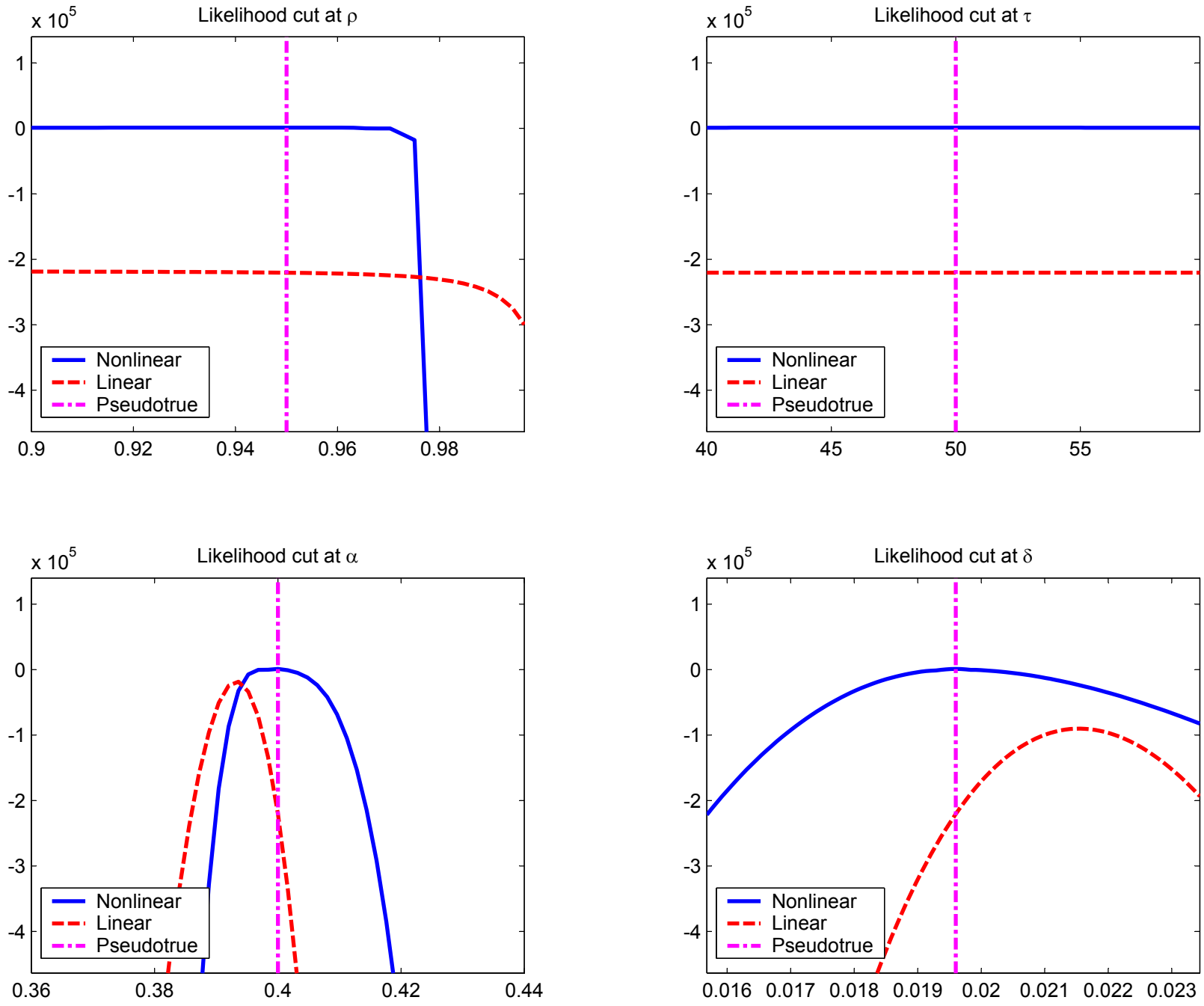
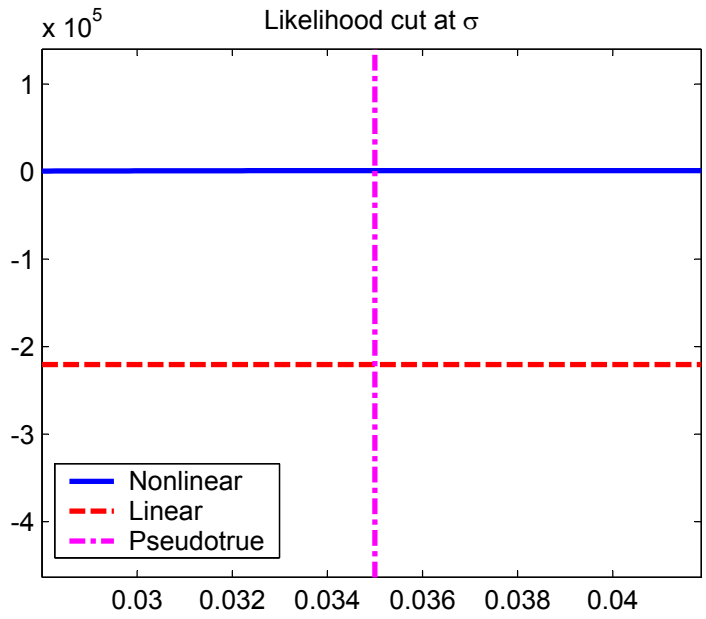
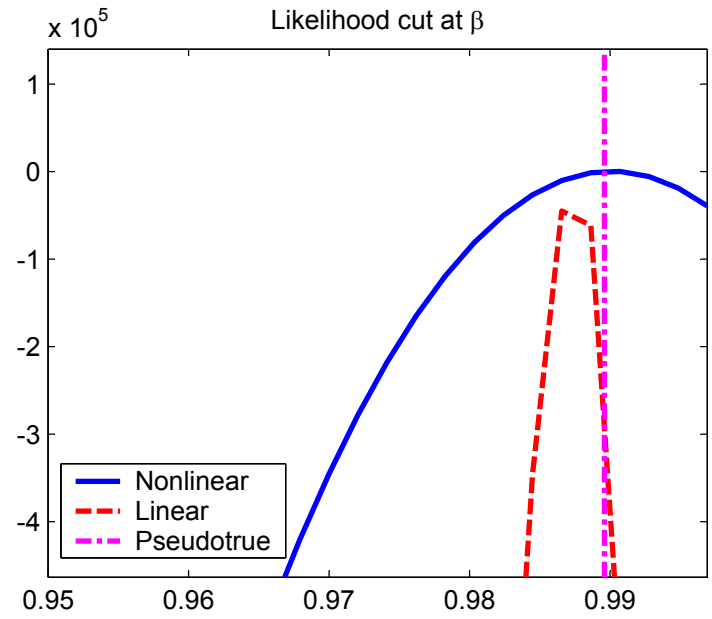
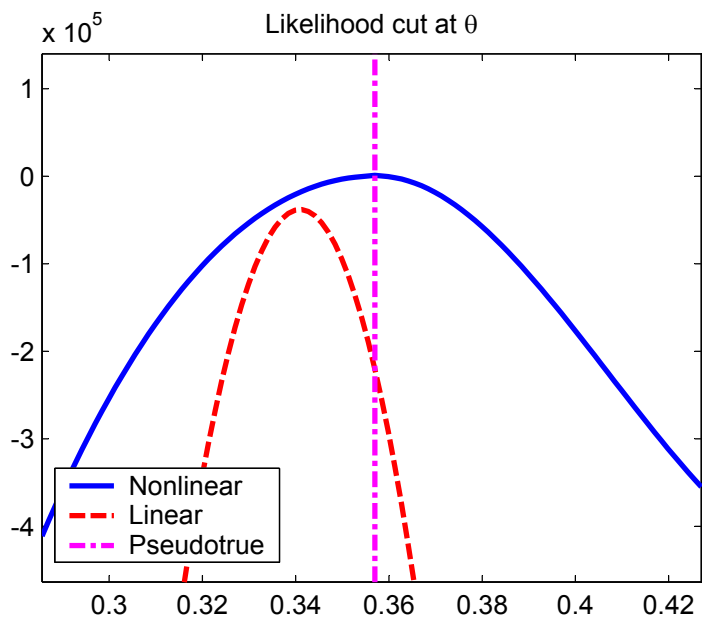

Likelihood cut at θ
5000
0
-5000
-10000
-15000
-20000
0.25
0.3
0.35
0.4
0.45
Nonlinear
Linear
Pseudotrue

Figure 5.2: Likelihood Function, Extreme Calibration




Likelihood cut at σ
x 10^5
1
0
-1
-2
-3
-4
0.03
0.032
0.034
0.036
0.038
0.04
Nonlinear
Linear
Pseudotrue


Likelihood cut at β
x 10^5
1
0
-1
-2
-3
-4
0.95
0.96
0.97
0.98
0.99
Nonlinear
Linear
Pseudotrue


Likelihood cut at θ
x 10^5
1
0
-1
-2
-3
-4
0.3
0.32
0.34
0.36
0.38
0.4
0.42
Nonlinear
Linear
Pseudotrue