FEDERAL RESERVE BANK of ATLANTA

Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood

Jesús Fernández-Villaverde and
Juan Francisco Rubio-Ramírez

Working Paper 2004-3
February 2004

WORKING PAPER SERIES

FEDERAL RESERVE BANK *of* ATLANTA | WORKING PAPER SERIES

Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood

Jesús Fernández-Villaverde and Juan Francisco Rubio-Ramírez

Working Paper 2004-3
February 2004

Abstract: This paper compares two methods for undertaking likelihood-based inference in dynamic equilibrium economies: a sequential Monte Carlo filter proposed by Fernández-Villaverde and Rubio-Ramírez (2004) 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. The authors 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, even if relatively small in absolute values, have important effects on the moments of the model. The authors conclude that the nonlinear filter is a superior procedure for taking models to the data.

JEL classification: C63, C68, E37

Key words: dynamic equilibrium economies, the likelihood function, the sequential Monte Carlo filter, the Kalman filter

---

The authors gratefully acknowledge Will Roberds, Tao Zha, and participants at several seminars for useful comments. They also thank the University of Minnesota Supercomputer Institute, which provided valuable computer assistance. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the authors' responsibility.

Please address questions regarding content to Jesús Fernández-Villaverde, Department of Economics, 160 McNeil Building, 3718 Locust Walk, University of Pennsylvania, Philadelphia, Pennsylvania 19104, 215-898-1504, jesusfv@econ.upenn.edu, or Juan Francisco Rubio-Ramírez, Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street, N.E., Atlanta, Georgia 30309-4470, 404-498-8057, juan.rubio@atl.frb.org.

Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood

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 (2003a and 2003b), 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 stochastic neoclassical growth model, which is nearly linear. We also report that, although the effects of linearization on point estimates are 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 the moments of the data.

Dynamic equilibrium models have become a standard tool in quantitative economics (see Cooley, 1995, or Ljungqvist and Sargent, 2000, for summaries of applications). An implication of these models is that they 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 this likelihood, inference is a direct exercise. In a classical environment we only need

to 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 only study an approximation to them, usually generated by a computer. The lack of a closed form for the solution of the model complicates the task 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 conditions that describe the equilibrium (first order conditions, resource constraints, laws of motion for exogenous variables and similar), 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.

A recently proposed alternative to linearization is to work instead with the nonlinear representation of the model and to apply a nonlinear filter to evaluate the likelihood. Fernández-

Villaverde and Rubio-Ramírez (2004) show how a Sequential Monte Carlo filter delivers a consistent evaluation of the likelihood function of a nonlinear and/or non-normal dynamic equilibrium model.

The presence of the two alternatives begets the following question: how different are the answers provided by each filter? 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 stochastic 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, ie., the probability that the model assigns to the data. This is true even if the differences in the point estimates of the parameters generated by the Sequential Monte Carlo filter and the Kalman filter are small.

Why is the marginal likelihood so much higher for the Sequential Monte Carlo? First, from a pure statistical perspective, the standard deviations of the posterior distributions are smaller. Given that we use flat priors in our estimation, the differences in the size of the standard deviations mean that the likelihood concentrates more mass around the pseudo-true value of the parameters in the nonlinear case. Second, and more importantly for macroeconomist, the Sequential Monte Carlo delivers points estimates for the parameters that imply model's moments closer to the moments of the data. This second result is crucial in applied work because these models are widely judged by their ability to match empirical moments.

Our finding is not the first in the literature that suggest 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 often 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 the stochastic neoclassical growth model and the linear and nonlinear solution methods that we choose. Section 4 discusses the estimation algorithm and section 5 reports our main findings with real and simulated data. Section 6 concludes. An appendix offers computational details.

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. The rest of the section is organized as follows. First, we present the state-space representation of a dynamic equilibrium model solved by nonlinear and linear methods. Second, we present 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 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 M at any γ:

$$L\left(y^T; \gamma\right) = p\left(y^T; \gamma\right), \tag{1}$$

where $\gamma \in \Upsilon$ is the vector collecting the structural parameters, those characterizing preferences, information and technology in model M.

Unfortunately, in general it is not possible to compute this function. Part of the reason is that most dynamic equilibrium 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 function (1).

There are two main routes to attack this problem. If we opt for a nonlinear solution method, we need to use the Sequential Monte Carlo algorithm as described in Fernández-Villaverde and Rubio-Ramírez (2004) to evaluate the likelihood. If we linearize the model, we can approximate (1) with the Kalman filter. We now describe both methodologies in more detail.

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:

$$S_t = f\left(S_{t-1}, W_t; \gamma\right) \tag{2}$$

where $\{W_t\}$ is a sequence of exogenous random variables.

The observable y_t is governed by the measurement equation:

$$Y_t = g\left(S_t, V_t; \gamma\right) \tag{3}$$

where $\{V_t\}$ is a sequence of exogenous independent random variables. The sequences $\{W_t\}$ and $\{V_t\}$ are independent of each other.[1] Along some dimension, the function g can be the identity mapping if a state is directly observed without noise.

The functions f and g depend on the equations that describe the equilibrium of the model - policy functions, laws of evolutions 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 need to 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] 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.

[2] See Fernández-Villaverde and Rubio-Ramírez (2004) for a more detailed discussion of stochastic singularity and how to fix it.

2.1.2. The Linear Solution of the Model

On the other hand, if we opt for a linear method to solve the same model, the state-space representation has the following linear form:

$$S_t = E(\gamma) + A(\gamma) S_{t-1} + B(\gamma) W_t \quad (4)$$

$$Y_t = F(\gamma) + C(\gamma) S_t + D(\gamma) V_t \quad (5)$$

where $A(\gamma)$, $B(\gamma)$, $C(\gamma)$, $D(\gamma)$, $E(\gamma)$, and $F(\gamma)$ are matrices with the required dimension which depend on the structural parameters of the model. Notice how this representation is nothing more than a particular case of (2) and (3). Also, we make the same assumptions regarding stochastic singularity as above.

We have presented two state-space representations of the same economy. Section 2.2 introduces a Sequential Monte Carlo filter to evaluate the likelihood function implied by (2) and (3). Section 2.3 exploits the Kalman filter to calculate the likelihood entailed by (4) and (5).

2.2. The Nonlinear Approach: A Sequential Monte Carlo Filter

Fernández-Villaverde and Rubio-Ramírez (2004) propose the following Sequential Monte Carlo method to evaluate the likelihood function of y^T induced by (2) and (3).

First, 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 $\dim(W_{2,t}) + \dim(V_t) = \dim(Y_t)$. If $\dim(V_t) = \dim(Y_t)$, we set $W_{1,t} = W_t \ \forall t$, i.e. $\{W_{2,t}\}$ to be a zero-dimensional sequence. Second, we set $W_{2,t} = W_t \ \forall t$, i.e., $\{W_{1,t}\}$ to be a zero-dimensional sequence, only if $\dim(W_t) + \dim(V_t) = \dim(Y_t)$.[3]

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 $y^0 = \{\emptyset\}$.

[3] We could make weaker assumptions, paying the cost of heavier notation.

We factor the likelihood function (1) as follows:

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

$$=\int\left(\prod_{t=1}^{T}\int p\left(y_t|W_1^t,y^{t-1},S_0;\gamma\right)p\left(W_1^t|y^{t-1},S_0;\gamma\right)dW_1^t\right)p\left(S_0;\gamma\right)dS_0. \tag{6}$$

Therefore, conditional on having N draws of $\{s_0^i\}_{i=1}^N$ from the density $p\left(S_0;\gamma\right)$ and N draws $\left\{\left\{w_1^{t|t-1,i}\right\}_{i=1}^N\right\}_{t=1}^T$ from the sequence of densities $\{p\left(W_1^t|y^{t-1},S_1;\gamma\right)\}_{t=1}^T$, the likelihood function (1) can be approximated by:

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

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 $\{p\left(W_1^t|y^{t-1},S_0;\gamma\right)\}_{t=1}^T$. The Sequential Monte Carlo filter accomplishes this objective.

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

Fernández-Villaverde and Rubio-Ramírez (2004) prove the following result that shows how to use $p\left(W_1^t|y^{t-1},S_0;\gamma\right)$ as an important sampling density to draw from $p\left(W_1^t|y^t,S_0;\gamma\right)$.

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

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

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

The proposition 1 shows how a draw $\left\{w_1^{t|t-1,i}\right\}_{i=1}^N$ from $p\left(W_1^t|y^{t-1},S_0;\gamma\right)$ can be used to get a draw $\left\{w_1^{t,i}\right\}_{i=1}^N$ from $p\left(W_1^t|y^t,S_0;\gamma\right)$. This result is key in the following Sequential Monte Carlo filter that generates a sequence of draws $\left\{\left\{w_1^{t|t-1,i}\right\}_{i=1}^N\right\}_{t=1}^T$ from the sequence of densities $\left\{p\left(W_1^t|y^{t-1},S_1;\gamma\right)\right\}_{t=1}^T$:

Step 0, Initialization: `Set` $t \rightsquigarrow 1$ `and generate` N `i.i.d. initial states` $\left\{s_0^i\right\}_{i=1}^N$ `from` $p\left(S_0;\gamma\right)$`. Initialize` $p\left(W_1^{t-1}|y^{t-1},S_0;\gamma\right)=1$.

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

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

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

The intuition of the algorithm is as follows. Given particles at $t-1$, $\left\{w_1^{t-1,i}\right\}_{i=1}^N$ distributed according to $p\left(W_1^{t-1}|y^{t-1},S_0;\gamma\right)$, step 1 generates draws $\left\{w_1^{t|t-1,i}\right\}_{i=1}^N$ from $p\left(W_1^t|y^{t-1},S_0;\gamma\right)$. Step 3 takes advantage of proposition 1 and resamples from $\left\{w_1^{t|t-1,i}\right\}_{i=1}^N$ a new swarm of particles, $\left\{w_1^{t,i}\right\}_{i=1}^N$distributed according to $p\left(W_1^t|y^t,S_0;\gamma\right)$. The output of the algorithm $\left\{s_0^i\right\}_{i=1}^N$ and $\left\{\left\{w_1^{t|t-1,i}\right\}_{i=1}^N\right\}_{t=2}^T$ is used to compute the likelihood:

$$p\left(y^T;\gamma\right) \simeq p_{SMC}\left(y^T;\gamma\right)=\frac{1}{N}\left(\prod_{t=1}^T\frac{1}{N}\sum_{i=1}^N p\left(y_t|w_1^{t|t-1,i},y^{t-1},s_0^i;\gamma\right)\right). \tag{7}$$

Step 3 is the key of the algorithm. 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 interested reader can find further details in Fernández-Villaverde and Rubio-Ramírez (2004). In particular they present all the technical details and discuss convergence, at both the theoretical and the practical level.

2.3. The Linear Approach: The Kalman filter

Now we describe how to evaluate the likelihood function implied by (4) and (5) using 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 $\widetilde{V}_t \sim N(0, R(\gamma))$.

Let us introduce some notation. First, we denote by $S_{t+1|t} = E(S_{t+1}|y^t)$ to be the linear projection of S_{t+1} on y^t and a constant, and call $y_{t+1|t} = E(Y_{t+1}|y^t) = F(\gamma) + C(\gamma) S_{t+1|t}$ to be the linear projection of Y_{t+1} on y^t and a constant. Also let $P_{t+1|t} = E\left(S_{t+1} - S_{t+1|t}\right)\left(S_{t+1} - S_{t+1|t}\right)'$, be the mean squared forecasting error when projecting S_{t+1}. Then the mean squared forecasting error when projecting Y_{t+1} is

$$\Sigma_{t+1|t} = E\left(Y_{t+1} - y_{t+1|t}\right)\left(Y_{t+1} - y_{t+1|t}\right)' = C(\gamma) P_{t+1|t} C(\gamma)' + R(\gamma),$$

and $E\left(Y_{t+1} - y_{t+1|t}\right)\left(S_{t+1} - S_{t+1|t}\right)' = C(\gamma) P_{t+1|t}$.

Given that the model is linear and all random variables are normally distributed we only need to keep track of their mean and variance-covariance matrix. Given $S_{t|t-1}$, $P_{t|t-1}$ and observation y_t, this is done by the Ricatti equations:

$$\begin{aligned} y_{t|t-1} &= F(\gamma) + C(\gamma) S_{t|t-1}, \\ \Sigma_{t|t-1} &= C(\gamma) P_{t|t-1} C(\gamma)' + R(\gamma) \\ S_{t+1|t} &= E(\gamma) + A(\gamma)\left(S_{t|t-1} + C(\gamma) P_{t|t-1} \Sigma_{t|t-1}^{-1}\left(y_t - y_{t|t-1}\right)\right) \end{aligned}$$

and

$$P_{t+1|t} = A(\gamma)\left(P_{t|t-1} - P_{t|t-1} C(\gamma)' \Sigma_{t|t-1}^{-1} C(\gamma) P_{t|t-1}\right) A(\gamma)' + Q(\gamma).$$

Since W_t and V_t are assumed to be normally distributed, the output of the previous equations can be used to calculate the likelihood function as follows:

$$p\left(y^T;\gamma\right) \simeq p_{KF}\left(y^T;\gamma\right) = \prod_{t=1}^{T} \frac{1}{\sqrt{(2\pi)^n \left|\Sigma_{t|t-1}\right|}} \exp\left(-\frac{\left(y_t - y_{t|t-1}\right)' \Sigma_{t|t-1}^{-1} \left(y_t - y_{t|t-1}\right)}{2}\right) \quad (8)$$

where $n = \dim\left(Y_t\right)$.[4]

3. An Application

Section 2 described two ways to approximate the likelihood function. If we take the nonlinear approach to solve the model, the Sequential Monte Carlo filter provides $p_{SMC}\left(y^T;\gamma\right)$. If we opt for a linear method, the Kalman filter delivers $p_{KF}\left(y^T;\gamma\right)$. This section presents a comparison between the two alternatives. We select the stochastic neoclassical growth model for that purpose. 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 stochastic 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 stochastic 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 Stochastic Neoclassical Growth Model

We work with the stochastic neoclassical growth model with leisure. Since this model is widely used (see Cooley and Prescott, 1995) we go through only the minimum exposition

[4]Notice that $S_{1|0}$ and $P_{1|0}$ have to be initialized. In general, they are set to the steady state values of the model.

required to fix notation.

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

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

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 a stochastic process affecting the technological progress. 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.

The solution is fully characterized by the following two stochastic partial differential equations, an Euler intertemporal condition:

$$\frac{\left(c_t^{\theta} \left(1 - l_t\right)^{1-\theta}\right)^{1-\tau}}{c_t} = \beta E_t \left\{ \frac{\left(c_{t+1}^{\theta} \left(1 - l_{t+1}\right)^{1-\theta}\right)^{1-\tau}}{c_{t+1}} \left(1 + \alpha e^{z_t + 1} k_{t+1}^{\alpha - 1} l_t^{\alpha} - \delta\right) \right\}, \tag{9}$$

and a static optimality condition:

$$\frac{1-\theta}{\theta} \frac{c_t}{1 - l_t} = (1-\alpha) e^{z_t} k_t^{\alpha} l_t^{-\alpha} \tag{10}$$

plus the stochastic process for productivity, the law of motion for capital, the economy resource constraint, and the boundary condition $c(0, z_t; \theta) = 0$.

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. The problem is simplified noting that we

only need to search for the solution $l(\cdot,\cdot)$ and find $c(\cdot,\cdot)$ using the static first order condition and $k'(\cdot,\cdot)$ using the resource constraint of the economy.

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 element method delivers an accurate, fast, and stable solution for a wide range of parameter values in a model exactly like the one consider here. Therefore, we choose this method for our nonlinear approach. Details of how to implement the finite element method are provided in the appendix. 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 discussed in the appendix.

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

This section describes the implementation of the Sequential Monte Carlo filter for the neoclassical growth model. Let $\gamma^1 \equiv (\theta,\rho,\tau,\alpha,\delta,\beta,\sigma_\epsilon) \in \Upsilon^1 \subset R^7$ be the structural parameters. 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 $W_t = \epsilon_t$. Let also $S_{ss} = (k_{ss},\tanh(0))$, the value of the states' variables at the deterministic steady state of the economy.

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 $l_{fem}(\cdot,\cdot;\gamma)$ where we make the dependence on the structural parameter values explicit.

The transition equation for this model is:

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

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:

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

It would be useful below to define the vector $x(S_t; \gamma)$ of predictions of the model regarding observables. Those are given by the measurement equation without the measurement errors, and they are equal to:

$$\begin{aligned}
x_1(S_t; \gamma) &= e^{\tanh^{-1}(\lambda_t)} k_t^{\alpha} l_{fem}\left(k_t, \tanh^{-1}(\lambda_t); \gamma\right)^{1-\alpha} \\
x_3(S_t; \gamma) &= l_{fem}\left(k_t, \tanh^{-1}(\lambda_t); \gamma\right) \\
x_3(S_t; \gamma) &= e^{\tanh^{-1}(\lambda_t)} k_t^{\alpha} l_{fem}\left(k_t, \tanh^{-1}(\lambda_t); \gamma\right)^{1-\alpha} \\
&\quad * \left(1 - \frac{\theta}{1-\theta}(1-\alpha)\frac{\left(1 - l_{fem}\left(k_t, \tanh^{-1}(\lambda_t); \gamma\right)\right)}{l_{fem}\left(k_t, \tanh^{-1}(\lambda_t); \gamma\right)}\right)
\end{aligned}$$

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 the fact that we have four sources of uncertainty, we set $\dim(W_{2,t}) = 0$ and $W_{1,t} =$

$W_t = \epsilon_t$. The likelihood function is given by:

$$p\left(y^T;\gamma\right) = \int \left(\prod_{t=1}^{T} \int p\left(y_t|W_1^t, y^{t-1}, S_0;\gamma\right) p\left(W_1^t|y^{t-1}, S_0;\gamma\right) dW_1^t\right) p\left(S_0;\gamma\right) dS_0. \tag{11}$$

Since $\dim(W_{2,t}) = 0$, $W_{1,t} = W_t$ and $S_t = g\left(S_{t-1}, W_t;\gamma\right)$ observe, first, that:

$$p\left(y_t|W_1^t, y^{t-1}, S_0;\gamma\right) = p\left(y_t|W^t, y^{t-1}, S_0;\gamma\right) = p\left(y_t|S_t;\gamma\right),$$

and second, that drawing from $p\left(W_1^t|y^{t-1}, S_0;\gamma\right)$ is equivalent to draw from $p\left(S_t|y^{t-1}, S_0;\gamma\right)$. This allow us to write the likelihood function (11) as:

$$p\left(y^T;\gamma\right) = \int \left(\prod_{t=1}^{T} \int p\left(y_t|S_t;\gamma\right) p\left(S_t|y^{t-1}, S_0;\gamma\right) dS_t\right) p\left(S_0;\gamma\right) dS_0. \tag{12}$$

But since our measurement equation implies that $p\left(y_t|S_t;\gamma\right) = (2\pi)^{-\frac{3}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{\omega(S_t;\gamma)}{2}}$ where we define the prediction errors to be $\omega(S_t;\gamma) = \left(y_t - x(S_t;\gamma)\right)' \Sigma^{-1} \left(y_t - x(S_t;\gamma)\right)$ $\forall t$, we can rewrite (12) as

$$p\left(y^T;\gamma\right) = (2\pi)^{-\frac{3T}{2}} |\Sigma|^{\frac{-T}{2}} \int \left(\prod_{t=1}^{T} \int e^{-\frac{\omega(S_t;\gamma)}{2}} p\left(S_t|y^{t-1}, S_0;\gamma\right) dS_t\right) p\left(S_0;\gamma\right) dS_1.$$

The last expression is simple to handle. With the particles $\left\{\left\{w_1^{t|t-1,i}\right\}_{i=1}^{N}\right\}_{t=1}^{T}$ and $\{s_0^i\}_{i=1}^{N}$ coming from our filter, we can 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 $s_0^i = S_{ss}$ $\forall i$. Therefore, the likelihood function is approximated by:

$$p\left(y^T;\gamma\right) \simeq p_{SMC}\left(y^T;\gamma\right) = (2\pi)^{-\frac{3T}{2}} |\Sigma|^{\frac{-T}{2}} \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} e^{-\frac{\omega(s_t^i;\gamma)}{2}} \tag{13}$$

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 linearization does not need to bound the perturbation space. Therefore $S_t = (k_t, z_t)$. Also, let $y_{ss}(\gamma), c_{ss}(\gamma), l_{ss}(\gamma)$ and $k_{ss}(\gamma)$ be deterministic steady state values for output, consumption, labor, and capital. Then $S_{ss}(\gamma) = (k_{ss}(\gamma), 0)$, the value of the deterministic states variables at the deterministic steady state of the model. After implementing the undetermined coefficients method we get:

$$\begin{aligned}
k_t &= k_{ss}(\gamma) + a_{11}(\gamma)(k_{t-1} - k_{ss}(\gamma)) + a_{12}(\gamma)(\rho z_{t-1} + W_t), \\
l_t &= l_{ss}(\gamma) + a_{21}(\gamma)(k_t - k_{ss}(\gamma)) + a_{22}(\gamma) z_t, \\
c_t &= c_{ss}(\gamma) + a_{31}(\gamma)(k_t - k_{ss}(\gamma)) + a_{32}(\gamma) z_t, \\
y_t &= y_{ss}(\gamma) + a_{41}(\gamma)(k_t - k_{ss}(\gamma)) + a_{42}(\gamma) z_t,
\end{aligned}$$

the equilibrium policy functions for capital, hours, consumption, and output.[5]

Then the transition equation for this model is:

$$\begin{aligned}
k_t &= k_{ss}(\gamma) + a_{11}(\gamma)(k_{t-1} - k_{ss}(\gamma)) + a_{12}(\gamma)(\rho z_{t-1} + W_t) \\
z_t &= \rho z_{t-1} + W_t
\end{aligned}$$

and with the same three observables as in the previous case, the measurement equation is:

$$\begin{aligned}
gdp_t &= y_t + V_{1,t} \\
hours_t &= l_t + V_{2,t} \\
inv_t &= y_t - c_t + V_{3,t}
\end{aligned}$$

[5] All the a's are functions of the structural parameters of the model. See the appendix for details.

In the notation of equations (4) and (5) we have:

$$E(\gamma) = \begin{bmatrix} (1 - a_{11}(\gamma)) k_{ss}(\gamma) \\ 0 \end{bmatrix}$$

$$A(\gamma) = \begin{bmatrix} a_{11}(\gamma) k_{t-1} & a_{12}(\gamma) \rho z_{t-1} \\ 0 & \rho \end{bmatrix}$$

$$B(\gamma) = \begin{bmatrix} a_{12}(\gamma) \\ 0 \end{bmatrix}$$

$$F(\gamma) = \begin{bmatrix} y_{ss}(\gamma) - a_{41}(\gamma) k_{ss}(\gamma) \\ l_{ss}(\gamma) - a_{21}(\gamma) k_{ss}(\gamma) \\ y_{ss}(\gamma) - c_{ss}(\gamma) - (a_{41}(\gamma) - a_{31}(\gamma)) k_{ss}(\gamma) \end{bmatrix}$$

$$C(\gamma) = \begin{bmatrix} a_{41}(\gamma) & a_{42}(\gamma) \\ a_{21}(\gamma) & a_{22}(\gamma) \\ a_{41}(\gamma) - a_{31}(\gamma) & a_{42}(\gamma) - a_{32}(\gamma) \end{bmatrix}$$

and $D(\gamma) = I_{3\times 3}$. Then we evaluate $p_{KF}\left(y^T; \gamma\right)$ as described in section 2.3.

4. The Estimation Algorithm

Now we explain how to incorporate the likelihood functions (7) and (8) in an 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 (7) or by (8), 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 (2) and (3), and $f_{KF}(\cdot, \cdot; \gamma_i)$ and $g_{KS}(\cdot, \cdot; \gamma_i)$ by (4) and (5).

The algorithm to draw a chain $\{\gamma_i\}_{i=1}^M$ from $\pi_j\left(\gamma | y^T\right)$, $\forall j \in \{SMC, FK\}$ 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 (7) or (8). Set $i \rightsquigarrow i+1$.

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

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 (7) or (8).

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.

Once we obtain $\{\gamma_i\}_{i=1}^M$, any moments of interest can be computed from $\pi_j\left(\gamma|y^T\right)$, $\forall j \in \{SMC, FK\}$.

The convergence of the algorithm depends on the fulfillment of a number of technical conditions. In practice, it is extremely important to adjust the variance of the innovation of the proposal density to get an appropriate acceptance rate.[6] If the rate is small, the chain does not visit the tails of the posterior. If the acceptance rate is high, the chain does not stay enough time at the high probability regions. Gelman, Roberts and Gilks (1996) suggest that a 20 percent acceptance rate tends to give the best performance. We found that a rate of around 30 percent outperformed different alternatives. A complete guide to convergence can be found in Mengersen, Robert and Guihenneuc-Jouyaux (1999).

In this paper, we concentrate on Bayesian inference because of space considerations. However, we could also perform classical inference. For that, once we obtain the likelihood, we can introduce it into a maximization routine. The output of the algorithm, $\widehat{\gamma}_{MLE}$, is the maximum likelihood point estimate. We can compute the asymptotic variance-covariance matrix of the parameters as:

$$var(\widehat{\gamma}_{MLE}) = -\left(\frac{\partial^2 L\left(y^T;\widehat{\gamma}_{MLE}\right)}{\partial\gamma\partial\gamma'}\right)^{-1}$$

[6] The acceptance rate is equal to the number of times the chain changes position divided by the number of iterations.

Since in general we cannot evaluate this second derivative directly, a numerical approximation needs to be used. Finally, the value of the likelihood function at its maximum is also useful building likelihood ratios for model comparison purposes.

5. Findings

We undertake two main exercises. Our first exercise as follows. First, 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 mean of each filter, we answer the first of these two questions. From the marginal likelihoods, we respond to the second.

Aruoba, Fernández-Villaverde and Rubio-Ramírez (2003) report that the difference between the policy functions implied by the finite element and the linear methods depends greatly on γ^*. If we consider high risk aversion and high variance of the productivity shock innovations, the policy function looks more different than in the case with low risk aversion and low variance. For these reason, we perform the described exercise for two different values of γ^*, one with low risk aversion and low variance, γ_l^*, and another with high risk aversion and high variance, γ_h^*.

Our second exercise uses real 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 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 real 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.[7] 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 Bayesian researcher that prefers more informative priors can always reweight the likelihood to accommodate her priors (see Geweke, 1998).

Table 5.1: Priors for the Parameters of the Model

Parameters	Distribution	Hyperparameters
θ	Uniform	0,1
ρ	Uniform	0.5,1
τ	Uniform	0,100
α	Uniform	0,1
δ	Uniform	0,0.1
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

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. This region implies a stationary distribution of the variables of the model[8] with a lower bound on no persistence. The parameter governing the elasticity of substitution, τ, follows a uniform between 0 and 100. That choice only rules out 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, ranges between 0.75 and 1, implying steady state annual interest

[7] Except for the very small issue of the bounded support of the priors.

[8] See Fernández-Villaverde and Rubio-Ramírez (2004) for a discussion on nonstationarity.

rates between 0 and 316 percent. The standard deviation of the innovation of productivity, σ_ϵ, follows a uniform 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 the discussion.

5.2. Results with "Artificial" Data

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

First, to make our experiment as realistic as possible, we calibrate the model following standard practices (Cooley and Prescott, 1995). We will call this the benchmark calibration. The discount factor $\beta = 0.9896$ matches an annual interest rate of 4.27 percent (McGrattan and Prescott, 2000). The risk aversion $\tau = 2$ is a common choice in the literature. $\theta = 0.357$ matches the microeconomic 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 a 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 of 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, that we will call extreme from now on, 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 then, 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,[9] we apply our priors and our likelihood evaluation algorithms. For the Sequential Monte Carlo filter we use 40,000 particles to get 50,000 draws from the posterior distribution. 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). Since we cannot draw a 10 dimensional figure, we plot in each panel the likelihood function for an interval of ± 20 percent of the calibrated value of the structural parameter, keeping all the other parameters fixed at their calibrated values.[10] 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.

Figure 5.1 reveals two points. First, for nearly all parameters (except θ), both likelihoods have the same shape, and they are roughly centered on the "true" value of the parameter. Note that since we are assuming flat priors, none of this curvature 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 Kalman filter. This is a first proof that the nonlinear model fits the data better even for this nearly linear economy.

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

[10] We do not draw the loglikelihood function when it takes values less than -2,000 to enhance the readability of the figure.

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	0.07×10^{-3}	θ	0.360	0.30×10^{-3}
ρ	0.950	0.34×10^{-3}	ρ	0.950	0.86×10^{-3}
τ	2.000	0.68×10^{-3}	τ	2.007	3.38×10^{-3}
α	0.400	0.09×10^{-3}	α	0.399	0.11×10^{-3}
δ	0.020	0.01×10^{-3}	δ	0.020	0.01×10^{-3}
β	0.990	0.02×10^{-3}	β	0.990	0.02×10^{-3}
σ_ϵ	0.007	0.09×10^{-4}	σ_ϵ	0.007	0.12×10^{-4}
σ_1	1.58×10^{-4}	5.75×10^{-8}	σ_1	1.58×10^{-4}	2.84×10^{-7}
σ_2	1.12×10^{-3}	6.43×10^{-7}	σ_2	1.12×10^{-3}	1.06×10^{-6}
σ_3	5.64×10^{-4}	6.49×10^{-7}	σ_3	8.69×10^{-4}	1.31×10^{-6}

Table 5.3 conveys similar information: the point estimates are approximately equal regardless of the filter. On the other hand, the standard deviations are bigger in the Kalman filter case. Since we use flat priors, the posterior is proportional to the likelihood. Consequently the Sequential Monte Carlo delivers a likelihood function more concentrated around the "true" value of the parameter. This result will have a dramatic impact on the marginal likelihood of the model.[11]

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 show the robustness of our finding to different values of this bound and following Geweke's (1998) advice, 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. This is just another way to summarize the differences observed in the levels of the likelihood plotted in figure 5.1. To put this number in perspective we may want to note that this difference is substantially

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

bigger than 7, a bound for DNA testing in forensic science, often accepted by courts of law as evidence beyond reasonable doubt (Evett, 1991).

Table 5.4: Logmarginal Likelihood Difference Benchmark Case

p	Nonlinear vs. Linear
0.1	163.045
0.3	164.082
0.5	164.465
0.7	164.615
0.9	164.661

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. For example, in the case of ρ, the maximum of the nonlinear approach is nearly the "true" value of the parameter while the Kalman filter delivers a maximum more than 20 percent below this "true" value. The case of θ is even more striking. The supports of the likelihoods are numerically disjointed and while the nonlinear likelihood is centered on the "true" value, the linear likelihood is numerically equivalent to zero at this point. Other parameters tell similar histories.

Table 5.5 recasts the same information in terms of means and standard deviations of the posteriors. As in the benchmark case, the standard deviations are bigger when we use the Kalman filter. This finding means that the linear filter provides the researcher with a more disperse likelihood function with the consequent impact on the marginal likelihood.

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	7.19×10^{-5}	θ	0.370	1.90×10^{-4}
ρ	0.950	1.88×10^{-4}	ρ	0.863	2.22×10^{-3}
τ	50.000	7.12×10^{-3}	τ	51.13	1.95×10^{-1}
α	0.400	4.80×10^{-4}	α	0.400	1.33×10^{-4}
δ	1.96×10^{-2}	3.53×10^{-6}	δ	1.98×10^{-2}	2.45×10^{-5}
β	0.990	8.69×10^{-6}	β	0.990	3.51×10^{-4}
σ_ϵ	3.50×10^{-2}	4.47×10^{-6}	σ_ϵ	3.47×10^{-2}	6.52×10^{-4}
σ_1	1.58×10^{-4}	1.87×10^{-8}	σ_1	1.61×10^{-4}	3.50×10^{-7}
σ_2	1.12×10^{-3}	2.14×10^{-7}	σ_2	1.11×10^{-3}	1.08×10^{-6}
σ_3	8.66×10^{-4}	2.33×10^{-7}	σ_3	8.69×10^{-4}	2.06×10^{-6}

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	80.731
0.3	80.734
0.5	80.713
0.7	80.661
0.9	80.613

As a conclusion, our exercise shows how even for a nearly linear case such as the stochastic 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.

5.3. Results with Real Data

Now we apply our procedure to estimate the stochastic 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 presents the results from the posterior distributions from 50,000 draws for each filter, again after a long burn-in period.

We briefly discuss some of our results. The discount factor, β, is estimated to be 0.997 with the nonlinear filter and 0.98 with the Kalman filter. This is an important difference (remember that we are using quarterly data). The parameter controlling the elasticity of substitution, τ, is estimated by the nonlinear filter to be 1.825 and by the Kalman filter to be 9.71. The linear model compensates for the lack of curvature induced by its certainty equivalence with higher risk aversion. The parameter α is close to the canonical value of one third in the case of the Sequential Monte Carlo, while it is a bit higher (0.4) in the case of the Kalman filter. Finally, we note how the standard deviation of the measurement error is estimated to be much higher when we use the nonlinear filter than when we employ the Kalman filter.

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	0.388	7.19×10^{-5}	θ	0.367	4.07×10^{-3}
ρ	0.969	1.88×10^{-4}	ρ	0.952	8.54×10^{-3}
τ	1.825	7.12×10^{-3}	τ	9.710	3.53×10^{-1}
α	0.323	4.80×10^{-4}	α	0.406	2.18×10^{-3}
δ	0.61×10^{-2}	3.53×10^{-6}	δ	1.87×10^{-2}	4.27×10^{-4}
β	0.997	8.69×10^{-6}	β	0.980	6.53×10^{-4}
σ_ϵ	2.35×10^{-2}	4.47×10^{-6}	σ_ϵ	1.65×10^{-2}	1.02×10^{-3}
σ_1	3.98×10^{-2}	1.87×10^{-8}	σ_1	1.51×10^{-4}	5.54×10^{-6}
σ_2	1.81×10^{-2}	2.14×10^{-7}	σ_2	8.00×10^{-3}	9.14×10^{-4}
σ_3	3.40×10^{-2}	2.33×10^{-7}	σ_3	4.26×10^{-2}	3.21×10^{-3}

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.[12]

We highlight two observations from table 5.8. First, the nonlinear model performs much better matching the data than the linearized model. 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 (23 percent), higher investment (43 percent) and higher hours worked (20 percent) 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

[12]The moments associated with each set of parameter values are nearly identical if we simulate the model using the linear or the nonlinear solution method. See Aruoba, Fernández-Villaverde and Rubio-Ramírez (2003) for details.

deviation of the economy are also important. The nonlinear economy is less volatile than the linearized model in terms of the standard deviation of output (0.087 versus 0.121) although the hours worked respond more to productivity shocks (standard deviation of 0.02 versus 0.007).

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	0.087	1.55	0.121
hours	0.36	0.015	0.36	0.020	0.30	0.007
inv	0.42	0.066	0.44	0.066	0.31	0.076

Finally 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 59. The differences in moments discussed above are one of the main driving force behind the finding. A second force is that, as in the case of artificial data, the likelihood function generated by the Sequential Monte Carlo is more concentrated than the one coming from the Kalman filter.

Table 5.9: Logmarginal Likelihood Difference Real Data

p	Nonlinear vs. Linear
0.1	58.78
0.3	58.86
0.5	58.98
0.7	59.03
0.9	59.05

6. Conclusions

We have compared the effects of estimating dynamic equilibrium models using a Sequential Monte Carlo filter proposed by Fernández-Villaverde 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 stochastic 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 relatively 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. Some papers have documented that nonnormalities may be important to account for the dynamics of macro data (see Geweke, 1994 among others). Future research will address how much accuracy is gained with the use of a Sequential Monte Carlo filter when estimating models with nonnormal innovations.

7. Appendix

This appendix presents further details about the computations in the paper. First it explains the finite element method. Second, it does the same for the undetermined coefficients method. Third, it describes the computation of the marginal likelihood. Finally, it comments on the programming language, hardware, and software used.

7.1. The Finite Element Method

We provide a brief exposition of the finite element method as applied in the paper. For a more detailed explanation the interested reader should consult the expositions in McGrattan (1999) and Aruoba, Fernández-Villaverde and Rubio-Ramírez (2003).

The first step in the finite element method is to note that we can rewrite the Euler equation for consumption as

$$U_c(k_t, z_t) = \frac{\beta}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \left[U_c(k_{t+1}, z_{t+1})(1 + \alpha e^{z_{t+1}} k_{t+1}^{\alpha-1} l(k_{t+1}, z_{t+1})^{1-\alpha} - \delta) \right] \exp(-\frac{\epsilon_{t+1}^2}{2\sigma^2}) d\epsilon_{t+1} \tag{14}$$

where $U_c(t) = U_c(k_t, z_t)$, $k_{t+1} = e^{z_{t+1}} k_t^{\alpha} l_t^{1-\alpha} + (1-\delta)k_t - c(k_t, z_t)$ and $z_{t+1} = \rho z_t + \epsilon_{t+1}$.

The problem is to find two policy functions $c(k, z) : R^+ \times [0, \infty] \to R^+$ and $l(k, z) : R^+ \times [0, \infty] \to [0, 1]$ that satisfy the model equilibrium conditions. Since the static first order condition gives a relation between the two policy functions, we only need to solve for one of them. For the rest of the exposition we will assume that we actually solve for $l(k, z)$ and then we find $c\left(l(k, z)\right)$.

First we bound the domain of the state variables to partition it in nonintersecting elements. To bound the productivity level of the economy we define $\lambda_t = \tanh(z_t)$. Since $\lambda_t \in [-1, 1]$ we can write the stochastic process as $\lambda_t = \tanh(\rho \tanh^{-1}(z_{t-1}) + \sqrt{2}\sigma v_t)$ where $v_t = \frac{\epsilon_t}{\sqrt{2}\sigma}$. Now, since $\exp(\tanh^{-1}(z_{t-1})) = \frac{\sqrt{1+\lambda_{t+1}}}{\sqrt{1-\lambda_{t+1}}} = \widehat{\lambda}_{t+1}$, we rewrite (14) as

$$U_c(t) = \frac{\beta}{\sqrt{\pi}} \int_{-1}^{1} \left[U_c(k_{t+1}, z_{t+1}) \left(1 + \alpha \widehat{\lambda}_{t+1} k_{t+1}^{\alpha-1} l(k_{t+1}, z_{t+1})^{1-\alpha} + \delta \right) \right] \exp(-v_{t+1}^2) dv_{t+1} \tag{15}$$

where $k_{t+1} = \widehat{\lambda}_{t+1} k_t^{\alpha} l\left(k_t, z_t\right)^{1-\alpha} + (1-\delta)k_t - c\left(l(k_t, z_t)\right)$ and $z_{t+1} = \tanh(\rho \tanh^{-1}(z_t) +$

$\sqrt{2}\sigma v_{t+1}$). For convenience we follow the same notation for $l(\cdot)$ in both (14) and (15), although they are not the same function since their domain is different. To bound the capital, we fix an upper bound $\overline{k}$, picked sufficiently high as a function of the steady state of the model that it will bind only with an extremely low probability.

Then define $\Omega = \left[0, \overline{k}\right] \times [-1, 1]$ as the domain of $l_{fe}(k, z; \overline{\theta})$ and divide Ω into nonoverlapping rectangles $[k_i, k_{i+1}] \times [z_j, z_{j+1}]$, where k_i is the ith grid point for capital and z_j is jth grid point for the technology shock. Clearly $\Omega = \cup_{i,j} [k_i, k_{i+1}] \times [z_j, z_{j+1}]$. These elements may be of unequal size. In our computations we define 14 unequal elements in the capital dimension and 10 on the λ axis. We have small elements in the areas of Ω where the economy spends most of the time while just a few large elements cover wide areas of the state space infrequently visited (see figure A.1 for our partition). Note that we define the elements in relation to the level of capital in the steady state of the model for each particular value of the parameters being used in that precise moment of the estimation. Consequently our mesh is endogenous to the estimation procedure, increasing efficiency and accuracy.

Next, we set $l_{fe}\left(k, z; \overline{\theta}\right) = \sum_{i,j} \overline{\theta}_{ij} \Psi_{ij}(k, z) = \sum_{i,j} \overline{\theta}_{ij} \widehat{\Psi}_i(k) \widetilde{\Psi}_j(z)$, where

$$\widehat{\Psi}_i(k) = \begin{cases} \frac{k - k_i}{k_{i+1} - k_i} & \text{if } k \in [k_{i-1}, k_i] \\ \frac{k_{i+1} - k}{k_{i+1} - k_i} & \text{if } k \in [k_i, k_{i+1}] \\ 0 \text{ elsewhere} \end{cases} \qquad \widetilde{\Psi}_j(z) = \begin{cases} \frac{z - z_j}{z_{j+1} - z_j} & \text{if } z \in [z_{j-1}, z_j] \\ \frac{z_{j+1} - z}{z_{j+1} - z_j} & \text{if } z \in [z_j, z_{j+1}] \\ 0 \text{ elsewhere} \end{cases}.$$

Note that $\Psi_{ij}(k, z) = 0$ if $(k, z) \notin [k_{i-1}, k_i] \times [z_{j-1}, z_j] \cup [k_i, k_{i+1}] \times [z_j, z_{j+1}]$ $\forall i, j$, i.e., the function is 0 everywhere except inside two elements. Also $l_{fe}(k_i, z_j; \overline{\theta}) = \overline{\theta}_{ij}$ $\forall i, j$, i.e., the values of $\overline{\theta}$ specify the values of c_{fe} at the corners of each subinterval $[k_i, k_{i+1}] \times [z_j, z_{j+1}]$.

Let us define $U_c(k_{t+1}, z_{t+1})_{fe}$ as the marginal utility of consumption evaluated at the finite element approximation values of consumption and leisure. In this case, from the Euler equation we have a residual equation:

$$R(k_t, z_t; \theta) = \frac{\beta}{\sqrt{\pi}} \int_{-1}^{1} \left[\frac{U_c(k_{t+1}, z_{t+1})_{fe}}{U_c(k_{t+1}, z_{t+1})_{fe}} \left(1 + \alpha \widehat{\lambda}_{t+1} k_{t+1}^{\alpha - 1} l_{fe}^{1-\alpha} - \delta \right) \right] \exp(-v_{t+1}^2) dv_{t+1} - 1 \quad (16)$$

A Galerkin scheme implies that we weight the residual function by the basis functions and solve the system of $\overline{\theta}$ equations

$$\int_{\left[0, \overline{k}\right] \times [-1,1]} \Psi_{i,j}(k, z) R(k, z; \overline{\theta}) dz dk = 0 \quad \forall i, j \quad (17)$$

on the $\overline{\theta}$ unknowns.

Since $\Psi_{ij}(k,z)=0$ if $(k,z)\notin[k_{i-1},k_i]\times[z_{j-1},z_j]\cup[k_i,k_{i+1}]\times[z_j,z_{j+1}]\ \forall i,j$ we can rewrite (17) as:

$$\int_{[k_{i-1},k_i]\times[z_{j-1},z_j]\cup[k_i,k_{i+1}]\times[z_j,z_{j+1}]}\Psi_{i,j}(k,z)\,R(k,z;\overline{\theta})dzdk=0\quad\forall i,j. \tag{18}$$

We evaluate the integral in the residual equation with a Gauss-Hermite method and the integrals in (18) with a Gauss-Legendre procedure. Finally, we solve the associated system of nonlinear equations with a quasi-Newton algorithm with a conservative update to avoid numerical instabilities.

7.2. Undetermined Coefficients Method

We provide a brief exposition of the method as applied in the paper. First we find the deterministic steady state of the model: $k_{ss}(\gamma)=\frac{\Psi}{\Omega+\varphi\Psi}$, $l_{ss}(\gamma)=\varphi k_{ss}(\gamma), c_{ss}(\gamma)=\Omega k_{ss}(\gamma)$ and $y_{ss}(\gamma)=k_{ss}(\gamma)^{\alpha}l_{ss}(\gamma)^{1-\alpha}$ where $\varphi=\left(\frac{1}{\alpha}\left(\frac{1}{\beta}-1+\delta\right)\right)^{\frac{1}{1-\alpha}}$, $\Omega=\varphi^{\frac{1}{\alpha}}-\delta$ and $\Psi=\frac{\theta}{1-\theta}(1-\alpha)\varphi^{-\alpha}$. From this point on, we will not make explicit the dependence of the steady-state value on γ, but it should be understood that $k_{ss}=k_{ss}(\gamma)$, $l_{ss}=l_{ss}(\gamma), c_{ss}=c_{ss}(\gamma)$ and $y_{ss}=y_{ss}(\gamma)$.

If we linearize the set of equilibrium conditions around the deterministic steady state:

$$\alpha_1(c_t-c_{ss})+\alpha_2(l_t-l_{ss})=E_t\left\{\alpha_1(c_{t+1}-c_{ss})+\alpha_3(l_{t+1}-l_{ss})+\alpha_4 z_{t+1}+\alpha_5(k_{t+1}-k_{ss})\right\}$$
$$(c_t-c_{ss})=c_{ss}z_t+\frac{\alpha}{k_{ss}}c_{ss}(k_t-k_{ss})+\alpha_6(l_t-l_{ss})$$
$$(c_t-c_{ss})+(k_{t+1}-k_{ss})=y_{ss}z_t+y_{ss}\frac{\alpha}{k_{ss}}(k_t-k_{ss})+\alpha_7(l_t-l_{ss})+(1-\delta)(k_t-k_{ss})$$
$$z_t=\rho z_{t-1}+\varepsilon_t$$

where

$$\begin{array}{ll}
\alpha_1=\frac{\theta(1-\tau)-1}{c_{ss}} & \alpha_2=-\frac{(1-\tau)(1-\theta)}{1-l_{ss}}\\
\alpha_3=\beta\frac{\alpha(1-\alpha)}{l_{ss}}k_{ss}^{\alpha-1}l_{ss}^{1-\alpha}-\frac{(1-\tau)(1-\theta)}{1-l_{ss}} & \alpha_4=\alpha\beta k_{ss}^{\alpha-1}l_{ss}^{1-\alpha}\\
\alpha_5=\beta\frac{\alpha(\alpha-1)}{k_{ss}}k_{ss}^{\alpha-1}l_{ss}^{1-\alpha} & \alpha_6=-\left(\frac{\alpha}{l_{ss}}+\frac{1}{(1-l_{ss})}\right)c_{ss}\\
\alpha_7=y_{ss}\frac{1-\alpha}{l_{ss}} & y_{ss}=k_{ss}^{\alpha}l_{ss}^{1-\alpha}
\end{array}$$

We group terms to eliminate one of the equations of the system and obtain the simpler system:

$$A\widehat{k}_{t+1} + B\widehat{k}_t + C\widehat{l}_t + Dz_t = 0$$
$$E_t\left(G\widehat{k}_{t+1} + H\widehat{k}_t + J\widehat{l}_{t+1} + K\widehat{l}_t + Lz_{t+1} + Mz_t\right) = 0$$
$$E_t z_{t+1} = Nz_t$$

where $A = 1$, $B = \frac{\alpha}{k_{ss}}c_{ss} - y_{ss}\frac{\alpha}{k_{ss}} - (1-\delta)$, $C = \alpha_6 - \alpha_7$, $D = c_{ss} - y_{ss}$, $G = \alpha_1\frac{\alpha}{k_{ss}}c_{ss} + \alpha_5$, $H = -\alpha_1\frac{\alpha}{k_{ss}}c_{ss}$, $J = \alpha_1\alpha_6 + \alpha_3$, $K = -(\alpha_1\alpha_6 + \alpha_2)$, $L = (\alpha_1 c_{ss} + \alpha_4)$, $M = -\alpha_1 c_{ss}$, $N = \rho$ and $\widehat{x}_t = x_t - x_{ss}$.

Now we can guess policy functions of the form $\widehat{k}_{t+1} = P\widehat{k}_t + Qz_t$ and $\widehat{l}_t = R\widehat{k}_t + Sz_t$, plug them into the linearized equilibrium conditions and solve for P, Q, R and S:[13]

$$R = -\frac{1}{C}(AP + B) = -\frac{1}{C}AP - \frac{1}{C}B$$
$$P = -\frac{1}{2}\left(-\left(\frac{B}{A} + \frac{K}{J} - \frac{GC}{JA}\right) \pm \sqrt{\left(\frac{B}{A} + \frac{K}{J} - \frac{GC}{JA}\right)^2 - 4\left(\frac{KB - HC}{JA}\right)}\right)$$
$$Q = \frac{-D(JN + K) + CLN + CM}{AJN + AK - CG - CJR}$$
$$S = \frac{-ALN - AM + DG + DJR}{AJN + AK - CG - CJR}.$$

In the notation used in section 3.4, we have:

$$\begin{array}{ll} a_{11}(\gamma) = P & a_{12}(\gamma) = Q \\ a_{21}(\gamma) = R & a_{22}(\gamma) = S \\ a_{31}(\gamma) = \frac{\alpha}{k_{ss}}c_{ss} + \alpha_6 a_{21}(\gamma) & a_{32}(\gamma) = c_{ss} + \alpha_6 a_{22}(\gamma) \\ a_{41}(\gamma) = \frac{\alpha}{k_{ss}}y_{ss} + \alpha_7 a_{21}(\gamma) & a_{42}(\gamma) = y_{ss} + \alpha_7 a_{22}(\gamma) \end{array}$$

[13] Since the equation defining P is quadratic we will have two possible solutions. Of course we pick the stable root of P.

7.3. The Posterior and the Marginal Likelihood

In genera,l we do not have a closed-form solution for the posterior distribution. Instead, we draw a sample of size M, $\{\gamma_1, \gamma_2, .., \gamma_M\}$, using a random walk Metropolis-Hastings algorithm.

We use the draw to estimate the marginal likelihood. This marginal likelihood determines the probability the model assigns to the observations and serves to compare models.

Following Gelfand and Dey (1994) note that for any density $h(\cdot)$ with support contained in Υ:

$$p\left(y^T\right)^{-1} = E_{L(y^T;\gamma)\pi(\gamma)}\left[\left.\frac{h(\gamma)}{L\left(y^T;\gamma\right)\pi(\gamma)}\right| y^T\right]$$

This expression is an unbiased and consistent estimator of the marginal likelihood and satisfies a central limit theorem if $\frac{\int_\Upsilon h^2(\gamma)d\gamma}{\int_\Upsilon L(y^T;\gamma)\pi(\gamma)d\gamma} < \infty$.

Then, we can use $\{\gamma_1, \gamma_2, .., \gamma_M\}$ to approximate the marginal likelihood as:

$$p\left(y^T\right)^{-1} \simeq \frac{1}{M}\sum_{i=1}^{M}\frac{h(\gamma_i)}{L\left(y^T;\gamma_i\right)\pi(\gamma_i)}$$

As a choice of h, we use Geweke's (1998) proposal. First, we compute $\widehat{\gamma} = \frac{1}{M}\sum_{i=1}^{M}\gamma_i$ and

$$\widehat{\Sigma_M} = \frac{1}{M}\sum_{i=1}^{M}\left(\gamma_i - \widehat{\gamma}\right)\left(\gamma_i - \widehat{\gamma}\right)'$$

Then, for a given $p \in (0,1)$ define the set $\Upsilon_M = \left\{\gamma : \left(\gamma - \widehat{\gamma}\right)\widehat{\Sigma_m}^{-1}\left(\gamma - \widehat{\gamma}\right)' \leq \chi^2_{1-p}(10)\right\}$ where $\chi^2_{1-p}(\cdot)$ is a chi-squared distribution with degrees of freedom equal to the number of parameters. Letting $I_{\Upsilon_M \cap \Upsilon}(\cdot)$ be the indicator function of a vector of parameters belonging to the intersection $\Upsilon_M \cap \Upsilon$, we define:

$$h(\gamma) = \frac{1}{\widehat{p}\,(2\pi)^{\frac{k}{2}}}\left|\widehat{\Sigma_m}\right|^{\frac{1}{2}} e^{-0.5*(\gamma-\widehat{\gamma})\widehat{\Sigma_m}^{-1}(\gamma-\widehat{\gamma})'} I_{\Upsilon_M\cap\Upsilon}(\gamma)$$

where $\widehat{p}$ is an appropriate normalizing constant. With this choice, if the posterior density is uniformly bounded away from zero on every compact subset of Υ, our computation approximates the marginal likelihood.

7.4. Computational Details

All programs needed for the computation of the model were coded in Fortran 95 and compiled in Compaq Visual Fortran 6.6 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 40,000 particles takes around 6.1 seconds. That implies a total of about 88 hours for each simulation of 50,000 draws. To put this number in perspective, note that the whole simulation from the linearized version of the model runs in one minute. All the code is available upon request from the corresponding author.

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] Evett, I.W. (1991). "Implementing Bayesian Methods in Forensic Science", paper presented at the Fourth Valencia International Meeting on Bayesian Statistics.

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

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

[11] Gelfand, A.E. and D.K. Dey (1994). "Bayesian Model Choice: Asymptotics and Exact Calculations". *Journal of the Royal Statistical Society Series B* 56, 501-514.

[12] Gelman, A., G.O. Roberts and W.R. Gilks (1996). "Efficiente Metropolis Jumping Rules" In J.O. Berger *et al.* (eds.) *Bayesian Statistics 5*, Oxford University Press.

[13] Geweke, J. (1994). "Priors for Macroeconomic Time Series and their Applications". *Econometric Theory* 10, 609-632.

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

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

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

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

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

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

[20] 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.

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

[22] Ljungqvist, L. and T.J. Sargent (2000). *Recursive Macroeconomic Theory.* MIT Press.

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

[24] McGrattan, E.R. (1999). "Application of Weighted Residuals Methods to Dynamic Economic Models" in R. Marimón and A. Scott (eds.) *Computational Methods for the Study of Dynamic Economies.* Oxford University Press.

[25] 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.

[26] McGrattan, E. and E.C. Prescott (2000). "Is the Stock Market Overvalued?". *Mimeo*, Federal Reserve Bank of Minneapolis.

[27] 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.

[28] 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.

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

[30] 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.*

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

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

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

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

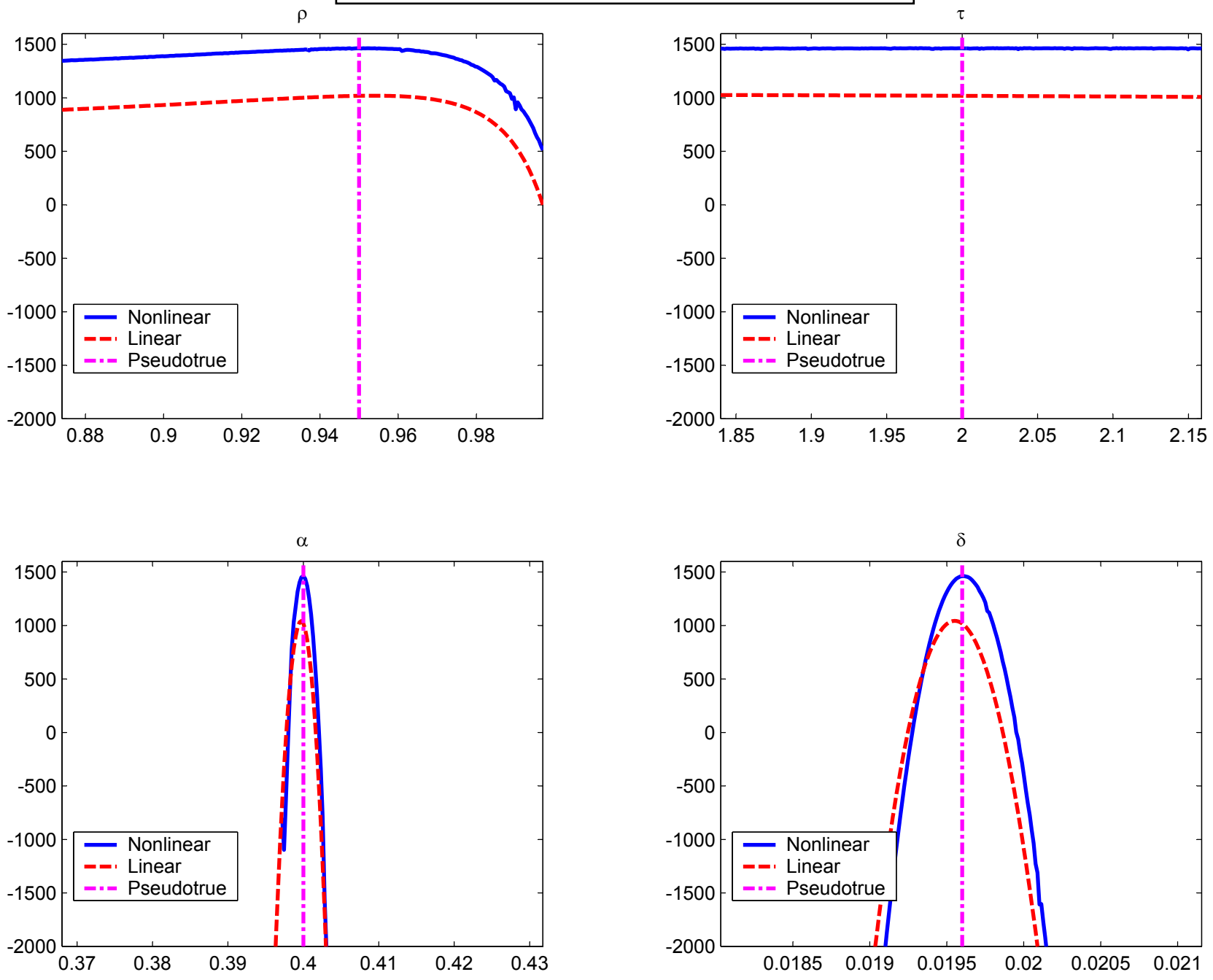
[35] Smets, F. and R. Wouters (2003a). "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area". *Nationale Bank van Belgie Working Paper No. 35.*

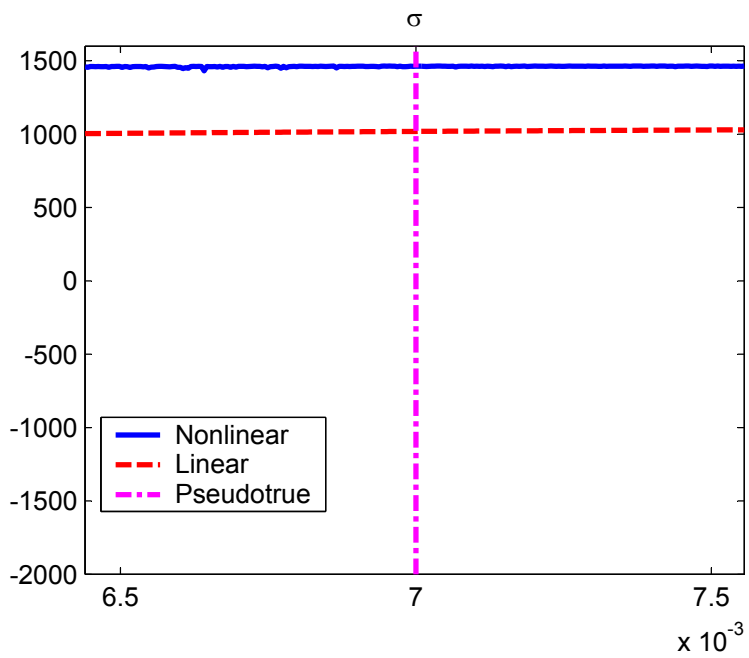
[36] Smets, F. and R. Wouters (2003b). "Shocks and Frictions in US Business Cycle Fluctuations: A Bayesian DSGE Approach". *Mimeo,* European Central Bank.

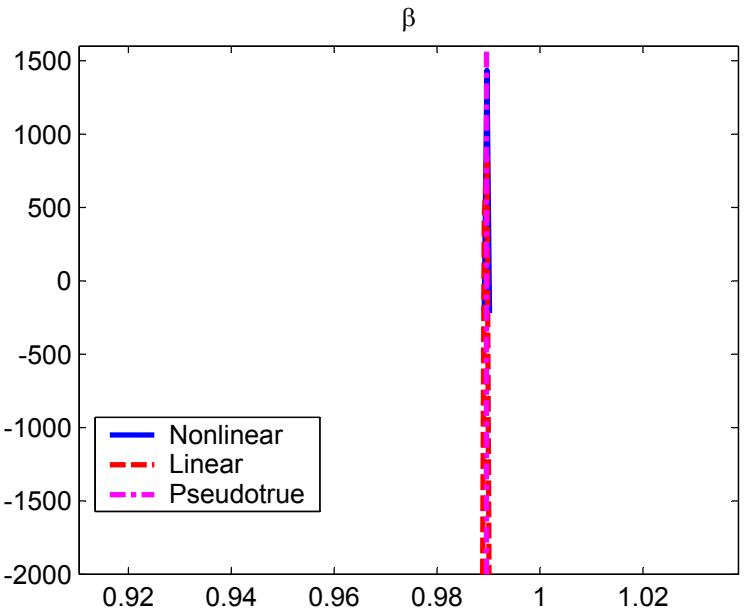
[37] Uhlig, H. (1999), "A Toolkit for Analyzing Nonlinear Dynamic Stochastic Models Easily" in R. Marimón and A. Scott (eds) *Computational Methods for the Study of Dynamic Economies.* Oxford University Press.

[38] 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




σ
1500
1000
500
0
-500
-1000
-1500
-2000
6.5
7
7.5
x 10^-3
Nonlinear
Linear
Pseudotrue


β
1500
1000
500
0
-500
-1000
-1500
-2000
0.92
0.94
0.96
0.98
1
1.02
Nonlinear
Linear
Pseudotrue

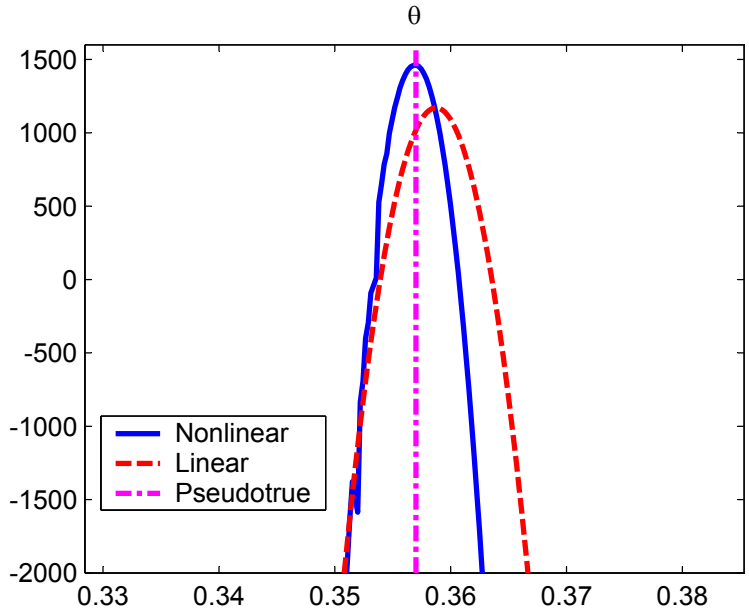
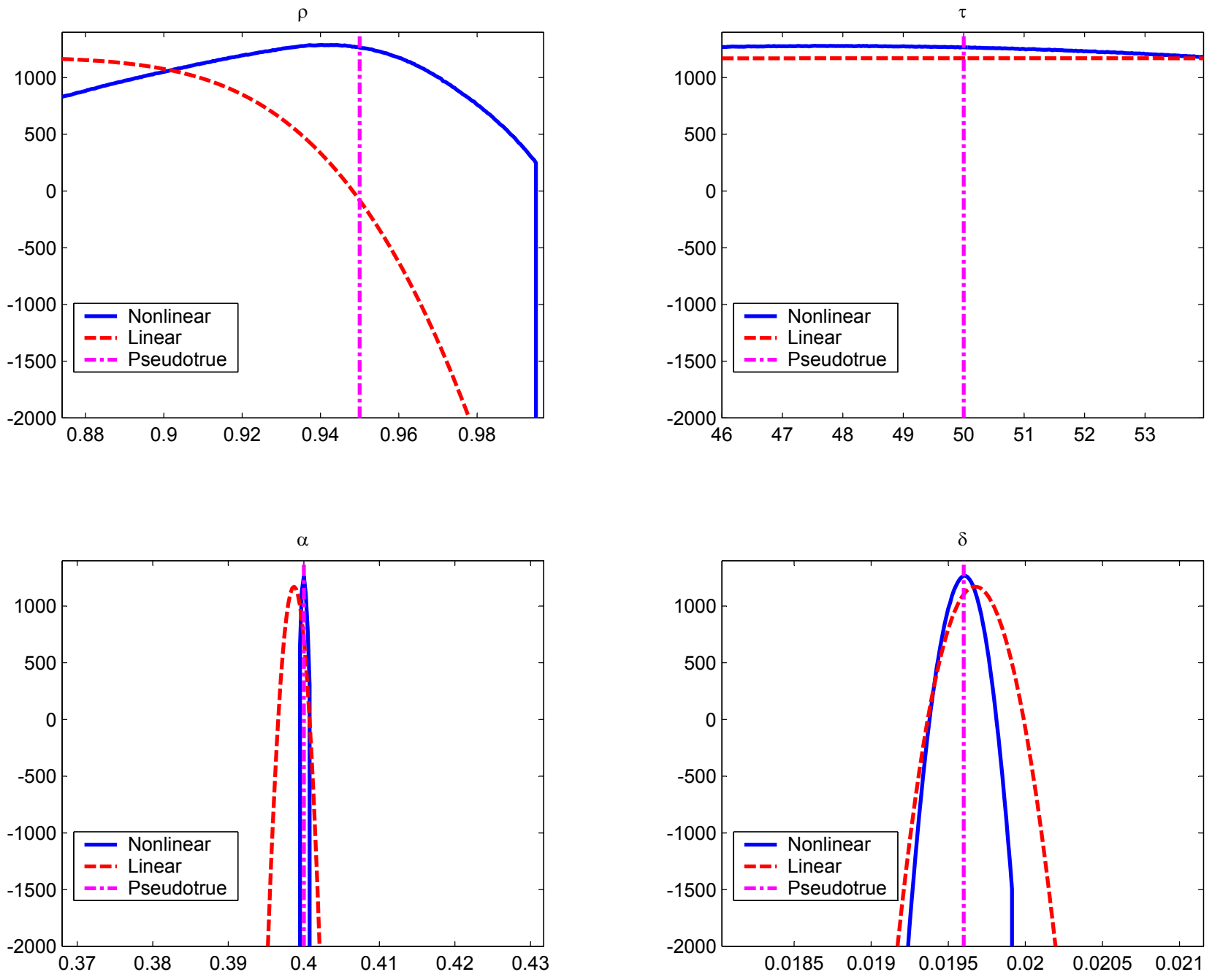
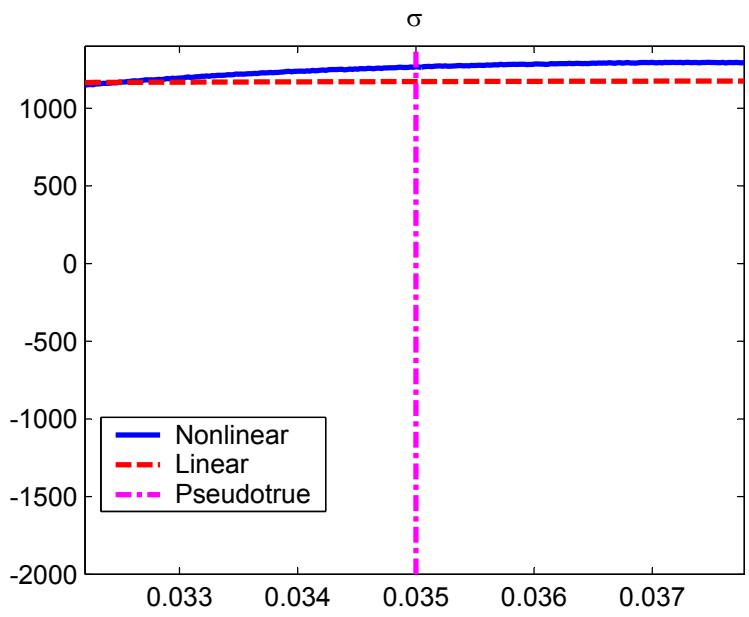
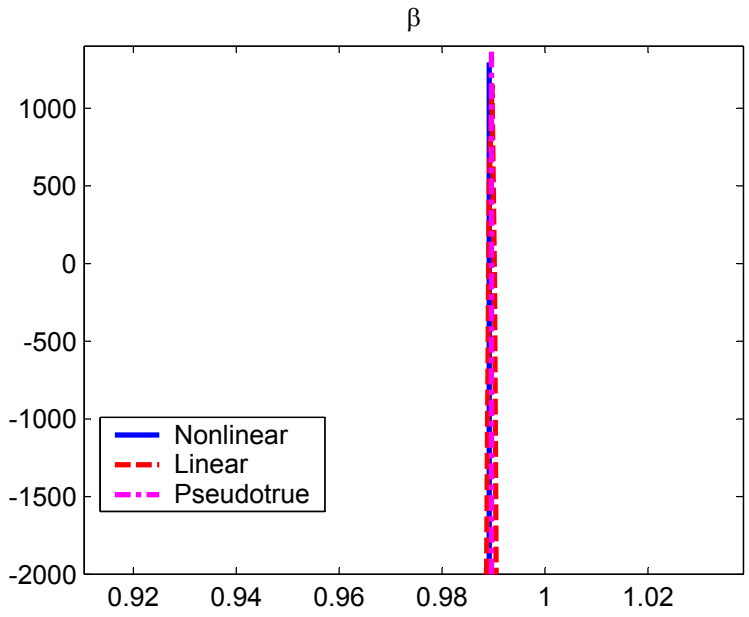
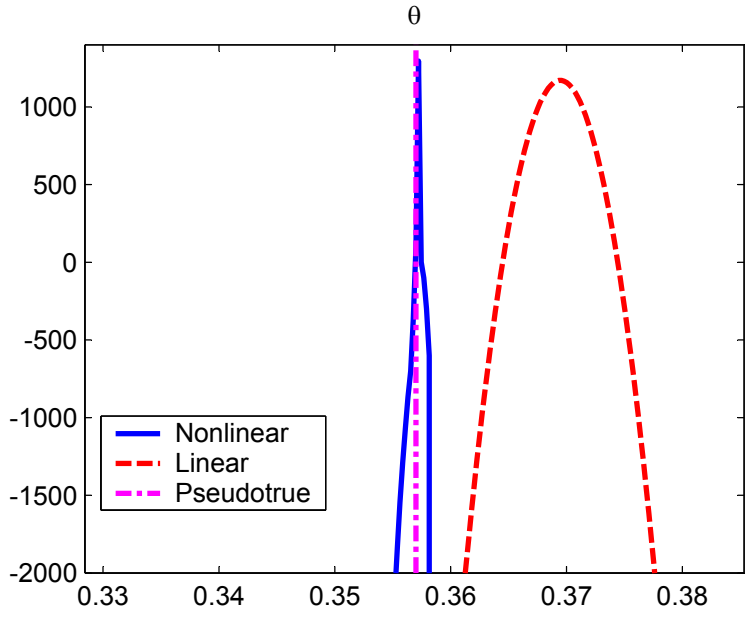

θ
1500
1000
500
0
-500
-1000
-1500
-2000
0.33
0.34
0.35
0.36
0.37
0.38
Nonlinear
Linear
Pseudotrue

Figure 5.2: Likelihood Function Extreme Calibration




σ
1000
500
0
-500
-1000
-1500
-2000
0.033
0.034
0.035
0.036
0.037
Nonlinear
Linear
Pseudotrue


β
1000
500
0
-500
-1000
-1500
-2000
0.92
0.94
0.96
0.98
1
1.02
Nonlinear
Linear
Pseudotrue


θ
1000
500
0
-500
-1000
-1500
-2000
0.33
0.34
0.35
0.36
0.37
0.38
Nonlinear
Linear
Pseudotrue