Estimating Nonlinear Dynamic Equilibrium Economies: A Likelihood Approach*

Jesús Fernández-Villaverde — University of Pennsylvania

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

January 6, 2004

Abstract

This paper presents a framework to undertake likelihood-based inference in nonlinear dynamic equilibrium economies. We develop a Sequential Monte Carlo algorithm that delivers an estimate of the likelihood function of the model using simulation methods. This likelihood can be used for parameter estimation and for model comparison. The algorithm can deal both with nonlinearities of the economy and with the presence of non-normal shocks. We show consistency of the estimate and its good performance in finite simulations. This new algorithm is important because the existing empirical literature that wanted to follow a likelihood approach was limited to the estimation of linear models with Gaussian innovations. We apply our procedure to estimate the structural parameters of the neoclassical growth model.

*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 Manuel Arellano, Will Roberds, Tao Zha, and participants at several seminars for useful comments. Valuable assistance was provided by University of Minnesota Supercomputer Institute. Beyond the usual disclaimer, we 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

This paper presents a method to undertake likelihood-based inference in nonlinear and/or non-normal dynamic equilibrium models. We show how we can use Sequential Monte Carlo methods to estimate the structural parameters of the model, those describing preferences and technology, and to compare different economies. Both tasks can be implemented from either a Bayesian perspective or a classical one.

Economists now routinely use dynamic general equilibrium economies to answer quantitative questions. To estimate these economies, the empirical literature has been forced to use either limited-information moment methods (in any of their different versions) or likelihood techniques on linearized versions of the model. This situation is unsatisfactory. Moment methods may suffer strong biases resulting from using small samples and may not use efficiently all the existing information. Linearization techniques depend crucially on the linear relation's accurate approximation of the true policy function's shape as well on the presence of normal shocks.

The main obstacle to a more standard likelihood-based inference is the difficulty in evaluating the likelihood function implied by a nonlinear and/or non-normal dynamic equilibrium economy. Beyond a few particular cases,[1] it is not possible to evaluate this function. Moment methods avoid the problem by moving away from full information approaches to inference. Linearization renounces evaluating the true likelihood function of the model and concentrates instead on the likelihood of a more tractable linear approximation to the economy.

We propose a Sequential Monte Carlo method to solve this problem. We describe how the technique can be applied to evaluate the likelihood function implied by the nonlinear solution of a dynamic equilibrium economy even if the driving shocks of the model are non-normal (although the algorithm is general enough to handle linear models with or without normal shocks).

To do so we borrow from a growing body of literature on nonlinear filtering (see the seminal paper by Gordon, Salmond and Smith, 1993 and the review in Doucet, de Freitas and Gordon, 2001). We adapt this know-how to deal with the likelihood functions of dynamic equilibrium models and we show how we get accurate and stable evaluations of the likelihood function. With these evaluations available, the door for likelihood-based inference opens, either by searching for a maximum of the function (quasi-maximum likelihood estimation) or by simulating the posterior distribution of the parameters using a Markov chain Monte Carlo algorithm (Bayesian estimation).

[1] Some of these cases are, however, important. For example there exist a popular literature on the estimation of dynamic discrete choice models that uses maximum likelihood methods. See Rust (1994) for a survey.

The general idea of the procedure follows. First, for a given set of parameter values, we compute the equilibrium policy functions of the model. Since we want to conduct inference in the nonlinear model, we rely on a nonlinear solution method to find the policy functions. With the policy functions we construct the state space representation of the model. Under certain mild conditions, we use this state space form and a Sequential Monte Carlo scheme to evaluate the likelihood function. Plugging this likelihood evaluation algorithm into an optimization or a Markov chain Monte Carlo routine we search the parameter space to perform likelihood-based inference, by either maximizing the likelihood function or, after specifying some priors on the parameters, finding posterior distributions. Finally, if we apply the algorithm to several models, we can compare models using the output of the model by building either likelihood ratios (Voung, 1989) or Bayes factors (Geweke, 1998), even if the models are misspecified and nonnested.

To illustrate our method we compute and estimate the benchmark dynamic equilibrium economy, the stochastic neoclassical growth model. After we solve the model nonlinearly, we estimate it using both Bayesian and maximum likelihood methods and we perform Monte Carlo analysis to evaluate the efficiency of our procedure.

Being able to perform likelihood-based inference is important for several reasons. From a theoretical perspective, the likelihood principle states that all the empirical evidence obtained from the data is contained in the likelihood function (Berger and Wolpert, 1988). From an applied position, likelihood-based inference is a simple way to deal with misspecified models (Monfort, 1996). Dynamic equilibrium economies are false by construction, and likelihood-based inference has both attractive asymptotic properties and good small-sample behavior even when models are misspecified (White, 1994 for quasi-maximum likelihood and Fernández-Villaverde and Rubio-Ramírez, 2003a, for Bayesian procedures). Finally, and for us the most compelling reason, likelihood inference allows us to compare models. Of course we do not want to imply that a likelihood approach is always preferable. For example we may only care about accounting for one particular dimension of the data, a task for which a moment method can be more suitable. We simply maintain that in numerous contexts, the likelihood function is an informative tool.

Our paper builds on the existing literature dealing with inference on dynamic equilibrium economies. Hansen's (1982) pioneered the use of moments methods, now widely applied.[2] Sargent (1989) uses the Kalman filter to evaluate the likelihood function of linear or linearized dynamic equilibrium economies with normal shocks. Altuğ (1989), also in a linear

[2]Variations include the Simulated Method of Moments (Lee and Ingramm 1991), the Efficient Method of Moments (Gallant and Tauchen, 1996), Indirect Inference (Gourieroux, Monfort and Renault, 1993 and Smith, 1993) and several information-based approaches (Kitamura and Stutzer, 1997 and Imbens, Spady and Johnson, 1998). We refer the reader to the special issue of the *Journal of Business and Economic Statistics* on the Generalized Method of Moments (Ghysels and Hall, 2002) for an overview of the literature.

framework, proposed to estimate the likelihood in the frequency domain. This spectral approach has been followed by Diebold, Ohanian and Berkowitz (1998) for estimation and by Watson (1993) to compare models with data. Christiano, Eichenbaum and Evans (2001) estimate dynamic equilibrium economies using the information in impulse-response functions of linearized solutions. Miranda and Rui (1997) exploit the structure of an asset pricing model to find the nonlinear likelihood function. Their method, however, involves the computation of the Jacobian of a transformation of variables and it is difficult to generalize. From a Bayesian perspective, DeJong, Ingram and Whiteman (2000) pioneered the Bayesian estimation of Real Business Cycles models using importance sampling. Landon-Lane (1999) and Otrok (2001) first applied the Markov chain Monte Carlo methods. Schorfheide (2000) formulates the impulse-response approach in the Bayesian framework. All those papers, though, stay within the linear framework. We also build on the contributions of the literature on nonlinear filtering. We discussion that literature in section 2.3.

The rest of the paper is organized as follows. In the next section we describe our framework to evaluate the likelihood function of the dynamic general equilibrium model for a given set of parameter values. Section 3 presents the stochastic neoclassical growth model and discusses how we can apply our Sequential Monte Carlo to it. Section 4 proposes estimation algorithms and section 5 reports findings both with simulated and with real data. Section 6 concludes. An appendix discusses computational details.

2. A Framework for Evaluating the Likelihood

In this section we develop a general framework to estimate and compare a large class of nonlinear dynamic equilibrium models using a likelihood approach. Examples of economies in this class are the stochastic neoclassical growth model (Cooley and Prescott, 1995), sticky prices models (Chari, Kehoe and McGrattan, 2000, Rotemberg and Woodford, 1997 and Woodford, 2003), asset pricing models (Mehra and Prescott, 1985), macro public finance models (Chari, Christiano and Kehoe, 1994) and regime switching models (Quadrini and Jermann, 2003), among many others.

All of these economies imply a different joint probability distribution function for observables given the model's structural parameters which describe preferences and technology. We refer to this density the likelihood function of the economy. The likelihood function is useful for two purposes. First, if we want to perform estimation, we can use an optimization routine to find the parameter values that maximize it or, if we specify a prior for the structural parameters, a Markov chain Monte Carlo to draw from the posterior. Second, if we are comparing several models, we can do so by building either likelihood ratios (Voung, 1989) or Bayes factors (Geweke, 1998).

The literature shows how to write the likelihood function of a dynamic equilibrium economy only in a few special cases. For example, we can evaluate the likelihood of a linear model with normal innovations using the Kalman filter. Unfortunately there is no general procedure for writing an analytic expression for this likelihood. As we discussed in the introduction, this problem has been a stumbling block to the application of likelihood-based methods to perform inference in dynamic equilibrium economies.

This section presents a Sequential Monte Carlo method to address the problem of evaluating the likelihood function of a nonlinear dynamic equilibrium economy. The rest of the section is organized as follows. First, we define the likelihood function of a dynamic equilibrium economy. Second, we present a simulation filter to evaluate that likelihood. We finish by comparing our approach with some alternatives.

2.1. The Likelihood Function of a Dynamic Equilibrium Economy

A large set of dynamic equilibrium models can be written in the following state space form. First the equilibrium of economy is characterized by some states S_t that change over time according to the following transition equation:

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

where $\{W_t\}$ is a sequence of exogenous independent random variables and $\gamma \in \Upsilon$ is the vector of structural parameters of the model.

Second the observables y_t are a realization of the random variable Y_t 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.[3] Along some dimension the function g can be the identity mapping if a state is directly observed without noise.

To summarize our notation: S_t are the states of the economy, W_t are the exogenous shocks that affect the states' law of motion, Y_t are the observables, and V_t are the exogenous perturbations that affect the observables but not the states.

The functions f and g come from the equations that describe the equilibrium of the model: policy functions, laws of motion for variables, resource and budget constraints, and so on. Dynamic equilibrium economies do not generally admit closed-form solutions for those functions. Our algorithm only requires a numerical procedure to approximate them.

[3] Assuming independence of $\{W_t\}$ and $\{V_t\}$ is only for notational convenience. Generalization to more involved structures for those stochastic processes is achieved by increasing the dimension of the state space.

To fix ideas, we now map $\{S_t\}$, $\{W_t\}$, $\{Y_t\}$, $\{V_t\}$, f and g into some examples of dynamic equilibrium economies. Consider first the example of the stochastic neoclassical growth model with leisure choice. The states of this economy are capital and the productivity level. Assume that our observables are output and labor supply but that labor supply is measured with some noise. Then S_t will be capital and productivity, W_t the shock to productivity, Y_t output and observed labor supply, V_t the measurement error of labor, f the policy function for capital and the law of motion for technology and g the production function plus the policy function for labor augmented by the measurement error. Consider also an economy with nominal rigidities in the form of overlapping contracts. This economy experiences both productivity and money growth shocks, and we observe output and inflation. Now the states S_t are the distribution of prices, capital, money and the productivity level, W_t includes the shocks to technology and money growth, Y_t is output and inflation, V_t is a degenerate distribution with mass at zero, f collects the policy functions for capital and prices as well as the laws of motion for technology and money growth and g is the aggregate supply function and the Phillips curve. Many more examples of dynamic economies can be fitted into this state space formulation.

To continue our analysis we make the following assumptions.

Assumption 1: $\dim(W_t) + \dim(V_t) \geq \dim(Y_t)$.

This assumption ensures that the model is not stochastically singular. We do not impose any restrictions on how those degrees of stochasticity are achieved.[4]

Assumption 2: We can partition $\{W_t\}$ into two 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. $\{W_{2,t}\}$ is a zero-dimensional sequence.[5]

Note that assumption 2 is in some sense implied by assumption 1 because with a slight abuse of notation we allow the dimension of any of the sequences $\{W_{1,t}\}$, $\{W_{2,t}\}$ and (V_t) to be zero.

[4] This paper does not contribute to the literature on how to solve the problem of stochastic singularity of dynamic equilibrium economies. Two routes are commonly used to fix this problem. One is to reduce the observables accounted for to the number of stochastic shocks present. This likelihood can be studied to evaluate the model (Landon-Lane, 1999) or to find posteriors for parameters or impulse response functions (Schorfheide, 2000). The second route, increasingly popular, is to fully specify a model rich in stochastic dynamics (for example, Smets and Wouters, 2003a and 2003b). This alternative is attractive to address practical policy questions like those of interest for Central Banks.

[5] Alternatively we could consider this more general alternative **Assumption 2a**: We can partition $\{W_t\}$ into two sequences $\{W_{1,t}\}$ and $\{W_{2,t}\}$, such that $W_t = (W_{1,t}, W_{2,t})$ and $\dim(W_{2,t}) + \dim(V_t) \geq \dim(Y_t)$. If $\dim(V_t) \geq \dim(Y_t)$ we set $W_{1,t} = W_t$ $\forall t$, i.e. $\{W_{2,t}\}$ is a zero-dimensional sequence.

The main structure of the algorithm would not change but notation will be heavier.

Assumption 3: We set $W_{2,t} = W_t \ \forall t$, i.e. $\{W_{1,t}\}$ is a zero-dimensional sequence only if $\dim(W_t) + \dim(V_t) = \dim(Y_t)$.

Assumption 3 is not really necessary, but it makes the implementation of the algorithm easier. As in the case of assumption 2 we can dispense with it at the price of heavier notation.

Now we make some definitions that will be useful in the rest of the paper. First, let $W_i^t = \{W_{i,m}\}_{m=1}^t$, and let w_i^t be a realization of the random variable W_i^t for $i = 1, 2$ and $\forall t$. Let $V^t = \{V_m\}_{m=1}^t$, and let v^t be a realization of the random variable V^t for $\forall t$. Let $S^t = \{S_m\}_{m=0}^t$, and let s^t be a realization of the random variable S^t for $\forall t$. Let $Y^t = \{Y_m\}_{m=1}^t$, and let y^t be a realization of the random variable Y^t for $\forall t$. We also define $W_i^0 = \{\emptyset\}$ and $y^0 = \{\emptyset\}$.

Our goal is to evaluate the likelihood function of the a sequence of realizations of the observable y^T at a particular parameter value γ:

$$L\left(y^T;\gamma\right) = p\left(y^T;\gamma\right). \tag{3}$$

Our first step is to factor the likelihood function as:

$$\begin{aligned} p\left(y^T;\gamma\right) &= \prod_{t=1}^{T} p\left(y_t|y^{t-1};\gamma\right) \\ &= \int \left(\prod_{t=1}^{T} \int p\left(y_t|W_1^t, y^{t-1}, S_0;\gamma\right) p\left(W_1^t|y^{t-1}, S_0;\gamma\right) dW_1^t\right) p\left(S_0;\gamma\right) dS_0, \end{aligned} \tag{4}$$

where S_0 is the initial state of the model, W_1^t is the history up to date t of $W_{1,t}$, and the p's represent the relevant densities.[6] To save on notation we assume herein that all the relevant Radon-Nykodim derivatives exist. Extending the exposition to the more general case is straightforward but cumbersome.

In general the factorized likelihood function (4) cannot be computed analytically. The Sequential Monte Carlo algorithm that we propose in the next subsection allows us to use simulation methods to estimate it. The basic idea of the our approach is as follows. First, we present a Sequential Monte Carlo algorithm to draw from $p\left(W_1^t|y^{t-1}, S_0;\gamma\right)$. Second, we use those draws to estimate (4) by Monte Carlo integration.

Before introducing the algorithm we need to make two additional technical assumptions.

Assumption 4: For any $\gamma \in \Upsilon$ and any y^t, we can evaluate the conditional densities $p\left(y_t|W_1^t, y^{t-1}, S_0;\gamma\right)$ for $\forall t$.

[6] Where we understand that in the trivial case that $\{W_{1t}\}$ has zero dimensions $\int p\left(y_t|W_1^t, y^{t-1}, S_1;\gamma\right) p\left(W_1^t|y^{t-1}, S_1;\gamma\right) dW_1^t = p\left(y_t|y^{t-1}, S_1;\gamma\right)$, for all t.

Assumption 4 implies that for any realizations s_0, w_1^t and y^t of the random variables S_0, W_1^t and Y^t, we can evaluate the probability of the model described by (1) and (2) generating the observables. In other words, assumption 4 implies that for any s_0, w_1^t and y^t the following system of equations

$$\begin{aligned} S_1 &= f\left(s_0, \left(w_{1,1}, W_{2,1}\right); \gamma\right) \\ y_m &= g\left(S_m, V_m; \gamma\right) \qquad \text{for } m = 1, 2, ...t \\ S_m &= f\left(S_{m-1}, \left(w_{1,m}, W_{2,m}\right); \gamma\right) \qquad \text{for } m = 2, 3, ...t \end{aligned}$$

has a unique solution (v^t, s^t, w_2^t) and that we can evaluate the probabilities $p(v^t)$ and $p(w_2^t)$, where $p\left(y_t | W_1^t, y^{t-1}, S_0; \gamma\right) = p\left(v_t\right) p\left(w_{2,t}\right) \ \forall t$.

Assumption 4 rules out the possibilities of sunspots or indeterminacy of equilibrium. To deal with sunspots and indeterminacy we could extend the results of Lubick and Schorfheide (2003) which indexed the multiple solutions that appear under indeterminacy through additional parameters.

Define the set $\Omega\left(S_0\right) = \{w_1^t : \exists w_2^t,\ v^t \text{ s.t. } p\left(y_t | w_1^t, y^{t-1}, S_0\right) > 0 \text{ for all } t\}$. Given some initial state S_0, this set defines the realizations, w_1^t, of the random variable W_1^t for which the model assigns positive probability to the data. To deal with an interesting problem we now make the assumption that this set is not empty.

Assumption 5: $\exists$ some initial state S_0 for which $\Omega\left(S_0\right) \neq \{\emptyset\}$. If $\dim\left(W_{1,t}\right) = 0$ the assumption holds if $p\left(y_t | y^{t-1}, S_0; \gamma\right) > 0$.

Therefore, if the five aforementioned assumptions hold and 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 corresponding sequence of densities $\{p\left(W_1^t | y^{t-1}, S_0; \gamma\right)\}_{t=1}^T$, the likelihood function (4) can be approximated by:

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

because of a law of large numbers.

This shows that the problem of evaluating the likelihood of a dynamic equilibrium economy is equivalent to the problem of drawing from $\{p\left(W_1^t | y^{t-1}, S_0; \gamma\right)\}_{t=1}^T$. We now propose a Sequential Monte Carlo algorithm to accomplish this objective.

2.2. A Sequential Monte Carlo Filter

We first fix some further notation. Let $\left\{w_1^{t-1,i}\right\}_{i=1}^N$ be a sequence of N i.i.d. draws from $p\left(W_1^{t-1}|y^{t-1},S_0;\gamma\right)$. Let $\left\{w_1^{t|t-1,i}\right\}_{i=1}^N$ 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 $\left\{w_1^{t-1,i}\right\}_{i=1}^N$ a swarm of particles. Let also $h\left(S_t\right)$ be any measurable functions for which the expectation

$$E_{p\left(W_1^t|y^t,S_0;\gamma\right)}\left(h\left(W_1^t\right)\right)=\int h\left(W_1^t\right)p\left(W_1^t|y^t,S_0;\gamma\right)dW_1^t$$

exists and is finite.

We now present a proposition that is close to previous results in importance sampling.

Proposition 1. *Let* $\left\{w_1^{t|t-1,i}\right\}_{i=1}^N$ *be a draw from* $p\left(W_1^t|y^{t-1},S_0;\gamma\right)$ *and the weights:*

$$q_t^i=\frac{p\left(y_t|w_1^{t|t-1,i},y^{t-1},S_0;\gamma\right)}{\sum_{i=1}^N p\left(y_t|w_1^{t|t-1,i},y^{t-1},S_0;\gamma\right)}.$$

Then:

$$E_{p\left(W_1^t|y^t,S_0;\gamma\right)}\left(h\left(W_1^t\right)\right)\simeq\sum_{i=1}^N q_t^i h\left(w_1^{t|t-1,i}\right).$$

Proof. By Bayes theorem:

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

Therefore if we use $p\left(W_1^t|y^{t-1},S_0;\gamma\right)$ as a important sampling function to draw from the density $p\left(W_1^t|y^t,S_0;\gamma\right)$, the result is a direct consequence of the law of large numbers (e.g. Geweke, 1989, Theorem 1). ■

Proposition 1 shows how we can use $p\left(W_1^t|y^{t-1},S_0;\gamma\right)$ as an important sampling density to draw from $p\left(W_1^t|y^t,S_0;\gamma\right)$ in the following way:

Corollary 2. *Let* $\left\{w_1^{t|t-1,i}\right\}_{i=1}^N$ *be a draw from* $p\left(W_1^t|y^{t-1},S_0;\gamma\right)$. *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 *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* $p\left(W_1^t|y^t,S_0;\gamma\right)$.

Corollary 2 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 $\left\{w_1^{t,i}\right\}_{i=1}^N$ form $p\left(W_1^t|y^t,S_0;\gamma\right)$. This corollary is key in the following Sequential

Monte Carlo algorithm that generates draws $\left\{\left\{w_1^{t|t-1,i}\right\}_{i=1}^N\right\}_{t=1}^T$ from the sequence of densities $\{p(W_1^t|y^{t-1}, S_0;\gamma)\}_{t=1}^T$:

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

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` $\{q_t^i\}_{i=1}^N$. `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 a swarm of particles up to period $t-1$, $\left\{w_1^{t-1,i}\right\}_{i=1}^N$, 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 $p(W_1^t|y^{t-1}, S_0;\gamma)$. Then step 3 takes advantage of corollary 2 and resamples from $\left\{w_1^{t|t-1,i}\right\}_{i=1}^N$ using the weights $\{q_t^i\}_{i=1}^N$ to draw a new swarm of particles up to period t, $\left\{w_1^{t,i}\right\}_{i=1}^N$distributed according to $p(W_1^t|y^t, S_0;\gamma)$. Notice that we use the output of the algorithm $\{s_0^i\}_{i=1}^N$ and $\left\{\left\{w_1^{t|t-1,i}\right\}_{i=1}^N\right\}_{t=2}^T$ to compute the likelihood

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

We emphasize that in the case where the $\dim(W_{1,t}) = 0$, the algorithm collapses to iterating step 2.

This algorithm derives from (but is not equal to) Sequential Monte Carlo algorithms for nonlinear filtering (see Fearhead, 1998 for an elegant review of the relevant literature). We modify existing procedures to deal with more general classes of state space representations than the ones addressed in the literature. In particular we can handle those cases, common in economics, where $\dim(V_t) < \dim(Y_t)$. We consider this more general applicability of our procedure an important advance.

Step 3 is the heart of the algorithm. A naive extension of basic Monte Carlo techniques without this step will diverge as T grows. All the sequences will wander away from the true (unobserved) shocks. To avoid this problem we do not carry over to the next period all the particles generated. We draw with replacement from them, giving a higher probability to those particles that are more likely. The fitting criterion is critical to the convergence of the procedure and very similar to the intuition behind why genetic algorithms work: we allow randomness to generate new simulations but we favor the survival of the particles that are more informative.[7]

Figure 2.1 may help to explain this point. Here we represent six different particles over six periods. We initialize all six particles at different values in period 1 and we move them to period 2. In the second period some of them are sampled (those represented by a green circle) while some are not (those that end with a red circle). As we just explained the sampling favors those draws that are "closer" in probability to the data. In the next period, the particles sampled (like particle 1) give birth to a number of continuations equal to the number of times they have been sampled. We iterate this procedure until the end of our observation sample.

Finally note that the algorithm does not require any assumption on the distribution of the shocks except the ability to evaluate $p\left(W_1^{t-1}|y^{t-1}, S_0; \gamma\right)$, either analytically or by simulation. This opens the door to dealing with models with a rich specification of non-normal innovations.

2.3. Comparison with Alternative Schemes

The algorithm outlined above is not the only procedure to numerically evaluate the likelihood of the data implied by nonlinear models. Our previous discussion highlighted how computing the likelihood amounts to solve a nonlinear filtering problem, i.e., to generate estimates of the values of $W_{1,t}$ so that the integral in (4) can be evaluated. Since this task is of interest in different fields, several alternative schemes have been proposed to handle this problem.

A first line of research has been in deterministic filtering. Historically the first procedure in this line was the Extended Kalman Filter (Jazwinski, 1973) which linearizes the transition and measurement equations and uses the Kalman Filter to estimate for the states and the shocks to the system. This approach suffers from the approximation error incurred by the linearization and by the inaccuracy incurred by the fact that the posterior estimates of the states are not Gaussian. As the sample size grows those problems accumulate and the filter diverges. Even refinements as the Iterated Extended Kalman Filter or the quadratic Kalman Filter cannot solve these problems.

[7] More sophisticated resampling schemes are available for variance reduction. See Doucet, De Freitas and Gordon (2001) and Pitt and Shephard (1999) for reviews of these alternatives.

A second approach in deterministic filtering is the Gaussian Sum approximations (Alspach and Sorenson, 1972) which approximates the different densities required to compute the likelihood with a mixture of normals. Under regularity conditions, as the number of normals increases, we will approximate the densities arbitrarily well. However, the approach suffers from an exponential growth in the number of components in the mixture and from the fact that we still need to use the Extended Kalman Filter to approximate the evolution of those different components.

A third alternative in deterministic filtering is the use of grid-based filters, based on deterministic numerical integration as proposed by Bucy and Senne (1974), to compute the different integrals. Their use is limited as grid-based filters turn out to be very difficult to implement, requiring a constant readjustment to small changes in the model or its parameter values, and they are too computationally expensive to be of any practical use beyond very low dimensions.[8]

Tanizaki (1996) investigates the performance of all those deterministic filters (Extended Kalman Filter, Gaussian Sum approximations and grid-based filters). He uses Monte Carlo evidence to document that all those approximations delivered a very poor performance when applied to real economic applications.

A second strategy is to think of the functions f and g as a change of variables of the innovations to the model and use the Jacobian of the transformation to evaluate the likelihood of the observables (Miranda and Rui, 1997). In general this approach is cumbersome and also difficult to implement since we need to approximate the derivatives in the (unknown) Jacobian. These approximations are costly and not very robust. Furthermore, technical conditions limit its applicability.

A third line of research is the use of Monte Carlo techniques. This approach was inaugurate by Kitagawa (1987). One of the first lessons from this literature was that importing basic simulation techniques is not straightforward because of convergence problems. For instance, a recursive extension of the Importance Sampling scheme is bound to fail as the number of observations grows (Robert and Casella, 1999).

The key innovation was proposed by Gordon, Salmond and Smith (1993). They pointed out that resampling from the simulated data could be performed using properly chosen weights. With these resampling it is feasible to solve efficiently and consistently the filtering problem. Our algorithm is a descendant of the original proposal by Gordon, Salmond and Smith (1993) and the following literature and includes theirs as a particular case when $\{W_{1,t}\} = \{W_t\}$.

[8] Another shortcoming of grid-based filters is that the grid points are fixed ex ante and the results are very dependent on that choice. In comparison we can think about our simulation filter as a grid-based filter where the grid points are chosen endogenously over time based on their ability to account for the data.

Other simulation algorithms include Mariano and Tanizaki (1995) and Geweke and Tanizaki (1999). Mariano and Tanizaki (1995) propose a version of rejection sampling. This method is difficult however because it depends on finding an appropriate density for the rejection test. This search is a time-consuming task that requires substantial work for each particular model. Geweke and Tanizaki (1999) use the whole joint likelihood and draw from the distribution of the whole set of states over the sample using a Metropolis-Hastings algorithm. This approach increases notably the dimensionality of the problem, especially for relatively long samples, and also requires good proposal densities and a good initialization of the chain.

3. An Application: The Stochastic Neoclassical Growth Model

In this section we present an application of our procedure to a dynamic equilibrium economy. We find it natural to use the stochastic neoclassical growth model for that purpose. First, it is a canonical example of a dynamic equilibrium model and it has been used, either directly or with small variations, to address a large number of questions in macroeconomics. Second, it is a relatively simple model, a fact that facilitates the illustration of the different parts of our procedure. In this paper we are more interested in showing the potential of our approach than in the empirical findings per se, and the growth model is the perfect laboratory for that purpose.

We are also aware that, since this a model that is nearly linear for a standard calibration, our procedure may be a bit of overkill. For example, a simpler procedure such as using the Kalman filter after linearizing the equilibrium conditions may deliver estimates that are nearly as good as those obtained while respecting the nonlinearities of the model. We actually see this fact as an advantage since it may help the reader to notice the differences of our algorithm from other alternatives and may allow to comparison of our results to the findings from the Kalman Filter (see Fernández-Villaverde and Rubio-Ramírez, 2003b for more details). Concurrent research applies our algorithm to models more explicitly nonlinear. For example, we investigate, among other examples, models with asset pricing and economies with nominal rigidities. We omit those results to keep this paper focused.

The rest of this section is divided into four parts. First, we present the stochastic neoclassical growth model. Second, we briefly describe how we solve the model numerically. Third, we explain how to evaluate the likelihood function. Finally, we explain how to introduce our Sequential Monte Carlo algorithm in an estimation procedure. We do this from both a Bayesian perspective and a classical one. Later in section 5 we report the results of our estimation for "artificial" and real data.

3.1. The Model

As just mentioned we work with the stochastic neoclassical growth model with leisure. This model is well known (see the textbook exposition of Cooley and Prescott, 1995). Consequently we only go through the minimum exposition required to fix notation.[9]

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

$$U = E_0 \sum_{t=0}^{\infty} \beta^t \frac{\left(c_t^{\theta} (1 - l_t)^{1-\theta}\right)^{1-\tau}}{1 - \tau},$$

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

There is one good in the economy 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 representing random technological progress. The stochastic process z_t follows an AR(1) process $z_t = \rho z_{t-1} + \epsilon_t$ with $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon)$. We restrict ourselves to cases where the process is stationary (i.e. $|\rho| < 1$). Capital's law of motion is $k_{t+1} = i_t + (1 - \delta) k_t$ where i_t is investment and the economy must satisfy the resource constrain $c_t + i_t = e^{z_t} k_t^{\alpha} l_t^{1-\alpha}$.

A competitive equilibrium can be defined in a standard way as a sequence of allocations and prices such that both the representative household and the firm maximize and markets clear. However since both welfare theorems hold in this economy, we can instead solve the equivalent and simpler social planner's problem that maximizes the utility of the representative household subject to the economy resource constraint, the law of motion for capital, the stochastic process and some initial conditions k_0 and z_0.

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

$$\frac{\left(c_t^{\theta} (1 - l_t)^{1-\theta}\right)^{1-\tau}}{c_t} = \beta E_t \left\{ \frac{\left(c_{t+1}^{\theta} (1 - l_{t+1})^{1-\theta}\right)^{1-\tau}}{c_{t+1}} \left(1 + \alpha e^{z_t+1} k_{t+1}^{\alpha-1} l_t^{\alpha} - \delta\right) \right\} \tag{5}$$

and a static optimality condition

$$\frac{1 - \theta}{\theta} \frac{c_t}{1 - l_t} = (1 - \alpha) e^{z_t} k_t^{\alpha} l_t^{-\alpha}, \tag{6}$$

[9] We avoid the case of the model with full depreciation and no leisure choice. Even if in this case the model has a closed-form solution, this form is loglinear and consequently is suited for estimation using the Kalman Filter. We want to deal with an explicitly nonlinear case to illustrate the generality of our procedure.

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 $k'(\cdot, \cdot)$, which deliver the optimal choices as functions of the two state variables, capital and the technology level. In practice, however, the problem is simpler because we only search for the solution $l(\cdot, \cdot)$ and find $c(\cdot, \cdot)$ using the static optimality condition and $k'(\cdot, \cdot)$ using the resource constraint of the economy.

3.2. Solving the Model

The previous system of equations does not have a known analytical solution and we need to use a numerical method to solve it. In a recent paper, Aruoba, Fernández-Villaverde and Rubio-Ramírez (2003) have documented that the Finite Element Method delivers a highly accurate, fast and numerically stable solution for a wide range of parameter values in a model exactly like the one considered here. In addition theoretical results ensure the convergence of the approximation to the true (but unknown) nonlinear solution of the economy. Details of how to implement the Finite Element Method in our application are provided in the appendix.

We emphasize, however, that nothing in the Sequential Monte Carlo filter stops us from using any other nonlinear solution method for the system of equations as perturbations (Guu and Judd, 1997), Chebyshev polynomials (Judd, 1992) or value function iteration. The appropriate choice of solution method should be dictated by the details of the particular model to be estimated.

3.3. The Likelihood Function

We assume that we have observed the following time series $y^T \in \times_{t=1}^{T} R^3$, where, for each t, the first component is output, gdp_t, the second is hours, $hours_t$ and the third is investment, inv_t. We make this assumption out of pure convenience. On the one hand we want to capture some of the main empirical predictions of the model. On the other hand, and again only for illustration purposes, we want to keep the dimensionality of the problem low. However the empirical analysis can be performed with very different combinations of data. Our choice should be understood just as an example of how to evaluate the likelihood function associated with a vector of observations.

Let $\gamma^1 \equiv (\theta, \rho, \tau, \alpha, \delta, \beta, \sigma_\epsilon) \in \Upsilon^1 \subset R^7$ be the structural parameters that describe the preferences and technology of the model. Also, as described in the appendix, our implementation of the Finite Element Method requires shocks bounded between -1 and 1. To achieve that goal we transform the productivity shock in the following way: $\lambda_t = \tanh(z_t)$. 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 in the steady state of the model.

Define $V_t \sim \mathcal{N}(0, \Sigma)$ as a vector of measurement errors for our three observables. To economize on parameters we assume that Σ is diagonal with diagonal elements 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 from the structural parameter values explicit.

The transition equation for this model is:

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

and 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 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 introduce measurement errors as the easiest way to avoid stochastic singularity (remember assumption 1). Nothing in our procedure depends on the presence of measurement errors. We could for example write a version of the model where in addition to shocks to technology we would have shocks both to preferences and to depreciation. This alternative might be more empirically relevant, but it would make the solution of the model much more involved. As we have reiterated several times, since our goal here is merely to illustrate how to use our Sequential Monte Carlo filter to evaluate the likelihood of the model in an example

as simple as possible, we prefer the "trick" of using measurement errors. We feel, however, that in a large number of empirical application, more structured alternatives to measurement errors are required.

Given the fact that we have four sources of uncertainty, and $\dim(V_t) = \dim(Y_t)$, we follow assumption 2 and set $\dim(W_{2,t}) = 0$ and $W_{1,t} = W_t = \epsilon_t$. Let $L\left(y^T;\gamma\right)$ be the likelihood function of the data. Remember that the likelihood was given by:

$$L\left(y^T;\gamma\right) = \int \left(\prod_{t=1}^{T} \int p\left(y_t|W_1^t, y^{t-1}, S_0;\gamma\right) p\left(W_1^t|y^{t-1}, S_0;\gamma\right) dW_1^t\right) p\left(S_0;\gamma\right) dS_0. \tag{7}$$

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 to draw 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 (7) as:

$$L\left(y^T;\gamma\right) = \int \left(\prod_{t=1}^{T} \int p\left(y_t|S_t;\gamma\right) p\left(S_t|y^{t-1}, S_0;\gamma\right) dS_t\right) p\left(S_0;\gamma\right) dS_0. \tag{8}$$

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 (8) as

$$L\left(y^T;\gamma\right) = (2\pi)^{-\frac{3T}{2}} |\Sigma|^{\frac{-T}{2}} \int \left(\prod_{t=1}^{T} \int e^{-\frac{\omega(S_t;\gamma)}{2}} p\left(S_t|y^{t-1}, S_0;\gamma\right) dS_t\right) p\left(S_0;\gamma\right) dS_1. \tag{9}$$

This last expression is simple to handle. Given 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 Sequential Monte Carlo filter, we can build the states $\left\{\{s_t^i\}_{i=1}^{N}\right\}_{t=1}^{T}$ and the prediction error $\left\{\{\omega(s_t^i;\gamma)\}_{i=1}^{N}\right\}_{t=1}^{T}$ implied by them. We set $s_0^i = S_{ss} \ \forall i$. Therefore, the likelihood function is approximated by:

$$L\left(y^T;\gamma\right) \simeq (2\pi)^{-\frac{3T}{2}} |\Sigma|^{\frac{-T}{2}} \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} e^{-\frac{\omega(s_t^i;\gamma)}{2}}. \tag{10}$$

Note that equation (10) is nearly identical to the likelihood function implied by the Kalman Filter (see for example equation 3.4.5 in Harvey, 1989) when applied to a linear model. The difference is that in the Kalman Filter the prediction errors $\omega(s_t^i;\gamma)$ come directly from the output of the Riccati equation, while in our filter those come from the output of the simulation.

4. Estimation Algorithms

We now explain how to use to the approximated likelihood function (10) to perform nonlinear likelihood-based estimation from both a Bayesian perspective and a classical one. First we will describe the Bayesian approach, then the classical.

In a Bayesian approach the main inference tool is the posterior distribution of the parameters given the data $\pi\left(\gamma|y^T\right)$. Once the posterior distribution is obtained, we can define a loss function and obtain a point estimate. The Bayes theorem tells us that the posterior density is proportional to the likelihood times the prior. Therefore, we need both 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 of the model is approximated by (10). The next step in Bayesian inference is to find the parameters' posterior. In general the posterior does not have a closed form. Therefore we use a Metropolis-Hasting algorithm to draw from it.[10] The algorithm to draw a chain $\{\gamma_i\}_{i=1}^M$ from $\pi\left(\gamma|y^T\right)$ is as follows:

Step 0, Initialization: `Set` $i \rightsquigarrow 0$ `and an initial` γ_i. `Solve the model for` γ_i `and compute` $f(\cdot,\cdot;\gamma_i)$ `and` $g(\cdot,\cdot;\gamma_i)$. `Evaluate` $\pi(\gamma_i)$ `and` $L\left(y^T;\gamma_i\right)$ `using (10). Set` $i \rightsquigarrow i+1$.

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

Step 2, Solving the Model: `Solve the model for` γ_i^* `and compute` $f(\cdot,\cdot;\gamma_i^*)$ `and` $g(\cdot,\cdot;\gamma_i^*)$.

Step 3, Evaluating the proposal: `Evaluate` $\pi(\gamma_i^*)$ `and` $L\left(y^T;\gamma_i^*\right)$ `using (10).`

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

Once $\{\gamma_i\}_{i=1}^M$ is obtained through this algorithm, any moments of interest of the posterior

[10] In other examples we could exploit the structure of the problem and use another, more efficient Markov chain Monte Carlo procedure.

can be computed as well as the marginal likelihood of the model.

On the classical side the main inference tool is the likelihood function and its global maximum. Once the likelihood is obtained using (10), we can introduce it inside a maximization loop suitable for the model being studied in the following way:

Step 0, Initialization: `Set` $i \rightsquigarrow 0$ `and an initial` γ_i. `Set` $i \rightsquigarrow i+1$

Step 1, Solving the Model: `Solve the model for` γ_i `and compute` $f(\cdot,\cdot;\gamma_i)$ `and` $g(\cdot,\cdot;\gamma_i)$.

Step 2, Evaluating the Likelihood: `Evaluate` $L\left(y^T;\gamma_i\right)$ `using (10) and get` γ_{i+1} `from a maximization routine.`

Step 3, Stopping Rule: `If` $\left\|\gamma_i - \gamma_{i+1}\right\| > \varepsilon$, `where` $\varepsilon > 0$ `is the accuracy level goal, set` $i \rightsquigarrow i+1$ `and go to step 1. Otherwise stop.`

The output of the algorithm, $\widehat{\gamma}_{MLE} = \gamma_i$, is the maximum likelihood point estimate. We can compute its asymptotic variance-covariance matrix as follows

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

Since in general we cannot evaluate this second derivative directly, we will use a numerical approximation using standard procedures.

Finally, although we do not use it in this paper, the value of the likelihood function at its maximum is also useful to build likelihood ratios for model comparison purposes.

5. Findings

In this section we conduct likelihood-based inference on our model. We undertake two exercises. In the first exercise we simulate "artificial" data from the model for a particular choice of values of γ. Then with these data, we compute the likelihood and estimate the parameters of the model using our Sequential Monte Carlo algorithm. This exercise documents how our filter delivers good estimates of the "true" parameter values. In this way we address two critical questions. First, since our procedure only produces an estimate of the likelihood function, we want to know if the numerical error incurred stops the filter from finding accurate parameter estimates. Working with simulated data avoids the problem of estimates being affected by model misspecification. Second, we can determine how many particles we need to obtain an accurate estimation. The theoretical arguments presented above rely on asymp-

totic arguments and they cast little light on the number of particles required in a particular application.

The second exercise takes the model to real data. We estimate it using real output per capita, average hours worked and real gross fixed investment per capita in the U.S. from 1964:Q1 to 2003:Q1. This exercises proves how the filter can be brought to "real life" applications and how it delivers sensible results.

We perform both exercises from a Bayesian perspective a from a classical one. For the Bayesian approach, we specify prior distributions over the parameters, evaluate the likelihood using our filter and draw from the posterior using a Metropolis-Hastings algorithm. However, since we specify flat priors, the posterior mean can be interpreted as the maximum likelihood estimate. In addition we perform a simulated annealing search to find "pure" maximum likelihood estimates.

We divide our exposition in four parts. First, we specify the priors for the parameters. Second, we present results from the "artificial" data experiment. Third, we present the results of the estimation with real data. Finally, we analyze some convergence issues of the Sequential Monte Carlo algorithm.

5.1. Specifying the Priors

The first step is to specify prior distributions for the different parameters of the model $\gamma \equiv (\theta, \rho, \tau, \alpha, \delta, \beta, \sigma_\epsilon, \sigma_1, \sigma_2, \sigma_3) \in \Upsilon$. We write $\pi(\gamma) : \Upsilon \rightarrow R^+$ when we denote the product of all the different priors.

We adopt flat priors for all ten parameters subject only to some boundary constraints to make the priors proper and to rule out parameter values that are either incompatible with the model (i.e. a negative value for a variance) or extremely implausible (the parameter governing the elasticity of substitution being bigger than 100). The looseness of those last constraints is shown by the fact that the simulations performed below never travel even close to those bounds.

Our choice of flat priors is motivated by two reasons. First, since we are going to undertake estimation on simulated data generated by some known parameter values, we do not want to bias the results in favor of our procedure by a careful choice of priors. Second, with a flat prior the posterior is proportional to the likelihood function.[11] Consequently our Bayesian results can be interpreted as a classical exercise where the mode of the likelihood function is the maximum likelihood estimate. Also, a researcher who prefers to use more informative

[11] The exception is the small issue of the bounded support of the priors. If we think about those bounds as frontiers of admissible parameter values in a classical perspective, the argument equating the posterior and likelihood holds exactly. Otherwise, it holds nearly exactly because the likelihood puts an negligible mass outside the support of the priors.

priors can always reweight the draws from the posterior to accommodate his favorite priors (see Geweke, 1998).[12]

Table 3.1: 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

We now describe the priors in more detail. The parameter governing labor supply, θ, follows a uniform distribution between 0 and 1. That range captures all the possible values for which leisure has positive marginal utility. The persistence of the technology shock, ρ, follows a uniform distribution between 0 and 1. This region implies a stationary distribution of the variables of the model[13] with a lower bound on no persistence. The parameter governing the elasticity of substitution, τ, follows a uniform between 0 (linear preferences) and 100. That choice encompasses all range of empirical estimates of the parameter and 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.[14] The prior for the technology parameter, α, is uniform between 0 and 1, including all values for which the marginal productivity of

[12] Note that we do not argue that our flat priors are uninformative. After a reparametrization of the model, a flat prior may become highly curved.

[13] This prior rules out almost surely the presence of a unit root in the output process. One attractive point of Bayesian inference is that, in contrast with classical methods, it is not necessary to use special tools to deal with unit roots (Sims and Uhlig, 1991). In the same way our filter can deal with these unit roots paying the cost of a somehow lower efficiency. As a consequence our prior choice is not motivated by any technical reason but out of our view of what is a reasonable characteristic of the data. We are using a version of the neoclassical growth model without long-run technological progress. As described below, we filter our data using a H-P filter before feeding them into the likelihood function. Since the H-P filter removes up to two unit roots (King and Rebelo, 1993), we are only ruling out the presence of three unit roots in output, a highly implausible hypothesis.

[14] As Lucas (1987) pointed out, in the steady state the product of τ and the rate of growth of output are equal to a constant plus the interest rate since the deterministic Euler condition states that $(1+g)^\tau = \beta(1+r)$ and then $\tau g = \log\beta + r$.

capital and labor are positive. 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 rates between 0 and 316 per cent. 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 3.1 summarizes the previous discussion.

5.2. Results with "Artificial" Data

As a first step to test our procedure we simulate observations from our model to use them as "artificial" data for the estimation. We will generate data from two different calibrations.

First, we select the benchmark calibration values for the stochastic neoclassical growth model according to the standard practice (Cooley and Prescott , 1995) to make our experiment as relevant as possible. The discount factor $\beta = 0.9896$ matches an annual interest rate of 4.27 per cent (see McGrattan and Prescott, 2000 for a justification of this number based on their measure of the return on capital and on the risk-free rate of inflation-protected U.S. Treasury bonds). The risk aversion $\tau = 2$ is a common choice in the literature. $\theta = 0.357$ matches the microeconomic evidence of labor supply. We set $\alpha = 0.4$ to match labor share of national income. The depreciation rate $\delta = 0.02$ fixes the investment/output ratio and $\rho = 0.95$ and $\sigma = 0.007$ match the stochastic 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 per cent of the steady state value of output, 0.35 per cent of the steady state value hours and 0.2 per cent of the steady state of investment based on our priors regarding the relative importance of measurement errors in National and Income Product Accounts. The chosen values are summarized in table 3.2.

Table 3.2: 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 keeps the same values for all the parameters except for τ and σ_ϵ. We increase τ to a value of 50 (implying a relative risk aversion of 24.5) and σ_ϵ to 0.035. The interaction between high risk aversion and high variance introduces a strong nonlinearity in the model that will help us to assess how the procedure does in a much more challenging environment. Our value for risk aversion is an order of magnitude higher than the usual values used in macroeconomics but not too far away from some numbers implied by practitioners in finance (see Cochrane and Hansen, 1992). However, we do not justify our choice based on

empirical relevance but on our desire to assess the performance of our algorithm under highly nonlinear circumstances.

We solve the model using our finite element method with 140 elements and we draw a sample of size 100 for each of the two calibrations.[15] We use our priors and our likelihood evaluation algorithm with 40,000 particles to get 50,000 draws from the posterior distribution using the Metropolis-Hastings algorithm.

We begin discussing the results for the "standard calibration." First, in figure 3.1, we plot the likelihood function in logs of the model given our simulated data. Since we deal with a high dimensional object, we plot in each panel the shape of the function for an interval of ± 20 per cent of the calibrated value of the structural parameter, keeping the rest of the parameters fixed at their calibrated values. For illustration purposes, the "true" value for the parameter corresponding to the direction being plotted is represented by the vertical red line. We can think of these plots as transversal cuts of the likelihood function. Since for some parameter values the likelihood function takes values less than -2,000 in log terms, roughly zero probability, we do not plot them to enhance the readability of the figure.

We see how the likelihood is very informative for the parameter α, δ, θ and β: the data clearly points out the most likely values for the parameters. Any estimation procedure, either Bayesian or classical, will quickly lead us to the peak of the likelihood. The situation is more complicated for the remaining three structural parameters, ρ, τ and σ_ϵ, which present very flat likelihoods. The finding for ρ is not very surprising. It is difficult to estimate precisely an autoregressive component, especially with only 100 observations. The parameter governing the elasticity of substitution τ is complicated to uncover because even relatively important changes in it will have very small changes in the behavior of agents. In the growth model, τ only enters in the policy function because of the presence of uncertainty (the steady state values of the model variables do not depend on it). Since the variability of the productivity shock in the standard calibration is low (and consequently the uncertainty in the data that will allow us to identify this parameter is also small), it is nearly impossible to get a very accurate estimate inside the region 1.8 and 2.2. Finally, σ_ϵ is confounded in the data with the measurement errors. This may be interpreted as a cautionary lesson for an indiscriminate use of measurement errors in empirical models.

We now present inference results. We graph our empirical posterior distributions in figure 3.2 (where the red line is again the calibrated value) and report the mean and standard deviations of these distributions in table 3.3. Under a quadratic loss function, the mean of the posterior distribution is the optimal point estimate of the parameter. Also, given our flat priors, the modes in figure 3.2 will be our maximum likelihood point estimates.

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

Table 3.3: Posterior Distributions Benchmark Calibration

Parameters	Mean	s.d.
θ	0.357	6.72×10^{-5}
ρ	0.950	3.40×10^{-4}
τ	2.000	6.78×10^{-4}
α	0.400	8.60×10^{-5}
δ	0.020	1.34×10^{-5}
β	0.989	1.54×10^{-5}
σ_ϵ	0.007	9.29×10^{-6}
σ_1	1.58×10^{-4}	5.75×10^{-8}
σ_2	1.12×10^{-2}	6.44×10^{-7}
σ_3	8.64×10^{-4}	6.49×10^{-7}

Table 3.3 reveals that our method does an excellent job of pinning down the values of the parameters, especially considering the low number of iterations. All the structural parameters except the standard deviation of the measurement error on output are estimated in an unbiased and tight way. That should not be a big surprised after inspecting the likelihood function in figure 3.1.

However, we mentioned before that for three parameters, ρ, τ and σ_ϵ, the likelihood was not that informative. How can we get such accurate answers? First, even for ρ and τ the likelihood displays an informative shape (the log scale may be deceptive). The case of σ_ϵ is more complicated because the likelihood is nearly flat. The result may be a consequence of our initial values for the Metropolis-Hastings.

Since we initialize our simulation close to the true parameter values we may be biasing the results in our favor. The problem of how to select the initial values for Markov chain Monte Carlo is well known but in our case it is particularly relevant because, ironically, we know the "true" parameter values and we can get very different answers with a careful choice of starting values and length of the chain. A first (but weak) lesson from our results above would be that our the procedure stays where it needs to stay when we begin at the right point of the parameter space. But of course a second lesson is that we need to assess how robust our findings are.

There are two possibilities to check the robustness of the estimates. One is to initiate the chain at the mean of the priors. Since our priors are flat over a large range, this choice implies initial values very far away from the true parameter values. The second alternative is to begin at a middle distance from the "true" parameter values (for example, 20 per cent off). We investigated both alternatives. We found that the algorithm quickly moves in the right

direction searching for higher likelihoods. The drawback is that we need a long burn-in period until the likelihood stabilizes. This observation is similar the teachings from the literature in applied Markov chain Monte Carlo: a careful exploration of the parameters' space for a good initial value is key to achieve a good performance in reasonable time. To illustrate how the algorithm searches for the right region, figure 3.3 plots the time series of the evolution of values of parameters in the simulation and how they converge from the initial value (signaled by an arrow) to the true value, represented by the red line.

Summarizing, we interpret the results from the different chains as follows. First, if we begin around the true parameter values, we stay in that neighborhood. Second, if we begin far away, after a long burn-in period, we converge to the right region.

An alternative to our Bayesian inference is to perform maximum likelihood. Given our previous exposition, such a task is relatively simple. We only need to plug in our maximization algorithm inside a maximization routine and let the procedure find a maximum of the function. However, our simulation procedure makes difficult to use a simple Newton-Raphson update scheme. Since we cannot compute derivatives analytically, we approximate them numerically. The sampling error associated with the likelihood function evaluation makes these numerical derivatives unstable[16] and the procedure faces difficulties in converging. We find, however, that using a simulated annealing scheme we get successful estimates of the parameter value nearly identical to the ones reported above.[17]

Table 3.4: Maximum Likelihood Estimates Benchmark Calibration

Parameters	MLE	s.d.
θ	0.357	8.19×10^{-6}
ρ	0.950	0.001
τ	2.000	0.020
α	0.400	2.02×10^{-6}
δ	0.002	2.07×10^{-5}
β	0.990	1.00×10^{-6}
σ_ϵ	0.007	0.004
σ_1	1.58×10^{-4}	0.007
σ_2	1.12×10^{-3}	0.007
σ_3	8.63×10^{-4}	0.005

[16] Unless we use a large number of particles. This would make the process too slow. For the evaluation of the standard deviations at the MLE we use 4,000,000 particles to assure convergence by brute force.

[17] That result is not surprising given the similarity in spirit of the Metropolis-Hastings and simulated annealing and the shape of the likelihood function reported in figures 3.1 and 3.4.

Table 3.4 presents the results from Maximum Likelihood. Direct inspection reveals that the point estimates are very similar (not surprising given the shape of the likelihood), but that the standard deviations are higher. Fernández-Villaverde and Rubio-Ramírez (2003a) document similar findings in details and argue that the shape of the likelihood function of dynamic equilibrium models (relatively flat close to the maximum) explains the higher standard deviations.

The results of the second calibration are reported in figure 3.4 (the likelihood transversal cuts), figure 3.5 (the posteriors), table 3.5 (mean and standard deviations of the posterior) and in table 3.6 (maximum likelihood estimates).

Table 3.5: Posterior Distributions Extreme Calibration

Parameters	Mean	s.d.
θ	0.357	7.19×10^{-4}
ρ	0.950	1.88×10^{-4}
τ	50.00	7.12×10^{-3}
α	0.400	4.80×10^{-5}
δ	0.020	3.52×10^{-6}
β	0.989	8.69×10^{-6}
σ_ϵ	0.035	4.47×10^{-6}
σ_1	1.58×10^{-4}	1.87×10^{-8}
σ_2	1.12×10^{-2}	2.14×10^{-7}
σ_3	8.65×10^{-4}	2.33×10^{-7}

Table 3.6: Maximum Likelihood Estimates Extreme Calibration

Parameters	MLE	s.d.
θ	0.357	2.42×10^{-6}
ρ	0.950	6.12×10^{-3}
τ	50.000	0.022
α	0.400	3.62×10^{-7}
δ	0.019	7.43×10^{-6}
β	0.990	1.00×10^{-5}
σ_ϵ	0.035	0.015
σ_1	1.58×10^{-4}	0.017
σ_2	1.12×10^{-3}	0.014
σ_3	8.66×10^{-4}	0.023

Our results are very similar to the previous ones. The likelihood is still very informative for the parameter α, δ, θ and β. Now, however, the data also tell us more about the other three parameters, especially ρ. This better information is reflected in table 3.5, which shows tighter and more consistent estimates in a highly nonlinear model than in table 3.3.

Finally let us mention that model comparison in this framework is straightforward. From the output of the algorithm we can either find the maximum value of the likelihood (to build likelihood ratios) or compute the marginal likelihood. This marginal likelihood determines the probability that the model assigns to the observations and serves to compare models. Geweke (1998) provides details regarding how to compute this marginal likelihood. Fernández-Villaverde and Rubio-Ramírez (2003b) use the marginal likelihood ratio to suggest that the data strongly favor a nonlinear version of the stochastic growth model over linearized approximations of the same economy.

5.3. Results with U.S. 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 a H-P filter. In this way we do not need to model explicitly the presence of a trend and possible changes to it.

Again we perform Bayesian and classical inferences. Table 3.7 presents the results from the posterior distribution from 50,000 draws and figure 3.6 displays the posteriors. In this case, to initialize the chain, we used the mean of the posterior computed from a linearized version of the model and the Kalman filter after 400 million iterations.[18]

We briefly discuss some of our results. The discount factor, β, goes very close to 1, a common finding when dynamic equilibrium economies are estimated. The parameter controlling the elasticity of substitution, τ, has a value of 1.825 and θ of 0.323. These two parameters imply an elasticity of substitution of 1.27. The estimated depreciation factor is very low, 0.006 since the estimation tries to compensate for the high desire of accumulation of capital implied by the very high discount factor. The parameter α is close to the canonical value of 0.4. Finally the autoregressive component, ρ, is estimated to be 0.969.

These numbers are close to the ones coming from a standard calibration exercise and to those generally accepted as reasonable after the accumulation of empirical evidence over the last two decades by very different empirical methods. Nearly as important, the standard deviations of the posterior are very low, indicating tight estimates. We interpret this finding

[18] Such a large of draws allows for an "overkill" in terms of convergence of the Metropolis-Hastings and thus for the elimination of the influence of initial guesses.

as another strong endorsement of the ability of the procedure to uncover sensible values for the structural parameters of dynamic equilibrium economies.

Table 3.7: Posterior Distributions Real Data

Parameters	Mean	s.d.
θ	0.323	7.976×10^{-4}
ρ	0.969	0.008
τ	1.825	0.011
α	0.388	0.001
δ	0.006	3.557×10^{-5}
β	0.997	9.221×10^{-5}
σ_ϵ	0.023	2.702×10^{-4}
σ_1	0.039	5.346×10^{-4}
σ_2	0.018	4.723×10^{-4}
σ_3	0.034	6.300×10^{-4}

The estimation delivers numbers a bit more problematic regarding the standard deviation of the productivity shock. In particular this shock seems to be much more variable than the number estimated directly from the Solow residual. At the same time the values for the standard deviations of the measurement errors seem high. The combination of these two results may be an indication of the lack of identification of the stochastic growth model along the dimension of the different shocks.

We conclude this section with table 3.8, which presents results using maximum likelihood. As in the previous cases we observe that results are nearly identical.

Table 3.8: Maximum Likelihood Estimates Real Data

Parameters	MLE	s.d.
θ	0.390	0.044
ρ	0.987	0.708
τ	1.781	1.398
α	0.324	0.019
δ	0.006	0.160
β	0.997	8.67×10^{-3}
σ_ϵ	0.023	0.224
σ_1	0.038	0.060
σ_2	0.016	0.061
σ_3	0.035	0.076

5.4. Convergence of the Sequential Monte Carlo

An important question to answer in practical applications is how many particles to use to evaluate the likelihood function. The theory only provides us with a convergence result as the number of particles goes to infinity but little guidance regarding finite behavior.

To explore this issue we compute 50 different times the likelihood of the model for different numbers of particles (i.e., we compute 50 estimations of the likelihood with 10,000 particles, 50 with 20,000 and so on).

Tables 3.6 to 3.8 report the mean and the standard deviation of the estimated loglikelihood at a particular parameter choice for the three different calibrations. For the benchmark case we estimate the loglikelihood at the mean of the posterior, while we choose different parameters values for the other two cases. We make this choice because of the different behavior of the estimated loglikelihood at those points. While the standard deviation for parameter choice that differs from the mean of the posterior is very low for any number of particles, the standard deviation increases when the loglikelihood is estimated around it. The reason for this increase in the variance is that at the mean of the posterior, the slope of the loglikelihood goes from positive to negative very quickly, making this estimation more difficult.

The robustness of the values of the loglikelihood in simulations with different parameter values and different sets of data justifies our choice of $N = 40,000$. Even in the worse case the standard deviation is less than 0.2 per cent of the value of the loglikelihood. Sensitivity analysis in our estimations also revealed that, after 20,000 particles, our posteriors and point estimates are nearly identical. As mentioned above, efficiency could be improved if we deal properly with the tails of the distribution but in the interest of simplicity we leave a careful evaluation of these refinements for future research.

Table 3.6: Convergence Benchmark Calibration

N	Mean	s.d.
10000	1459.163	6.4107
20000	1461.928	2.8298
30000	1462.078	1.5415
40000	1462.031	0.9900
50000	1462.636	0.7168
60000	1462.696	0.6353

Table 3.7: Convergence Extreme Calibration

N	Mean	s.d.
10000	831.493	0.1954
20000	831.471	0.1347
30000	831.489	0.0971
40000	831.508	0.0836
50000	831.509	0.0882
60000	831.532	0.0607

Table 3.8: Convergence Real Data

N	Mean	s.d.
10000	1014.558	0.3296
20000	1014.600	0.2595
30000	1014.653	0.1829
40000	1014.666	0.1604
50000	1014.688	0.1465
60000	1014.664	0.1347

We can also explore the response of the simulation to changes in the number of particles with figures 3.7 to 3.9. These figures represent the C.D.F. for the weights q_t^i as defined in proposition 1 for a particular t and the three models. Figure 3.7 draws the C.D.F. for the benchmark case, figure 3.8 for the extreme Calibration and figure 3.9 for the real data case. The optimal behavior in terms of informational content of the different paths will be $q_t^i = q_t^j$ for t, i and j. This case will imply a straight C.D.F. with slope $\frac{1}{N}$ and equal weight for all particles. The further away from this straight line the higher the weight on a small set of particles (i.e. most particles carry very little information) and the higher the standard deviation of the estimated loglikelihood. As we could have guessed from the previous tables, the actual C.D.F. almost matches the straight line both for the extreme calibration and the real data. In the benchmark calibration case, the actual C.D.F. is someway farther away from the straight line.

6. Conclusions

We have presented a simple, general and efficient algorithm to perform likelihood-based inference in nonlinear and/or non-normal dynamic equilibrium economies. We have shown how parameter estimation and model comparison can be undertaken, either from a classical perspective or from a Bayesian one when we work with a nonlinear solution of the model. Also we can perform this inference regardless of whether or not we have normal innovations to the

model. The key result has been the development of an algorithm to evaluate the likelihood function of the dynamic model. To achieve this goal we have worked on the tradition of nonlinear filtering theory to develop a Sequential Monte Carlo algorithm that can be applied to a large class of economies. The intuition of the procedure is to simulate different paths for the states of the model but to ensure convergence resampling from them using some appropriately built weights.

Our simulation results and the application to real data of the stochastic neoclassical growth model suggest that the procedure works superbly in delivering accurate and consistent estimates. In a companion paper (Fernández-Villaverde and Rubio-Ramírez, 2003b) we compare point estimates using linear and nonlinear approximations to the solution of the model. We also use the marginal likelihood ratio to suggest that the data favor the nonlinear version of the stochastic growth model over linearized approximations of the same economy. Our current research applies the algorithm to fully nonlinear models of asset pricing, to the study nominal rigidities, to the evaluation of the importance of non-normal innovations to dynamic models (see Geweke 1994 for some suggestive evidence), to regime-switching models and to economies with multiplicity of equilibria.

7. Appendix

This appendix presents further details about this paper's computations. First, it explains our implementation of the Finite Element method. Second, it offers further details of the Metropolis-Hastings algorithm used and its convergence. Finally, it discusses the computing language and the software used.

7.1. The Finite Element Method

We provide a brief exposition of the finite elements method as applied in the paper. For 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 Elements method is to note that we can rewrite the Euler equation for consumption as

$$U_c(k_t, z_t) = \frac{\beta}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} \left[U_c(k_{t+1}, z_{t+1})(1 + \alpha e^{z_{t+1}} k_{t+1}^{\alpha-1} l(k_{t+1}, z_{t+1})^{1-\alpha} - \delta)\right] \exp(-\frac{\epsilon_{t+1}^2}{2\sigma^2}) d\epsilon_{t+1} \tag{11}$$

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 (11) as

$$U_c(t) = \frac{\beta}{\sqrt{\pi}} \int_{-1}^{1} \left[U_c(k_{t+1}, z_{t+1}) \left(1 + \alpha \widehat{\lambda}_{t+1} k_{t+1}^{\alpha-1} l(k_{t+1}, z_{t+1})^{1-\alpha} + \delta\right)\right] \exp(-v_{t+1}^2) dv_{t+1} \tag{12}$$

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 use the same notation for $l\left(\cdot\right)$ in both (11) and (12) 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 only bind 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 fourteen unequal elements in the capital dimension and ten 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 with 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

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

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:

$$R(k_t, z_t; \theta) = \frac{\beta}{\sqrt{\pi}} \int_{-1}^{1} \left[\frac{U_c(k_{t+1}, z_{t+1})_{fe}}{U_c(k_{t+1}, z_{t+1})_{fe}} \left(1 + \alpha \widehat{\lambda}_{t+1} k_{t+1}^{\alpha-1} l_{fe}^{1-\alpha} - \delta\right) \right] \exp(-v_{t+1}^2) dv_{t+1} - 1 \quad (13)$$

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

$$\int_{\left[0,\overline{k}\right] \times [-1,1]} \Psi_{i,j}(k, z) R(k, z; \overline{\theta}) dz dk = 0 \quad \forall i, j \quad (14)$$

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 (14) as

$$\int_{[k_{i-1}, k_i] \times [z_{j-1}, z_j] \cup [k_i, k_{i+1}] \times [z_j, z_{j+1}]} \Psi_{i,j}(k, z) R(k, z; \overline{\theta}) dz dk = 0 \quad \forall i, j. \quad (15)$$

We evaluate the integral in the residual equation with a Gauss-Hermite method and the integrals in (15) 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. The Metropolis-Hastings Algorithm

Given the prior and the likelihood function we need to find the posterior distribution, $p(\gamma|y^T)$. As described in the main body of the paper, since we do not have a closed-form solution for the posterior, we draw a sample of size of M, $\{\gamma_i\}_{i=1}^M$, using a Random Walk Metropolis-Hastings algorithm.

The success 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.[19] 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 per cent acceptance rate tends to give the best performance. We found that rates between 25 and 30 per cent outperformed different alternatives.

We monitored convergence using standard techniques. A complete guide to convergence can be found in Mengersen, Robert and Guihenneuc-Jouyaux (1999).

7.3. 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 with 40,000 particles takes around 6.1 seconds. That implies a total of around 88 hours for each simulation of 50,000 draws. To put this number in perspective note that the linearized version of the model runs 400 million draws in 12 hours. Versions paralellized with MPI directives to be run in the IBM-SP facilities at the Minnesota Supercomputer Institute were also prepared but they were not used in the final computations of the paper. All the code, both in serial and parallel versions, is available upon request from the corresponding author.

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

References

[1] Alspach, D.L. and H.W. Sorenson (1972). "Non-linear Bayesian Estimation using Gaussian Sum Approximations". *IEEE Transaction on Automatic Control* 17, 439-447.

[2] Altuğ, S. (1989). "Time-to-Build and Aggregate Fluctuations: Some New Evidence". *International Economic Review* 30, 889-920.

[3] 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.*

[4] Berger, J.O. and R.L. Wolpert (1988). *The Likelihood Principle.* Institute of Mathematical Statistics, Lecture Notes volume 6.

[5] Bucy, R.S. and H. Youssef (1974). "Nonlinear Filter Representatio via Spline Functions". *5th Symposium on Nonlinear Estimation,* 51-60.

[6] Chari, V.V., L.J. Christiano and P.J. Kehoe (1994). "Optimal Fiscal Policy in a Business Cycle Model." *Journal of Political Economy* 102, 617-652.

[7] Chari, V.V., P. Kehoe and E.R. McGrattan (2000). "Sticky Price Models of the Business Cycle: the Persistence Problem". *Econometrica* 68, 1151-1179.

[8] Christiano, L.J., M. Eichenbaum and C.L. Evans (2001). "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy". *Mimeo*, Northwestern University.

[9] Cochrane, J.H. and L.P. Hansen (1992). "Asset Pricing Explorations for Macroeconomics". In Olivier Blanchard and S. Fisher (eds), *NBER Macroeconomics Annual 7*, 115-165.

[10] 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.

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

[12] Diebold, F.X., L.E. Ohanian and J. Berkowitz (1998). "Dynamic Equilibrium Economies: A Framework for Comparing Model and Data", *Review of Economic Studies* 65, 433-51.

[13] Doucet. A., N. de Freitas and N. Gordon (2001). *Sequential Monte Carlo Methods in Practice.* Springer Verlag.

[14] Fearnhead, P. (1998). *Sequential Monte Carlo Methods in Filter Theory.* DPhil Thesis, Oxford University.

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

[16] Fernández-Villaverde, J. and J. Rubio-Ramírez (2003b). "Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood". *Mimeo*, University of Pennsylvania.

[17] Gallant, A.R. and G. Tauchen (1996). "Which Moments to Match?". *Econometric Theory* 12, 657-681.

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

[19] 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.

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

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

[22] Geweke, J. and H. Tanizaki (1999). "On Markov Chain Monte Carlo Methods for Nonlinear and Non-Gaussian State-Space Models". *Communications in Statistics, Simulation and Computation* 28, 867-894.

[23] Ghysels, E. and A.R. Hall, (2002). "Editor's Introduction to JBES Twentieht Anniversary Issue on the Generalized Method of Moments". *Journal of Business and Economic Statistics* 20, 441.

[24] Gordon, N.J., D.J. Salmond and A.F.M. Smith (1993). "Novel Approaches to Nonlinear/Non-Gaussian Bayesian State Estimation". *IEE Proceedings-F* 140, 107-113.

[25] Gourieroux, C., A. Monfort and E. Renault (1993). "Indirect Inference". *Journal of Applied Econometrics* 8, S85-S118.

[26] Guu S.M. and K.L. Judd (1997). "Asymptotic Methods for Aggregate Growth Models". *Journal of Economic Dynamics and Control* 21, 1025-1042.

[27] Hansen, L.P. (1982). "Large Sample Properties of Generalized Method of Moments Estimation". *Econometrica* 50, 1029-1054.

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

[29] Imbens, G., R. Spady and P. Johnson (1998). "Information Theoretic Approaches in Moment Conditions Models". *Econometrica* 66, 333-357.

[30] Jazwinski, A.H. (1973). *Stochastic Processes and Filtering Theory.* Academic Press.

[31] Jermann, U. and V. Quadrini (2003). "Stock Market Boom and the Productivity Gains of the 1990's". *Mimeo*, University of Pennsylvania.

[32] Judd, K.L. (1992). "Projection Methods for Solving Aggregate Growth Models". *Journal of Economic Theory* 58, 410-452.

[33] King, R.G. and S.T. Rebelo (1993). "Low Frequency Filtering and Real Business Cycles". Journal of Economic Dynamics and Control 77, 207-231.

[34] Kitagawa, G. (1987). "Non-Gaussian State-Space Modeling of Nonstationary Time Series". *Journal of the American Statistical Association* 82, 1032-1063.

[35] Kitamura, Y. and M. Stutzer (1997). "An Information-theoretic Alternative to Generalized Method of Moment Estimation". *Econometrica* 65, 861-874.

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

[37] Lee, B. and B.F. Ingram (1991). "Simulated Estimation of Time-Series Models". *Journal of Econometrics* 47, 197-205.

[38] Lubick, T.A. and F. Schorfheide (2003). "Testing for Indeterminacy: an Application to U.S. Monetary Policy". *American Economic Review*, forthcoming.

[39] Lucas, R.E. (1987). *Models of Business Cycles*. Basil Blackwell.

[40] Mariano, R.S. and H. Tanizaki (1995). "Prediction, Filtering and Smoothing Techniques in Nonlinear and Nonnormal Cases using Monte-Carlo Integration". In H.K. VanDijk, A. Monfort and B.W. Brown (eds), *Econometric Inference Using Simulation Techniques*. John Wiley & Sons.

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

[42] Mehra, R. and E.C. Prescott (1985). "The Equity Premium: a Puzzle". *Journal of Monetary Economics* 15, 145-161.

[43] 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.

[44] Miranda, M.J. and X. Rui (1997). "Maximum Likelihood Estimation of the Nonlinear Rational Expectations Asset Pricing Model". *Journal of Economic Dynamics and Control* 21, 1493-1510.

[45] Monfort, A. (1996). "A Reappraisal of Missespecified Econometric Models". *Econometric Theory* 12, 597-619.

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

[47] Pitt, M.K. and N. Shephard (1999). "Filtering via Simulation: Auxiliary Particle Filters". *Journal of the American Statistical Association* 94, 590-599.

[48] Robert, C.P. and G. Casella (1999). *Monte Carlo Statistical Methods*. Springer-Verlag.

[49] Rotemberg, J.L. and M. Woodford (1997). "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy". In Ben Bernanke and J.L. Rotemberg (eds) *NBER Macroeconomic Annual 1997*. MIT Press.

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

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

[52] Smith, A.A. (1993). "Estimating Nonlinear Time-series Models Using Simulated Vector Autoregressions". *Journal of Applied Econometrics* 8, S63-84.

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

[54] Sims, C.A. and H. Uhlig (1991). "Understanding Unit Rooters: A Helicopter Tour". *Econometrica* 59, 1591-1599.

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

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

[57] Tanizaki, H. (1996). *Nonlinear Filters: Estimation and Application*. Second Edition. Springer Verlag.

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

[59] Watson, M.W. (1993). "Measures of Fit for Calibrated Models". *Journal of Political Economy* 101, 1011-1041.

[60] White, H. (1994), *Estimation, Inference and Specification Analysis*. Cambridge University Press.

[61] Woodford, M. (2003). *Interest and Prices*. Princeton University Press.

Figure 2.1: Particles Evolving over Time

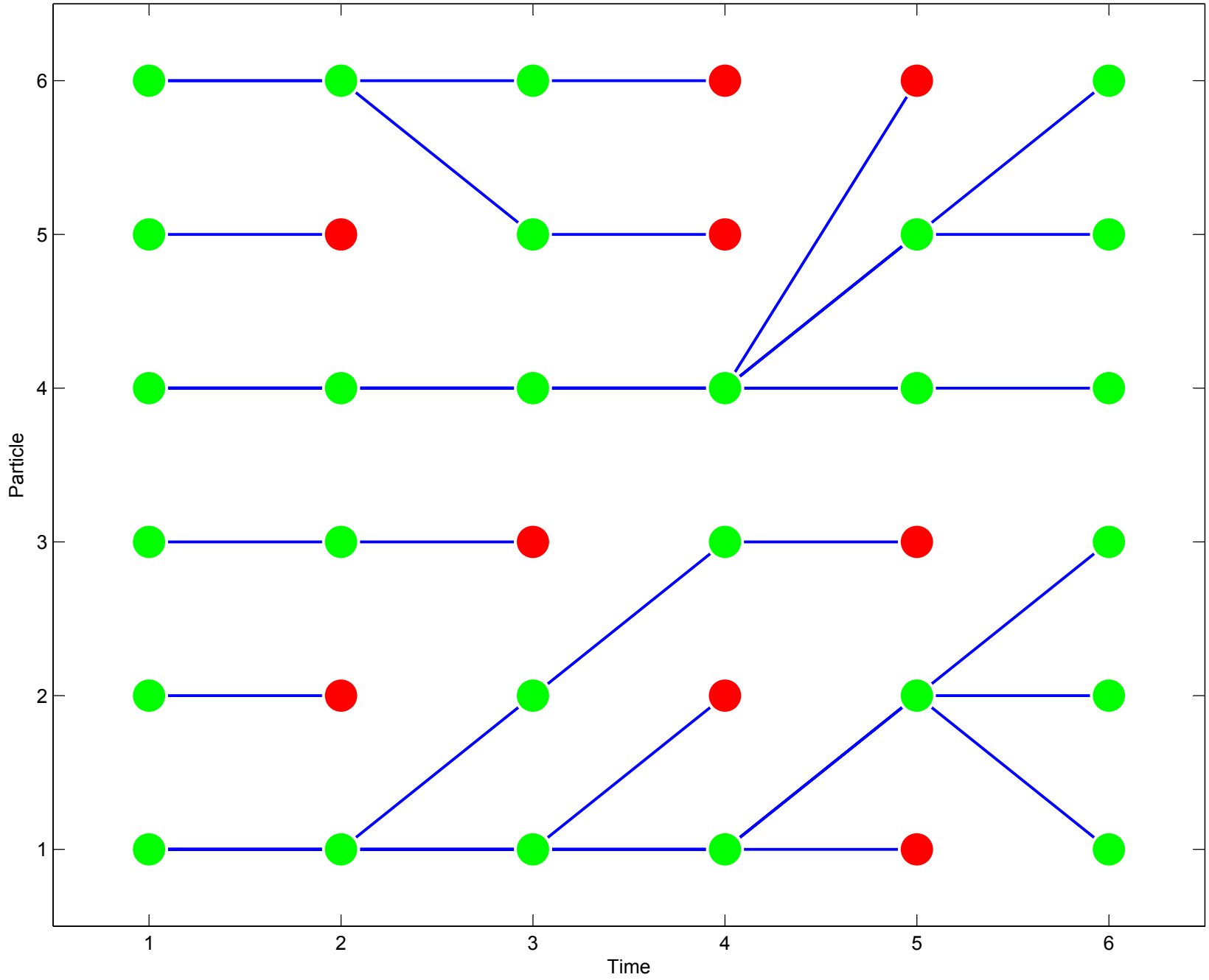
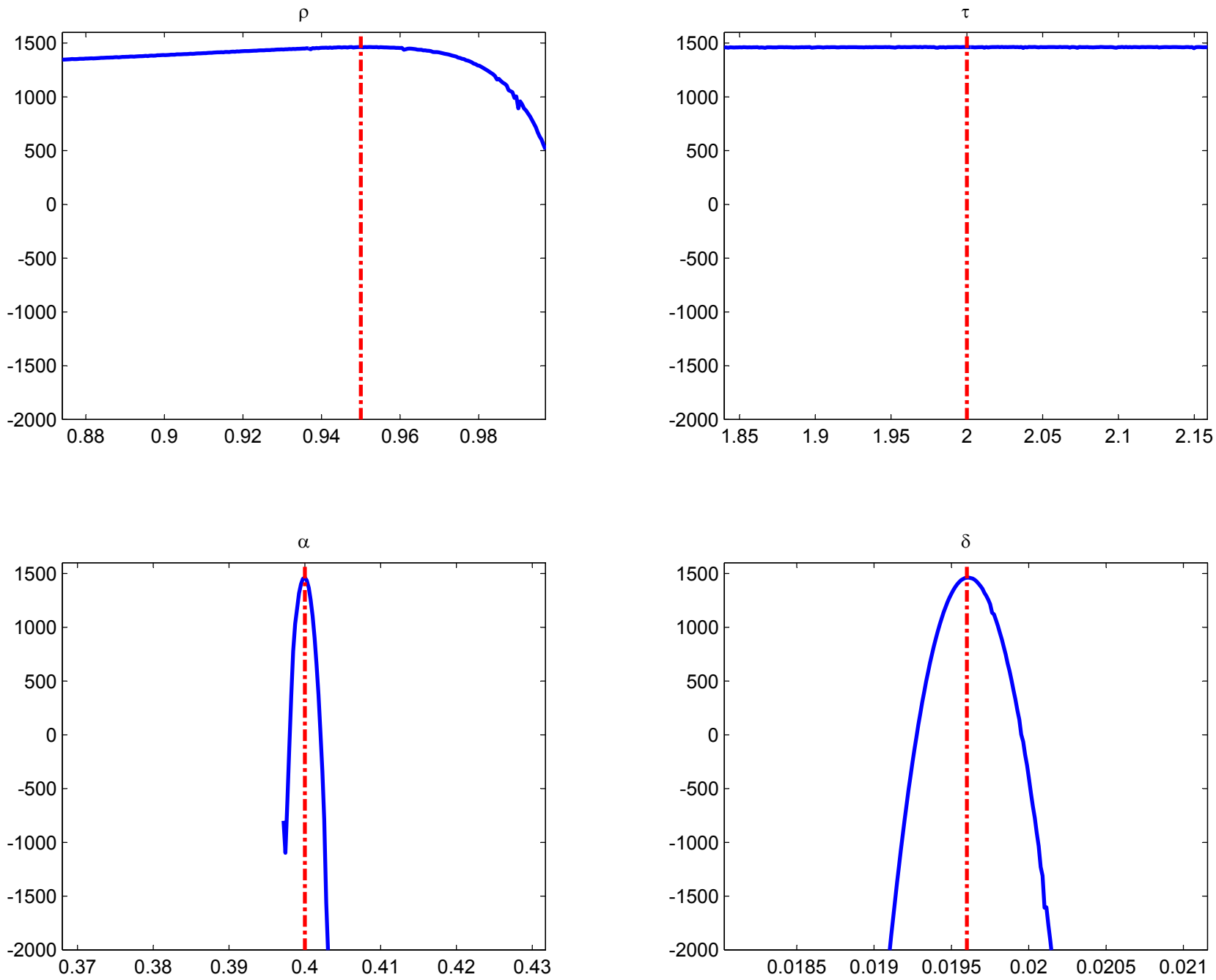
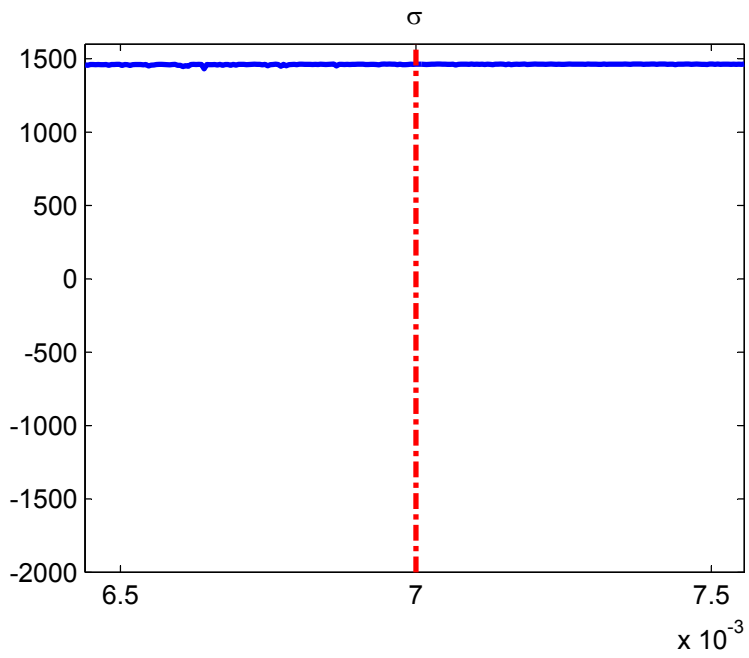
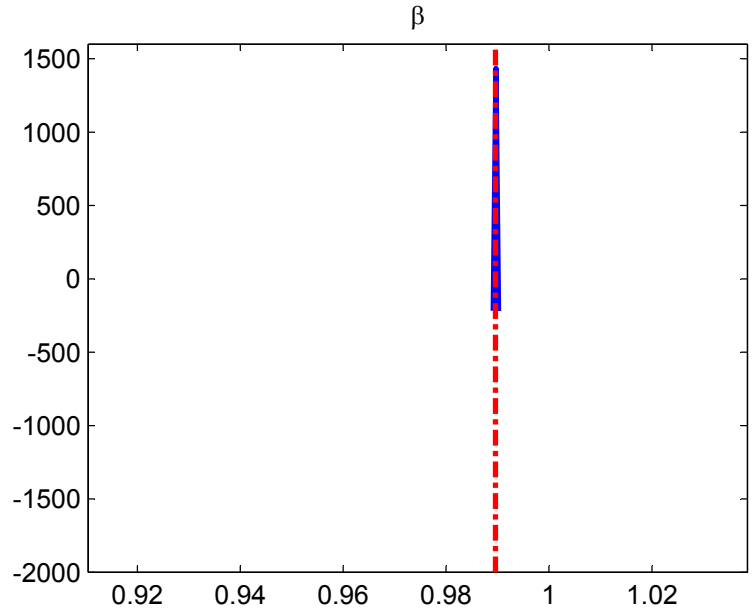


Figure 3.1: Likelihood Benchmark Calibration




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


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

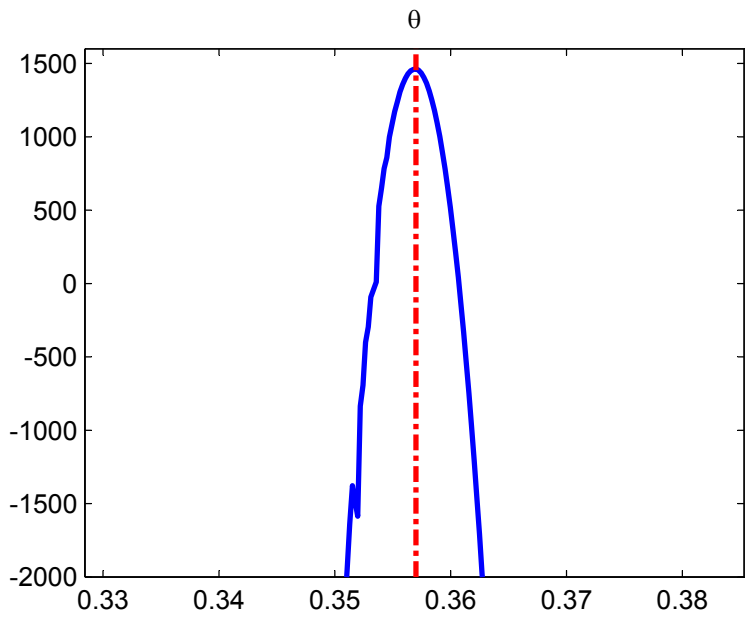
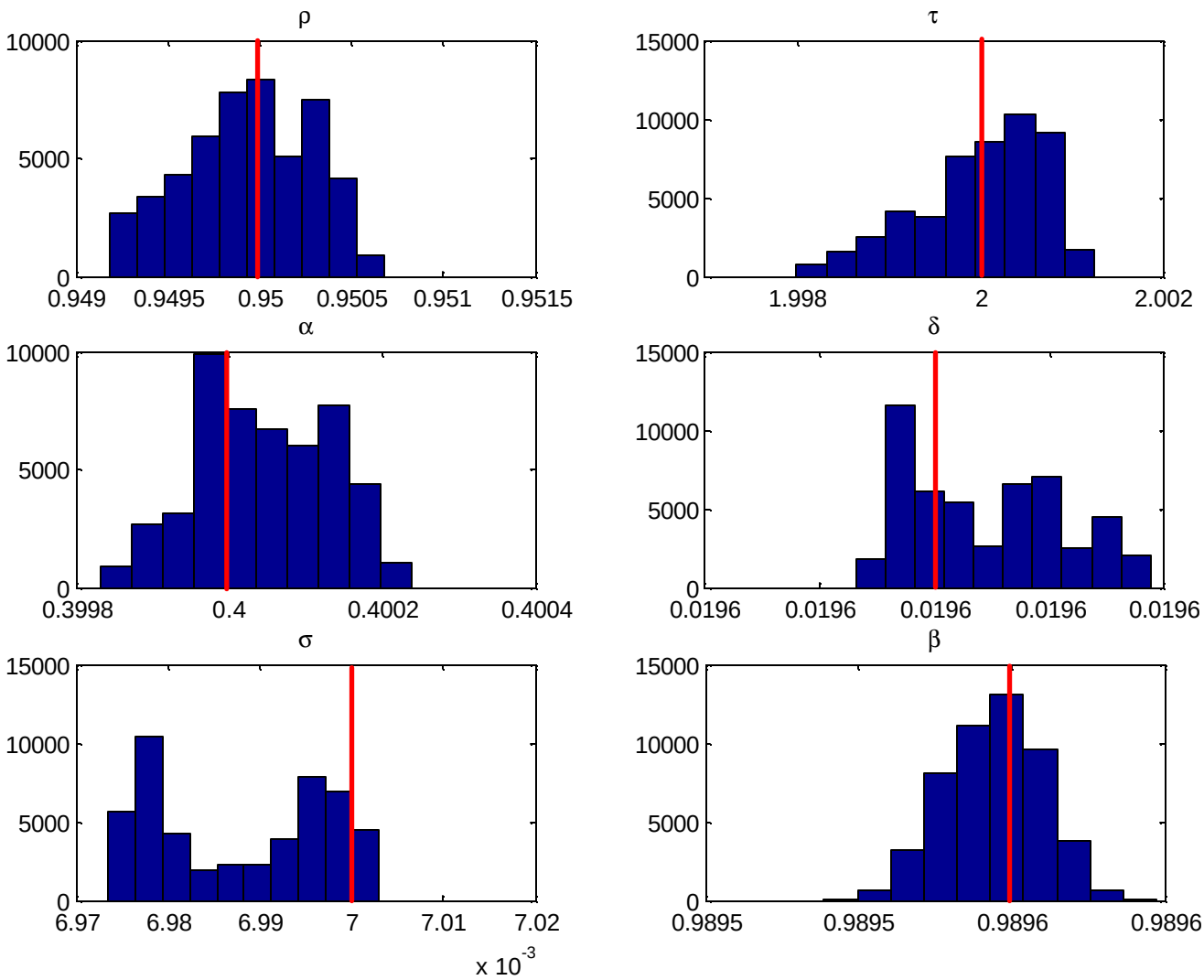

θ
1500
1000
500
0
-500
-1000
-1500
-2000
0.33
0.34
0.35
0.36
0.37
0.38

Figure 3.2: Posteriors Benchmark Calibration



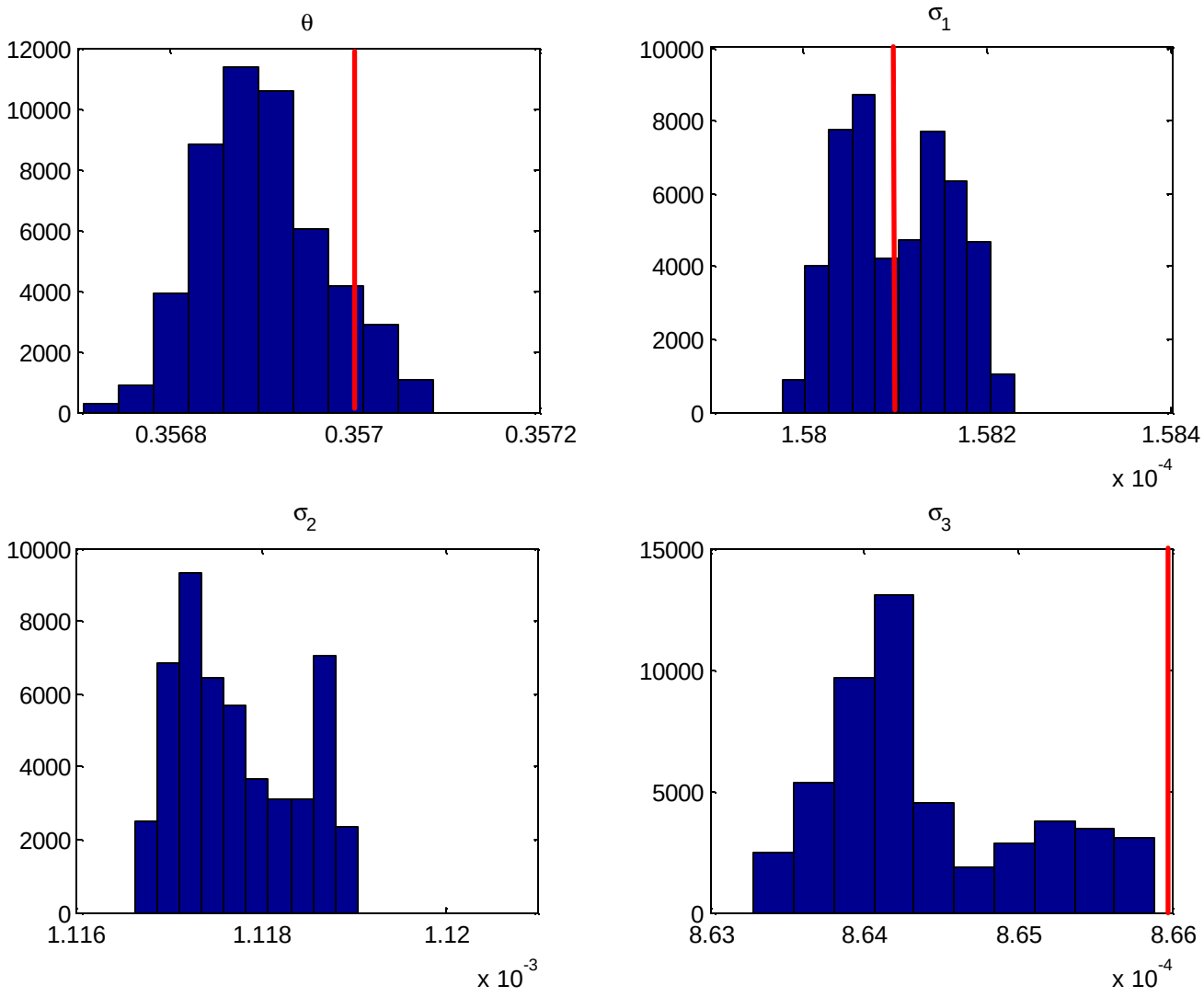
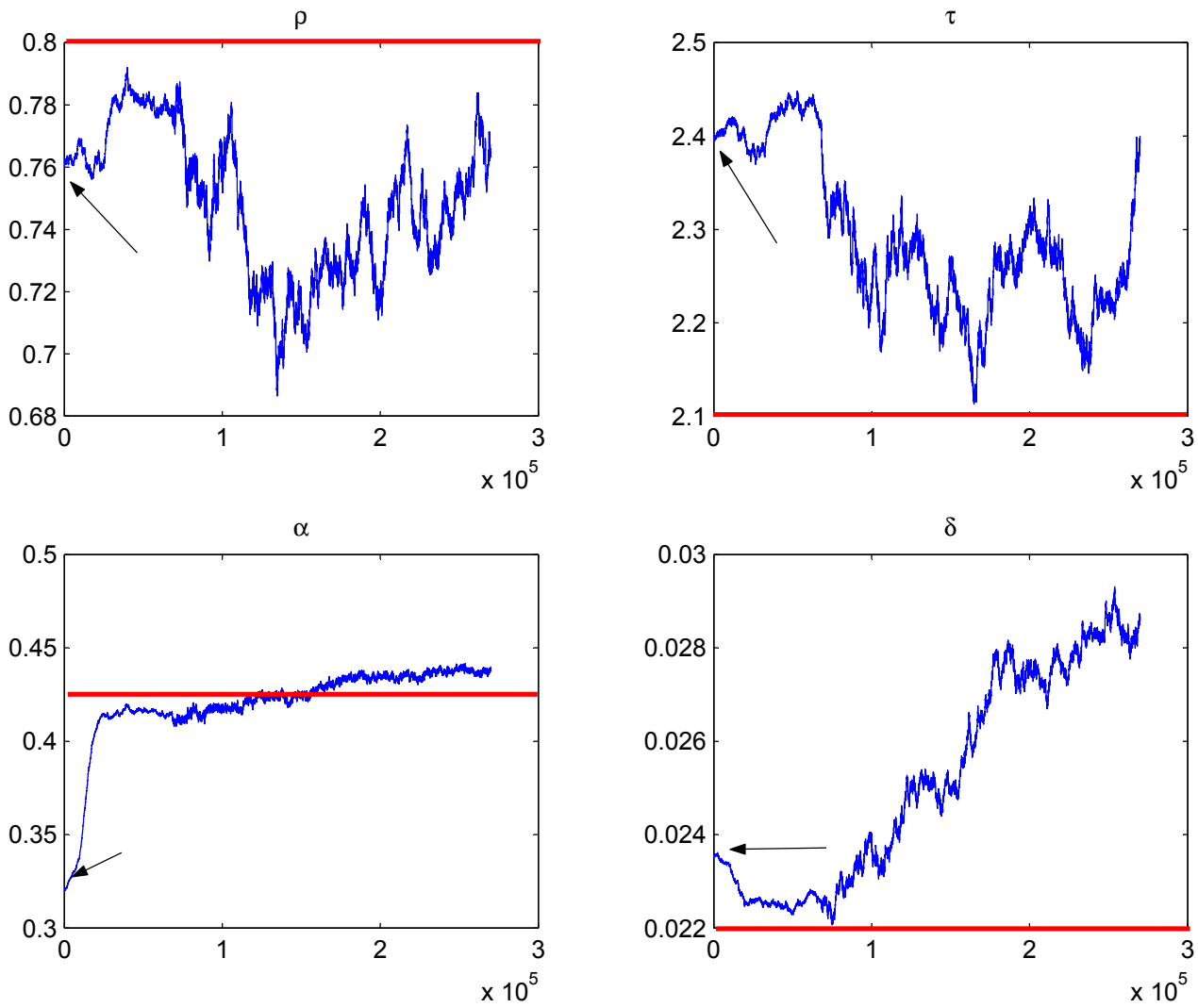

θ
12000
10000
8000
6000
4000
2000
0
0.3568
0.357
0.3572
σ1
10000
8000
6000
4000
2000
0
1.58
1.582
1.584
x 10-4
σ2
10000
8000
6000
4000
2000
0
1.116
1.118
1.12
x 10-3
σ3
15000
10000
5000
0
8.63
8.64
8.65
8.66
x 10-4

Figure 3.3: Convergence of Estimates Benchmark Calibration



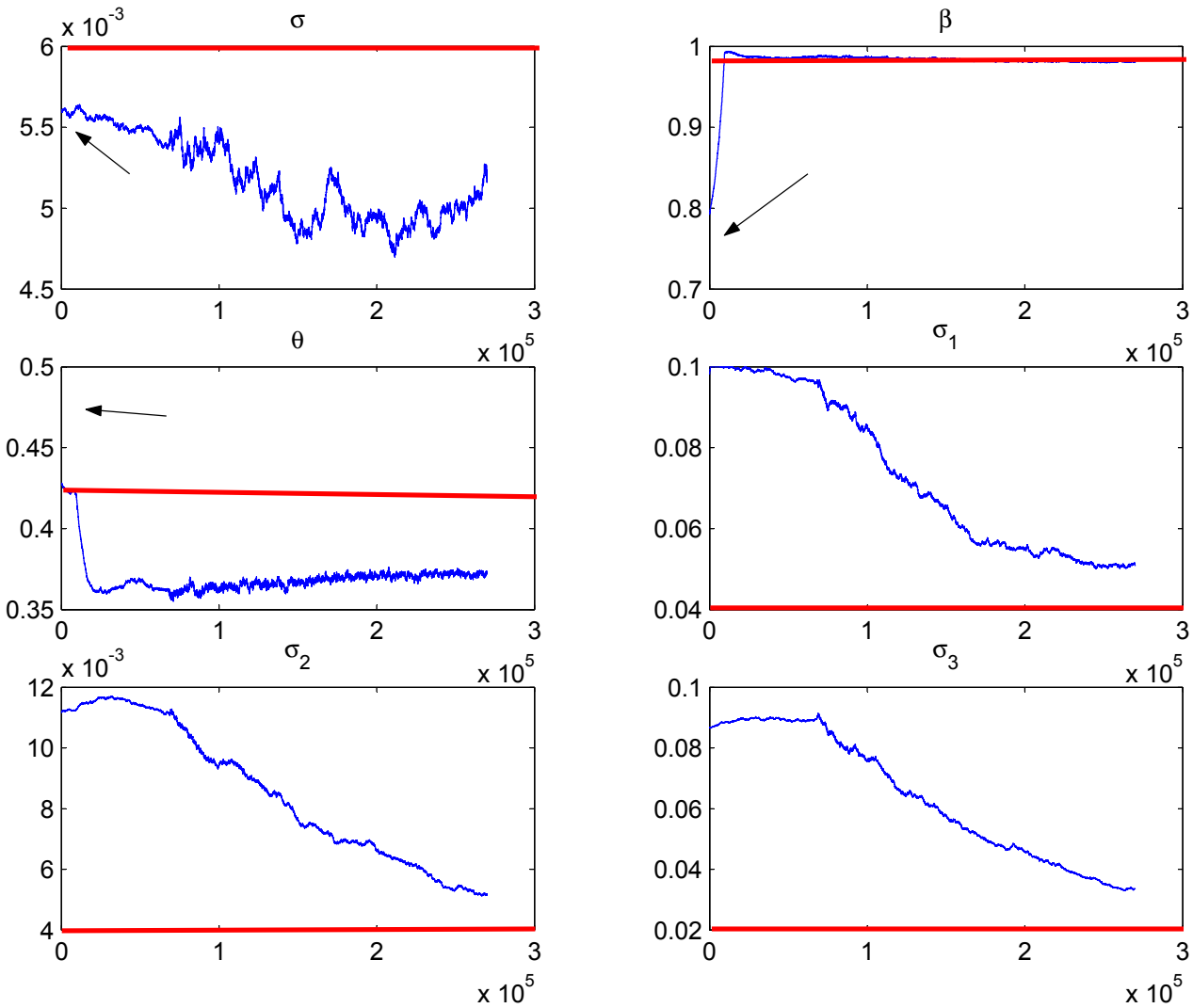
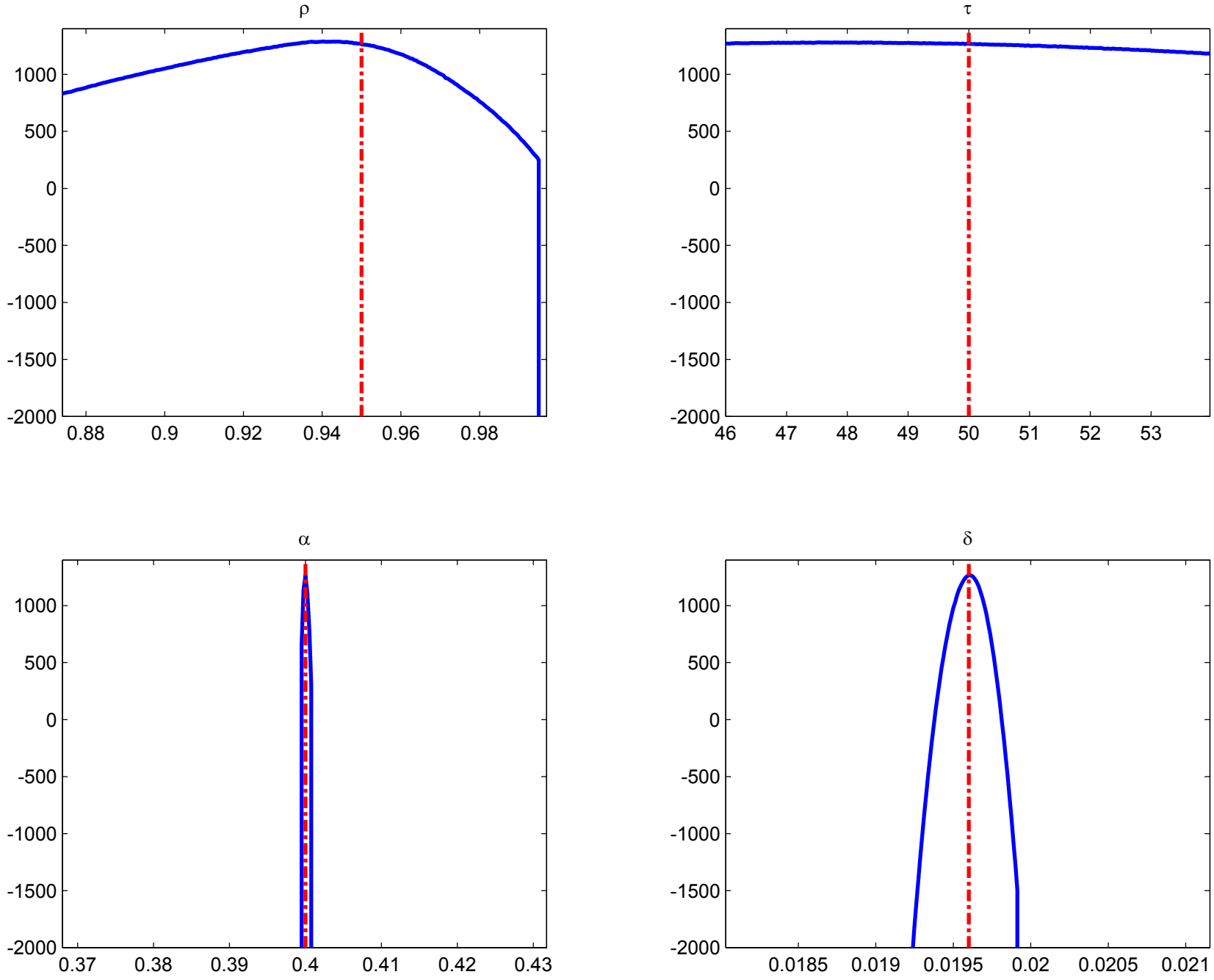
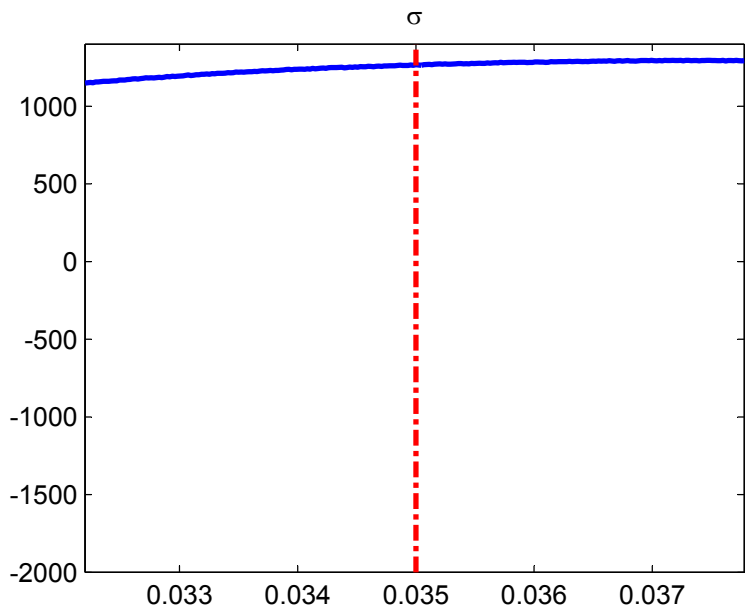
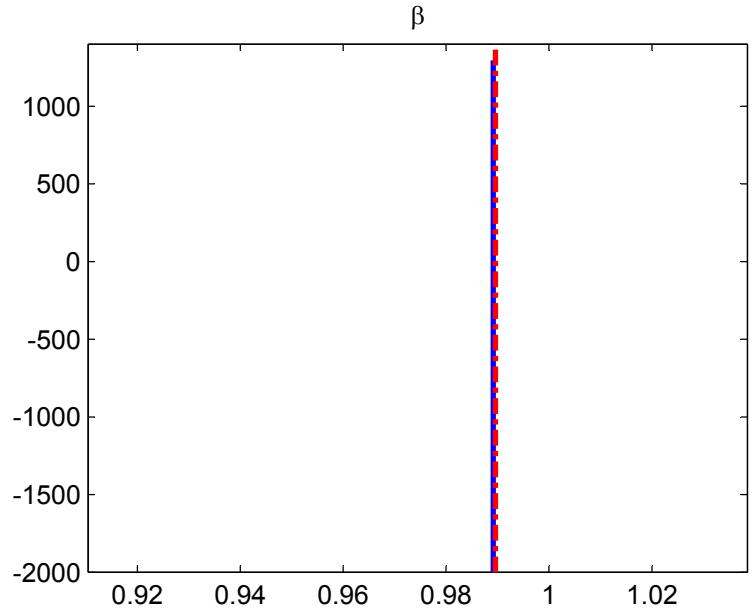

σ
x 10^-3
6
5.5
5
4.5
0
1
2
3
β
1
0.9
0.8
0.7
θ
x 10^5
0.5
0.45
0.4
0.35
σ1
0.1
0.08
0.06
0.04
σ2
12
10
8
6
4
σ3
0.02

Figure 3.4: Likelihood Extreme Calibration




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


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

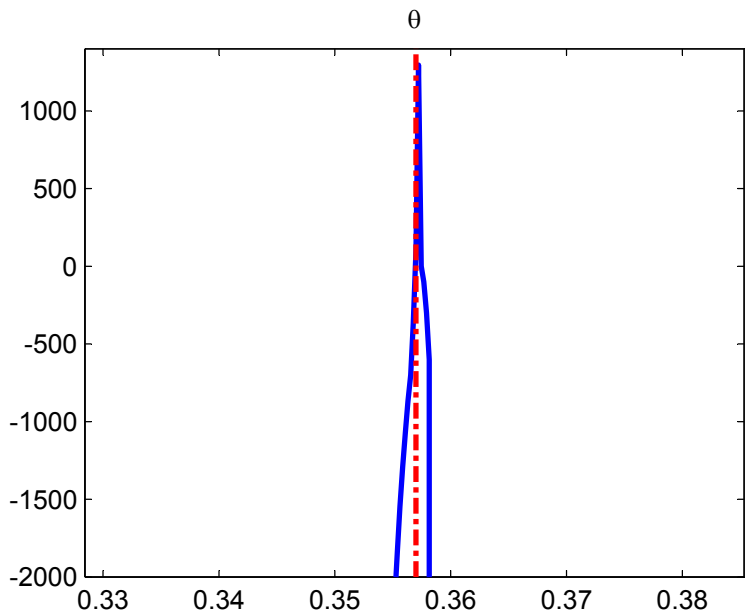
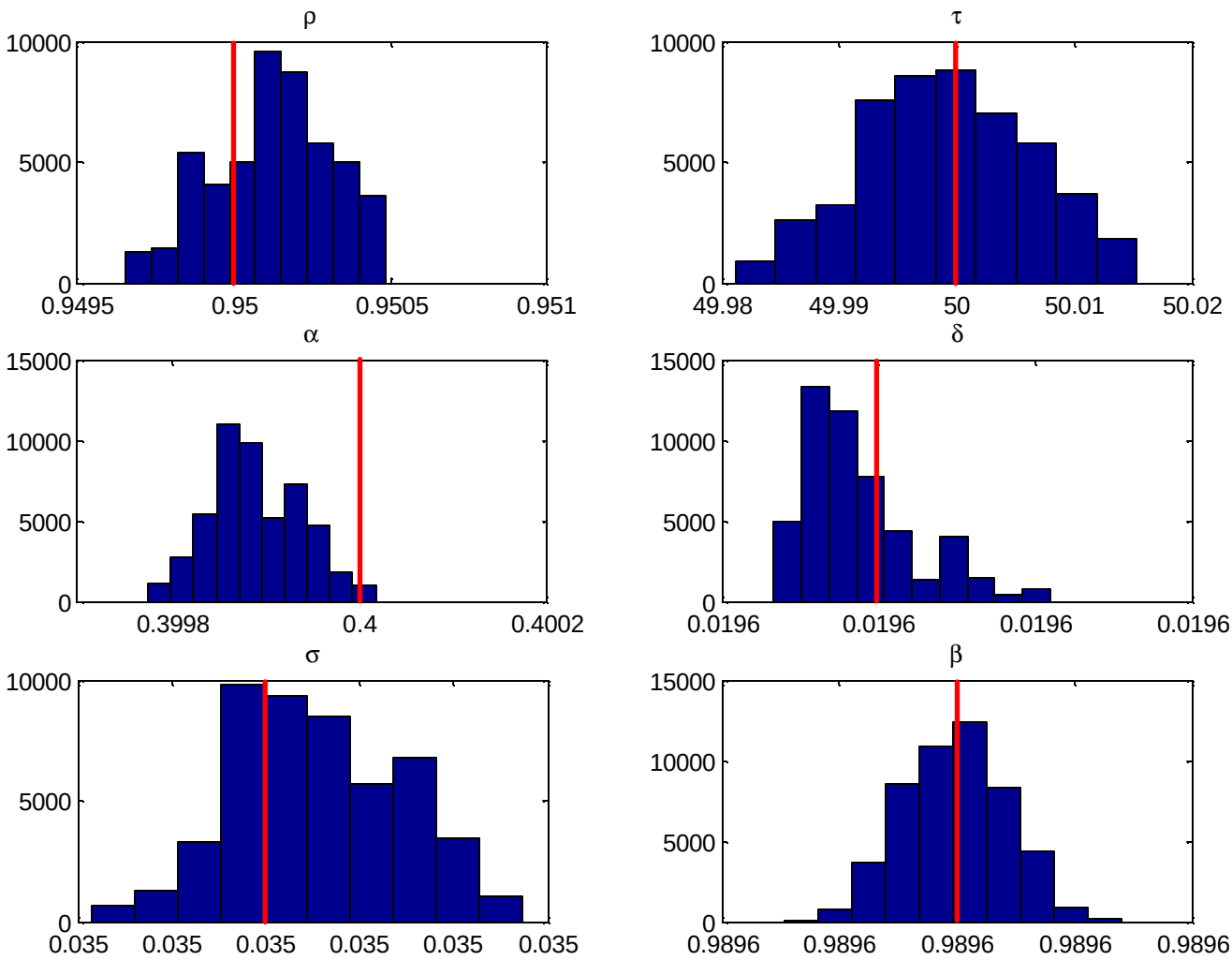

θ
1000
500
0
-500
-1000
-1500
-2000
0.33
0.34
0.35
0.36
0.37
0.38

Figure 3.5: Posteriors Extreme Calibration



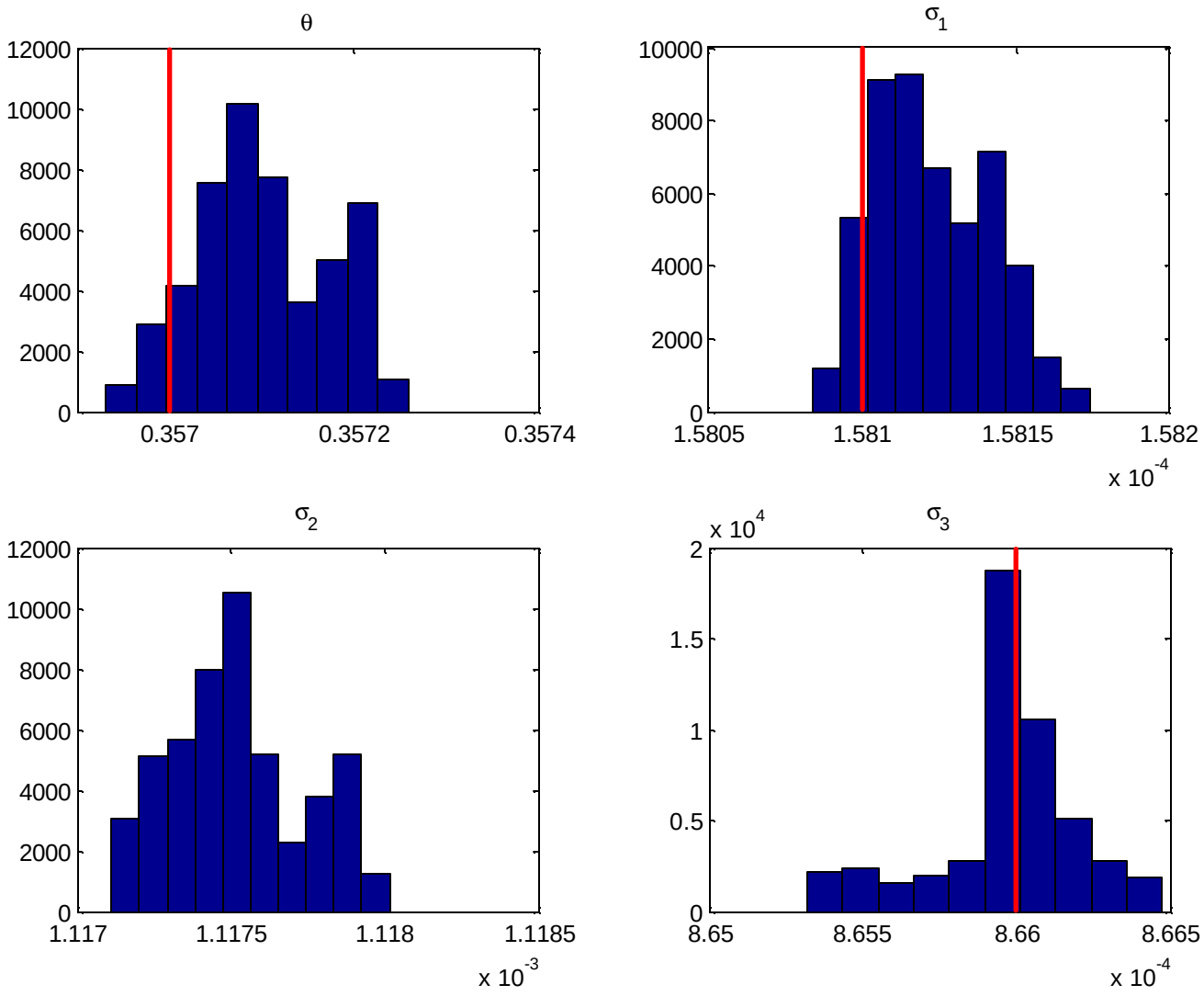
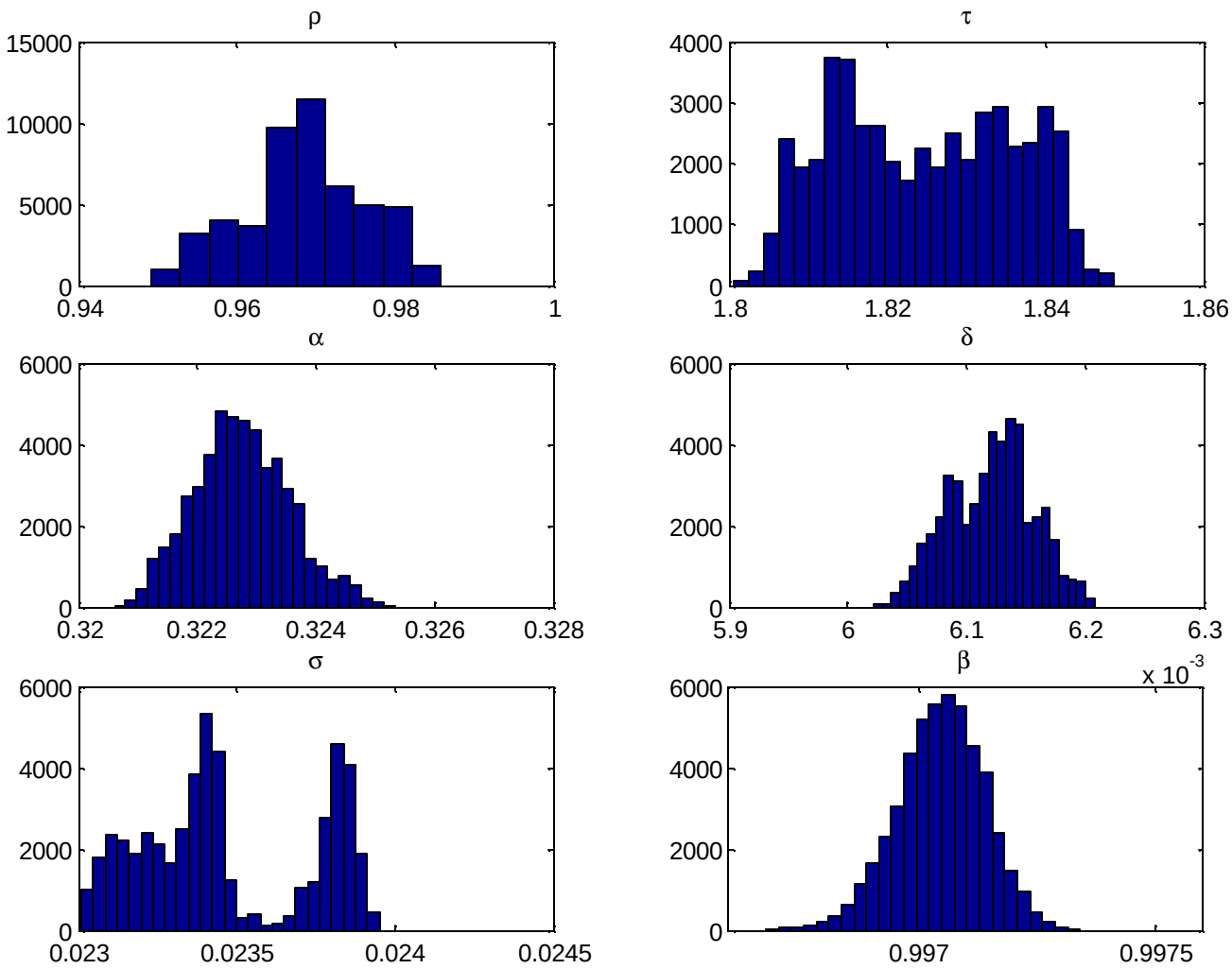

θ
12000
10000
8000
6000
4000
2000
0
0.357
0.3572
0.3574
σ1
10000
8000
6000
4000
2000
0
1.5805
1.581
1.5815
1.582
x 10-4
σ2
12000
10000
8000
6000
4000
2000
0
1.117
1.1175
1.118
1.1185
x 10-3
σ3
x 104
2
1.5
1
0.5
0
8.65
8.655
8.66
8.665
x 10-4

Figure 3.6: Posteriors Real Data



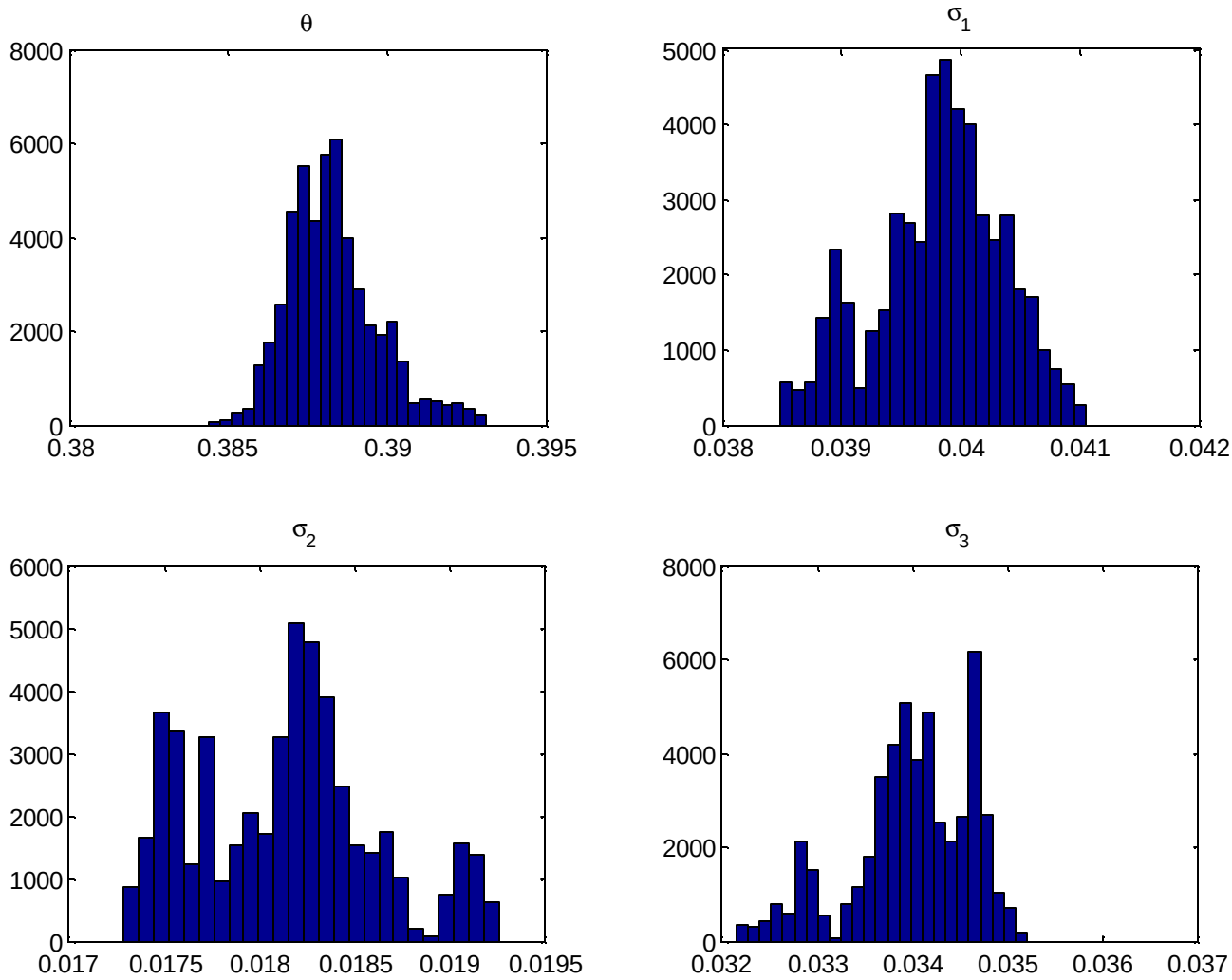

θ
8000
6000
4000
2000
0
0.38
0.385
0.39
0.395
σ1
5000
4000
3000
2000
1000
0
0.038
0.039
0.04
0.041
0.042
σ2
6000
5000
4000
3000
2000
1000
0
0.017
0.0175
0.018
0.0185
0.019
0.0195
σ3
8000
6000
4000
2000
0
0.032
0.033
0.034
0.035
0.036
0.037

Figure 3.7: CDF Benchmark Calibration

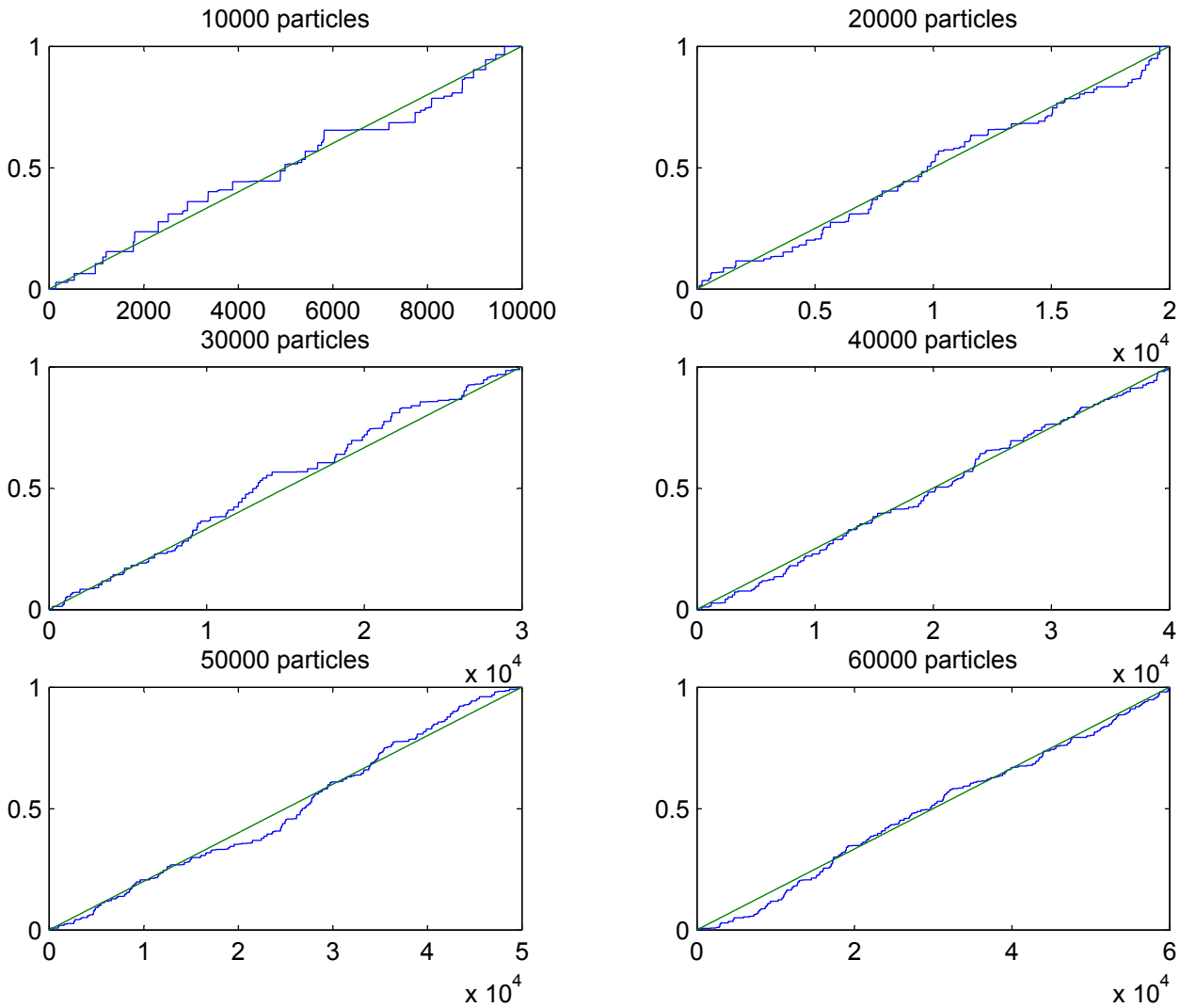


Figure 3.8: CDF Extreme Calibration

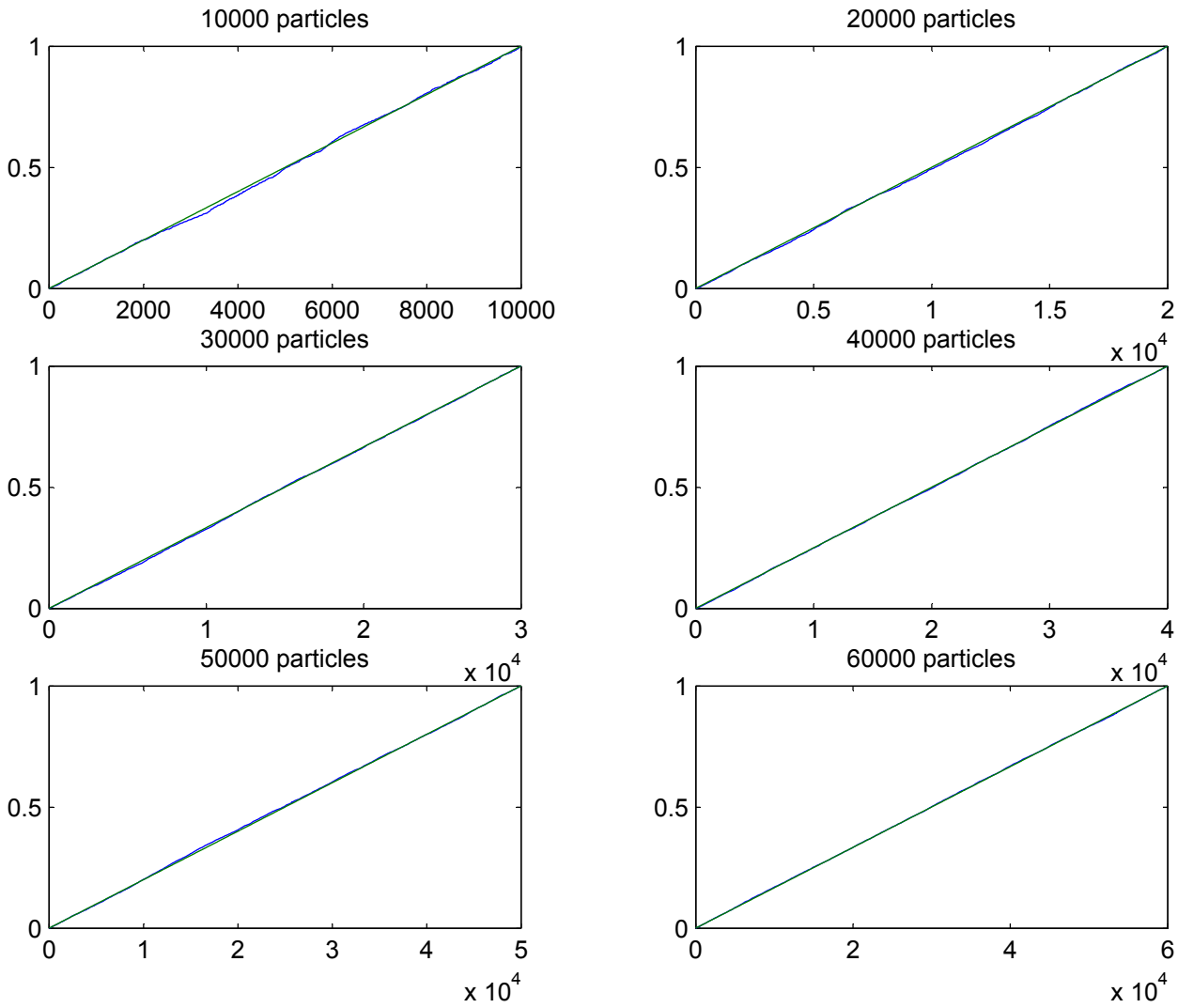


Figure 3.9: CDF Real Data

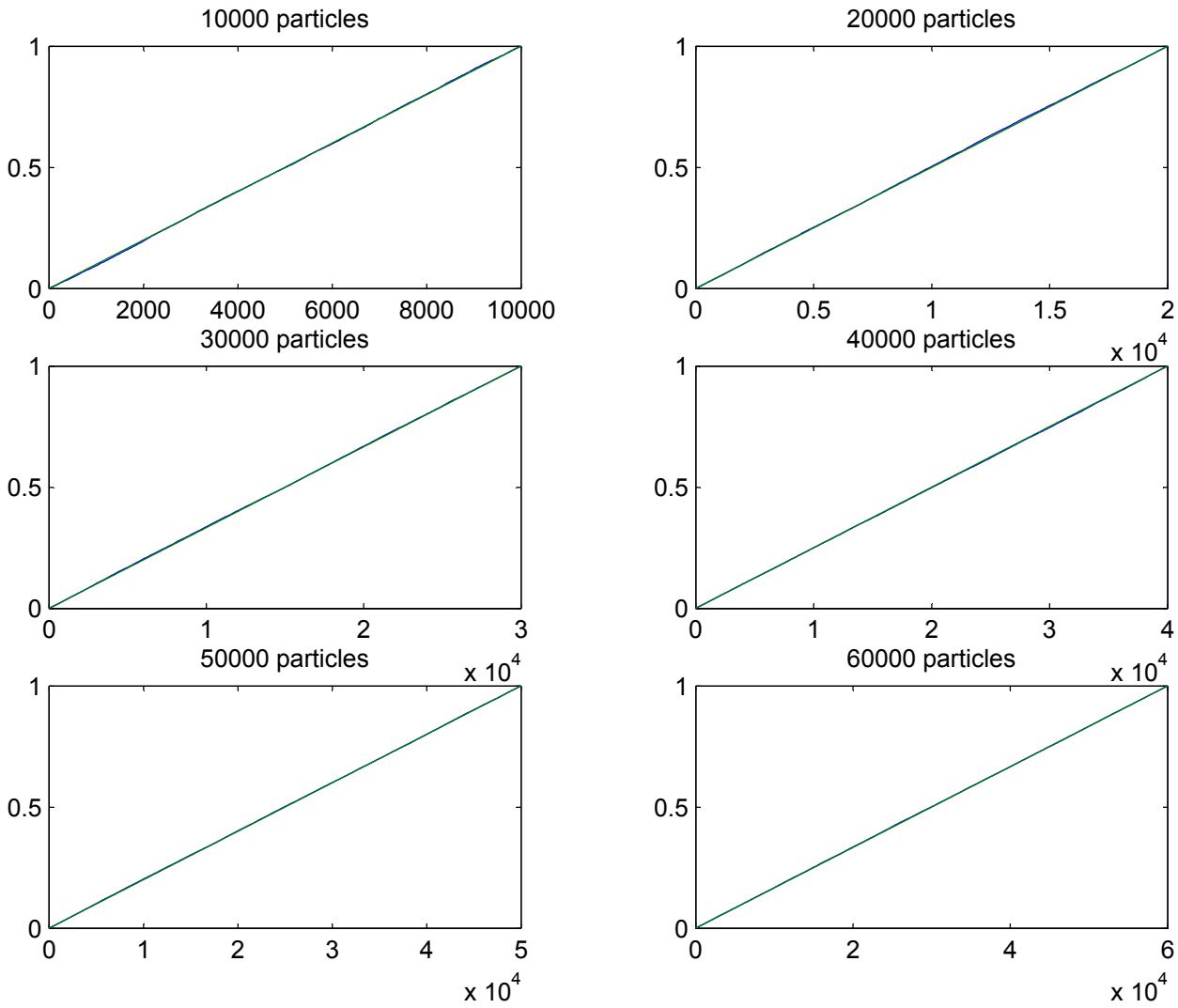
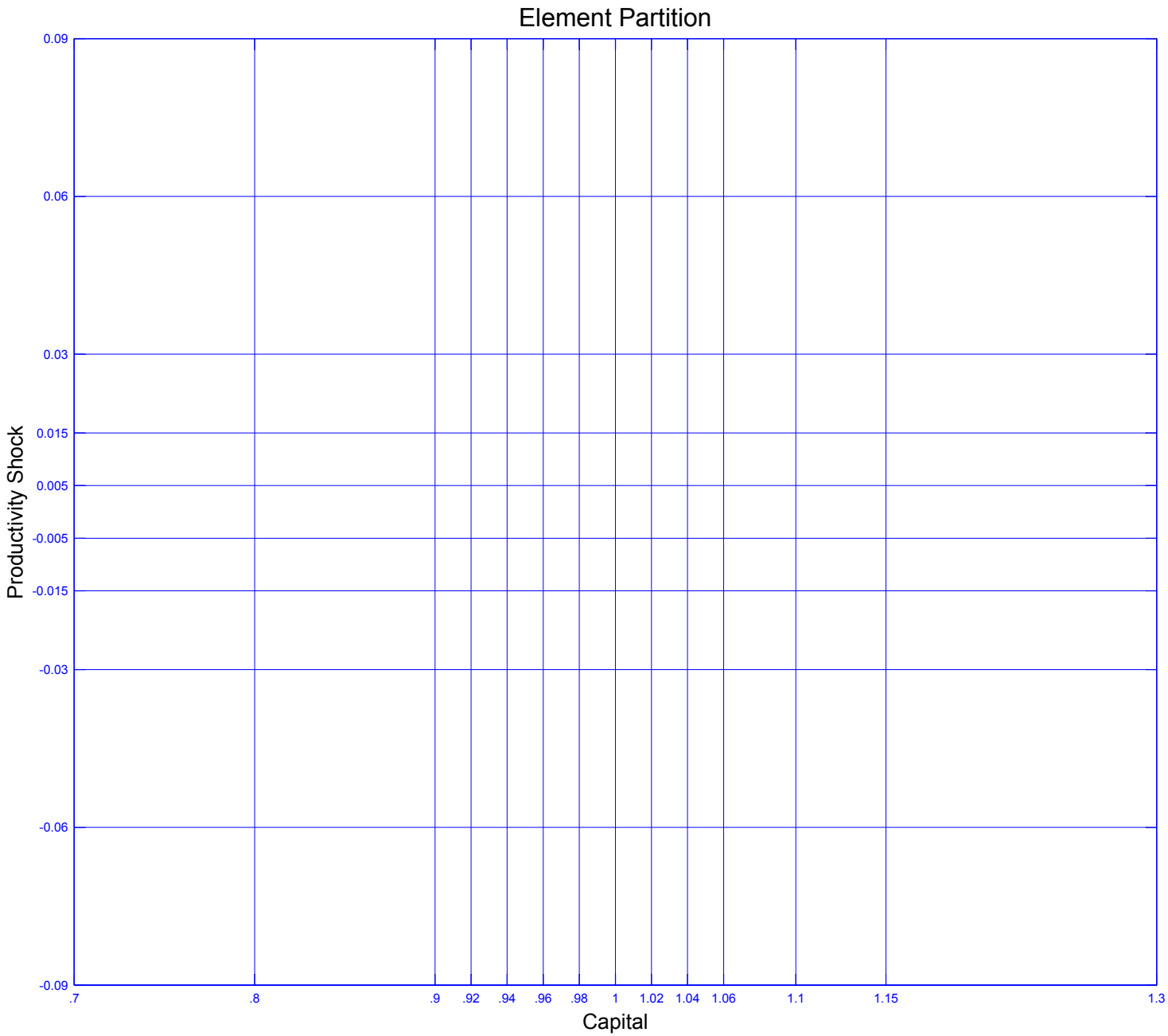



Figure A1: Finite Element Partition