Estimating Macroeconomic Models: A Likelihood Approach*

Jesús Fernández-Villaverde

University of Pennsylvania, NBER, and CEPR

Juan F. Rubio-Ramírez

Federal Reserve Bank of Atlanta

January 24, 2006

*Corresponding author: Juan F. Rubio-Ramírez, Federal Reserve Bank of Atlanta, Research Department, 1000 Peachtree St NE, Atlanta, GA 30309-4470, USA. E-mail: juan.rubio@atl.frb.org. We thank Manuel Arellano, Will Roberds, Eric Renault, Barbara Rossi, Stephanie Schmitt-Grohé, Enrique Sentana, Chris Sims, Tao Zha, two referees, and participants at several seminars for comments. We also thank Jonas Fisher for providing us with his investment deflator. Beyond the usual disclaimer, we must note that any views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

Abstract

This paper shows how particle filtering allows us to undertake likelihood-based inference in dynamic macroeconomic models. The models can be nonlinear and/or non-normal. We describe how to use the output from the particle filter to estimate the structural parameters of the model, those characterizing preferences and technology, and to compare different economies. Both tasks can be implemented from either a classical or a Bayesian perspective. We illustrate the technique by estimating a business cycle model with investment-specific technological change, preference shocks, and stochastic volatility.

Keywords: Dynamic Macroeconomic Models, Particle Filtering, Nonlinear and/or Non-normal Models, Business Cycle, Stochastic Volatility.

JEL classification Numbers: C11, C15, E10, E32.

1. Introduction

This paper shows how particle filtering allows us to undertake likelihood-based inference in dynamic macroeconomic models. The models can be nonlinear and/or non-normal. We describe how to use the particle filter to estimate the structural parameters of the model, those characterizing preferences and technology, and to compare different economies. Both tasks can be implemented from either a classical or a Bayesian perspective. We illustrate the technique by estimating a business cycle model with investment-specific technological change, preference shocks, and stochastic volatility. We report three main results. First, there is strong evidence for the presence of stochastic volatility on U.S. data. Second, the decline in aggregate volatility has been a gradual trend and not the result of an abrupt change in the mid 1980s, as suggested by the literature. It started in the late 1950s, was interrupted in the late 1960s and 1970s, and resumed around 1979. Third, variations in the volatility of preferences shocks can account for most of the variation in the volatility of growth in U.S. real output per capita over the last 50 years.

Macroeconomists now routinely build dynamic models to answer quantitative questions. To estimate these economies, the literature has been forced either to exploit methods of moments or to linearize the model solution and evaluate the implied approximated likelihood with the Kalman filter. This situation is unsatisfactory. Methods of moments suffer from small-sample biases and may not efficiently employ the available information. Linearization techniques depend on the accurate approximation of the exact policy function by a linear relation and on the presence of normal shocks.

The impact of linearization is grimmer than it appears. Fernández-Villaverde, Rubio-Ramírez, and Santos (2006) prove that second order approximation errors in the solution of the model have first order effects on the likelihood function. Moreover, the error in the approximated likelihood gets compounded with the size of the sample. Period by period, small errors in the policy function accumulate at the same rate at which the sample size grows. Therefore, the likelihood implied by the linearized model diverges from the likelihood implied by the exact model. In Fernández-Villaverde and Rubio-Ramírez (2005), we document how those insights are quantitatively relevant for real-life applications. Using U.S. data, we estimate the neoclassical growth model with two methods: the particle filter described in this paper and the Kalman filter on a linearized version of the model. We uncover significant differences on the parameter estimates, on the level of the likelihood, and on the moments generated by the model. These findings are relevant because they highlight how linearization has a tremendous impact on inference, even for nearly linear economies as the neoclassical growth model.

Finally, the assumption of normal shocks precludes investigating models with fat tails distributions (for example, with student-t's innovations), time-varying volatility, autoregressive conditional duration, and others that are of interest to address many empirical questions.

The main obstacle to likelihood-based inference is the difficulty in evaluating the likelihood function implied by a nonlinear and/or non-normal macroeconomic model. Beyond a few particular cases, it is not possible to perform this evaluation analytically or numerically.[1] Methods of moments avoid the problem by moving away from the likelihood. Linearization fails to evaluate the exact likelihood function of the model and computes instead the likelihood of a linear approximation to the economy.

We use a particle filter to solve the problem of evaluating the likelihood of nonlinear and/or non-normal macroeconomic models (although the algorithm is general enough to handle linear models with or without normal shocks). To do so, we borrow from the growing literature on Sequential Monte Carlo methods (see the book-length review by Doucet, de Freitas, and Gordon, 2001). In economics, particle filters have been applied by Pitt and Shephard (1999) and Kim, Shephard, and Chib (1998) to the estimation of stochastic volatility models in financial econometrics. We adapt this know-how to handle the peculiarities of the likelihood of macroeconomic models. We propose and exploit in our application a novel partition of the shocks that drive the model. This partition facilitates the estimation of some models while being general enough to encompass existing particle filters.

The general idea of the procedure follows. First, for given values of the parameters, we compute the optimal policy functions of the model with a nonlinear solution method. The researcher can employ the solution method that best fits her needs in terms of accuracy, complexity, and speed. With the policy functions, we construct the state space representation of the model. Under mild conditions, we apply a particle filter to this state space form to evaluate the likelihood function of the model. Then, we either maximize the likelihood function or, after specifying priors on the parameters, find posterior distributions with a Markov chain Monte Carlo (McMc) algorithm. If we carry out the procedure with several models, we could compare them by building either likelihood ratios (Rivers and Vuong, 2002) or Bayes factors (Fernández-Villaverde and Rubio-Ramírez, 2004), even if the models are misspecified and nonnested.

Particle filtering is both reasonably general purpose and asymptotically efficient. Therefore, it is an improvement over approaches that rely on features of a particular model, like Miranda and Rui (1997) or Landon-Lane (1999), and hence are difficult to generalize. It is

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

also an improvement over methods of moments, which are asymptotically less efficient than the likelihood (except in the few cases pointed out by Carrasco and Florens, 2002). Fermanian and Salanié's procedure (2004) shares the general-purpose and asymptotically efficiency characteristics of particle filters. However, the particle filter avoids the kernel estimation required by their nonparametric simulated likelihood method, which is difficult to implement in models with a large number of observables.

Being able to perform likelihood-based inference is important for several additional reasons. First, the likelihood principle states that all the evidence in the data is contained in the likelihood function (Berger and Wolpert, 1988). Second, likelihood-based inference is a simple way to deal with misspecified models (Monfort, 1996). Macroeconomic models are false by construction, and likelihood-based inference has both attractive asymptotic properties and good small-sample behavior under misspecification (see White, 1994, for a classical approach and Fernández-Villaverde and Rubio-Ramírez, 2004, for Bayesian procedures). Furthermore, likelihood inference allows us to compare models. We do not argue that a likelihood approach is always preferable. There are many instances where because of computational simplicity, or robustness, or because the model is incompletely specified, a method of moments is more suitable. We simply maintain that, in numerous contexts, the likelihood is an informative tool.

To illustrate our discussion, we estimate a business cycle model of the U.S. economy. Greenwood, Herkowitz, and Krusell (1997 and 2000) have vigorously defended the importance of technological change specific to new investment goods for understanding postwar U.S. growth and aggregate fluctuations. We follow their lead and estimate a version of the neoclassical growth model modified to include a shock to investment, a shock to preferences, two unit roots, cointegration relations derived from the balanced growth path properties of the model, and stochastic volatility on the economic shocks that drive the dynamics of the economy.

Introducing stochastic volatility is convenient for two reasons. First, the evidence accumulated by Kim and Nelson (1999), McConnell and Pérez-Quirós (2000), Stock and Watson (2002), and Sims and Zha (2005) among others strongly suggests that an assessment of volatility is of first order transcendence for modelling U.S. aggregate time series. This makes the application of interest per se. Second, stochastic volatility induces both fundamental nonlinearities in the law of motion for states and non-normal distributions. If we linearized the laws of motions for shocks to apply the Kalman filter, the stochastic volatility terms would drop, killing any possibility of exploring this mechanism. Thus, the Kalman filter not only induces an approximation error, but more important, it makes it impossible to learn about time-varying volatility. With our business cycle model, we demonstrate how the particle filter is an important tool to address empirical questions at the core of macroeconomics.

In our estimation, we identify the process driving investment-specific technology shocks through the relative price of new equipment to consumption and the neutral technology and preference shock from the log difference of real output per capita, the real gross investment per capita, and the level of hours worked per capita. The data reveal three patterns. First, there is compelling evidence that stochastic volatility is key to understanding the dynamics of U.S. aggregate time series. Second, the decline in aggregate volatility has been a gradual process since the late 1950s, interrupted only by the turbulence of the 1970s. Third, the reduction in the preference shock is a plausible explanation for the increase and posterior fall in the volatility of growth in U.S. real output per capita over the last 50 years. In addition, we provide evidence of how inference is affected both by the nonlinear component of the solution and by the stochastic volatility part, reinforcing the message of Fernández-Villaverde, Rubio-Ramírez, and Santos (2006) and Fernández-Villaverde and Rubio-Ramírez (2005).

Methodologically, our paper builds on the literature on likelihood-based inference on macroeconomic models, as reviewed, for instance, by An and Schorfheide (2005). Our paper is also related to the literature on simulated likelihood and simulated pseudo-likelihood applied to macroeconomic models. Important examples are Laroque and Salanié (1989, 1993, and 1994). The approach taken in these papers is to minimize a distance function between the observed variables and the conditional expectations, weighted by their conditional variances. We, instead, consider the whole set of moments defined by the likelihood function.

With respect to the application, we are aware of only one other paper that deals with stochastic volatility using a dynamic equilibrium model: the important and fascinating contribution of Justiniano and Primiceri (2005).[2] Their innovative paper estimates a rich New Keynesian model of the business cycle with nominal rigidities and adjustment costs. One difference between our papers is that the particle filter allows us to characterize the nonlinear behavior of the economy induced by stochastic volatility that Justiniano and Primiceri cannot handle. We document how including this nonlinear component is quantitatively important for inference. Moreover, we provide smooth estimates of stochastic volatility.

The rest of the paper is organized as follows. In section 2, we describe the particle filter, how it evaluates the likelihood function of a macroeconomic model, and how to apply it for filtering and smoothing. We present our application in sections 3 to 5 and our results in sections 6 and 7. We discuss computational details in section 8. We conclude in section 9.

[2] There is also a relevant literature on VARs that incorporate time-varying volatility. See, for example, Uhlig (1997), Bernanke and Mihov (1988), Cogley and Sargent (2005), Primiceri (2005), and Sims and Zha (2005). See Laforte (2005) for a related dynamic equilibrium model with Markov regime switching.

2. A Framework for Likelihood Inference

In this section, we describe a framework to estimate and compare a large class of nonlinear and/or non-normal dynamic macroeconomic models using a likelihood approach. Examples of economies in this class are the neoclassical growth model (Cooley and Prescott, 1995), sticky prices models (Woodford, 2003), asset pricing models (Mehra and Prescott, 1985), macro public finance models (Chari, Christiano, and Kehoe, 1994), and regime-switching models (Jermann and Quadrini, 2003), among many others.

All of these economies imply a different joint probability density function for observables given the model's structural parameters. We refer to this density as the likelihood function of the model. The likelihood function is useful for two purposes. First, if we want to estimate the model, we can either obtain point estimates by maximizing the likelihood or, if we specify a prior, find the posterior of the parameters with an McMc algorithm. Second, if we are comparing macroeconomic models, we can do so by building either likelihood ratios or Bayes factors.

In the past, the literature has shown how to write the likelihood function of dynamic macroeconomic models only in a few special cases. For example, we knew how to evaluate the likelihood of a linear model with normal innovations using the Kalman filter. In comparison, particle filtering allows us to evaluate the likelihood of macroeconomic models in a general case, removing a stumbling block for the application of likelihood methods to perform inference.

We structure this section as follows. First, we define the likelihood function of a dynamic macroeconomic model. Second, we present a particle filter to evaluate that likelihood. Third, we link the filter with the estimation of the structural parameters of the model. Fourth, we compare particle filtering with some alternatives. We finish by discussing the smoothing of unobserved states.

2.1. The Likelihood Function of a Dynamic Macroeconomic Model

A large set of dynamic macroeconomic models can be written in the following state space form. First, the equilibrium of the economy is characterized by some states S_t that evolve 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 parameters of the model.

Second, the observables $\mathcal{Y}_t$ are a realization of the random variable $\mathbb{Y}_t$ governed by the measurement equation:

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

where $\{V_t\}$ is a sequence of exogenous independent random variables. The sequences $\{W_t\}$ and $\{V_t\}$ are independent of each other. The random variables W_t and V_t are distributed as $p\left(W_t; \gamma\right)$ and $p\left(V_t; \gamma\right)$. We only require the ability to evaluate these densities. It should be clear that $\gamma \in \Upsilon$ also included any parameters characterizing the distributions of W_t and V_t. Assuming independence of $\{W_t\}$ and $\{V_t\}$ is only for notational convenience. Generalization to more involved stochastic processes is achieved by increasing the dimension of the state space. 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, $\mathbb{Y}_t$ 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 behavior of the model: policy functions, laws of motion for exogenous variables, resource and budget constraints, and so on. Along some dimension, the function g can be the identity mapping if a state is observed without noise. Dynamic macroeconomic models do not generally admit closed-form solutions for functions f and g. Our algorithm requires only a numerical procedure to approximate them.

To fix ideas, we map $\{S_t\}$, $\{W_t\}$, $\{\mathbb{Y}_t\}$, $\{V_t\}$, f, and g into some examples of dynamic macroeconomic models. Consider first the example of the neoclassical growth model. 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 noise. Thus, S_t will be capital and productivity, W_t the shock to productivity, $\mathbb{Y}_t$ 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, $\mathbb{Y}_t$ 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 macroeconomic models can be fitted into the state space formulation.

To continue our analysis we make the following assumptions.

Assumption 1. $\dim(W_t) + \dim(V_t) \geq \dim(\mathbb{Y}_t)$ *for all* t.

This assumption is a necessary condition for the model not to be stochastically singular. We do not impose any restrictions on how those degrees of stochasticity are achieved.[3]

Now we provide some definitions that will be useful in the rest of the paper. To be able to deal with a larger class of macroeconomic models, we partition $\{W_t\}$ into two sequences $\{W_{1,t}\}$ and $\{W_{2,t}\}$, such that $W_t = (W_{1,t}, W_{2,t})$ and $\dim(W_{2,t}) + \dim(V_t) = \dim(\mathbb{Y}_t)$. The sequence $\{W_{2,t}\}$ is the part of $\{W_t\}$ necessary to keep the system stochastically nonsingular. If $\dim(V_t) = \dim(\mathbb{Y}_t)$, we set $W_{1,t} = W_t\ \forall t$, i.e., $\{W_{2,t}\}$ is a zero-dimensional sequence. If $\dim(W_t) + \dim(V_t) = \dim(\mathbb{Y}_t)$, we set $W_{2,t} = W_t$ for $\forall t$, i.e., $\{W_{1,t}\}$ is a zero-dimensional sequence. Also, 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 for $\forall t$. 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 $\mathcal{Y}^t = \{\mathcal{Y}_m\}_{m=1}^t$ and let $\mathcal{Y}^t$ be a realization of the random variable $\mathbb{Y}^t = \{\mathbb{Y}_m\}_{m=1}^t$ for $\forall t$. Finally, we define $W_i^0 = \{\emptyset\}$ and $\mathcal{Y}^0 = \{\emptyset\}$.

Our goal is to evaluate the likelihood function of a sequence of realizations of the observable $\mathcal{Y}^T$ at a particular parameter value γ:

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

In general the likelihood function (3) cannot be computed analytically. The particle filter relies on simulation methods to estimate it. Our first step is to factor the likelihood as:

$$p\left(\mathcal{Y}^T;\gamma\right) = \prod_{t=1}^{T} p\left(\mathcal{Y}_t|\mathcal{Y}^{t-1};\gamma\right) = \prod_{t=1}^{T} \int\int p\left(\mathcal{Y}_t|W_1^t, S_0, \mathcal{Y}^{t-1};\gamma\right) p\left(W_1^t, S_0|\mathcal{Y}^{t-1};\gamma\right) dW_1^t dS_0, \tag{4}$$

where S_0 is the initial state of the model, the p's represent the relevant densities, and where in the case $\{W_{1t}\}$ has zero dimensions $\int\int p\left(\mathcal{Y}_t|W_1^t, S_0, \mathcal{Y}^{t-1};\gamma\right) p\left(W_1^t, S_0|\mathcal{Y}^{t-1};\gamma\right) dW_1^t dS_0 = \int p\left(\mathcal{Y}_t|S_0, \mathcal{Y}^{t-1};\gamma\right) p\left(S_0|\mathcal{Y}^{t-1};\gamma\right) dS_0$. 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.

[3] This paper does not contribute to the literature on how to solve the problem of stochastic singularity of dynamic macroeconomic models. There are two routes 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 specify a model rich in stochastic dynamics (Smets and Wouters, 2003). This alternative is attractive for addressing practical policy questions like those of interest to central banks.

The previous expression shows that the problem of evaluating the likelihood (3) amounts to solving an integral, with conditional densities $p\left(\mathcal{Y}_t|W_1^t, S_0, \mathcal{Y}^{t-1};\gamma\right)$ and $p\left(W_1^t, S_0|\mathcal{Y}^{t-1};\gamma\right)$ that are difficult to characterize. When the state space representation is linear and normal, the integral simplifies notably because all the relevant densities are conditionally normal. Then, tracking the mean and variance-covariance matrix of the densities is enough to compute the likelihood. The Kalman filter accomplishes this objective efficiently through the Riccati equations. However, when the state representation is nonlinear and/or non-normal, the conditional densities are not any longer normal, and we require a more powerful tool than the Kalman filter to evaluate the likelihood.

Before continuing, we present two additional technical assumptions.

Assumption 2. *For all γ, s_0, w_1^t, and t, the following system of equations:*

$$S_1 = f\left(s_0, \left(w_{1,1}, W_{2,1}\right);\gamma\right)$$

$$\mathcal{Y}_m = g\left(S_m, V_m;\gamma\right) \qquad \text{for } m = 1, 2, ...t$$

$$S_m = f\left(S_{m-1}, \left(w_{1,m}, W_{2,m}\right);\gamma\right) \qquad \text{for } m = 2, 3, ...t$$

has a unique solution, $\left(v^t\left(w_1^t, s_0, \mathcal{Y}^t;\gamma\right), s^t\left(w_1^t, s_0, \mathcal{Y}^t;\gamma\right), w_2^t\left(w_1^t, s_0, \mathcal{Y}^t;\gamma\right)\right)$, and we can evaluate the probabilities $p\left(v^t\left(w_1^t, s_0, \mathcal{Y}^t;\gamma\right);\gamma\right)$ and $p\left(w_2^t\left(w_1^t, s_0, \mathcal{Y}^t;\gamma\right);\gamma\right)$.

Assumption 2 implies that we can evaluate the conditional densities $p\left(\mathcal{Y}_t|w_1^t, s_0, \mathcal{Y}^{t-1};\gamma\right)$ for all γ, s_0, w_1^t, and t. To simplify the notation, we write $\left(v^t, s^t, w_2^t\right)$, instead of the more cumbersome $\left(v^t\left(w_1^t, s_0, \mathcal{Y}^t;\gamma\right), s^t\left(w_1^t, s_0, \mathcal{Y}^t;\gamma\right), w_2^t\left(w_1^t, s_0, \mathcal{Y}^t;\gamma\right)\right)$. Then, we have:

$$p\left(\mathcal{Y}_t|w_1^t, s_0, \mathcal{Y}^{t-1};\gamma\right) = p\left(v_t, w_{2,t}|w_1^t, s_0, \mathcal{Y}^{t-1};\gamma\right)\left|dy\left(v_t, w_{2,t};\gamma\right)\right| \tag{5}$$

for all γ, s_0, w_1^t, and t, where $\left|dy\left(v_t, w_{2,t};\gamma\right)\right|$ stands for the determinant of the jacobian of $\mathbb{Y}_t$ with respect to V_t and $W_{2,t}$ evaluated at v_t and $w_{2,t}$. Note that assumption 2 requires only the ability to evaluate the density; it does not require having a closed form for it. Thus, we may employ numerical or simulation methods for this evaluation if this is convenient.

The most important implication of (5) is that, to compute $p\left(\mathcal{Y}_t|w_1^t, s_0, \mathcal{Y}^{t-1};\gamma\right)$, we only need to solve a system of equations and evaluate the probability of observing the solution to the system, $p\left(v_t, w_{2,t}|w_1^t, s_0, \mathcal{Y}^{t-1};\gamma\right)$, times the determinant of a jacobian evaluated at the solution, $\left|dy\left(v_t, w_{2,t};\gamma\right)\right|$. The evaluation of $p\left(v_t, w_{2,t}|w_1^t, s_0, \mathcal{Y}^{t-1};\gamma\right)$ is always possible by assumption. How difficult is to evaluate the jacobian? Often, this is a simple task because the jacobian depends on f and g, which are functions that we can evaluate numerically, and γ.

For example, if, given γ, we employ a second order perturbation method to solve the model and get f and g, the jacobian is a constant matrix that comes directly from the solution procedure.

To avoid trivial problems, we assume that the model assigns positive probability to the data, y^T. This is formally reflected in the following assumption:

Assumption 3. *For all $\gamma \in \Upsilon$, s_0, w_1^t, and t, the model gives some positive probability to the data $\mathcal{Y}^T$, i.e.,*

$$p\left(\mathcal{Y}_t|w_1^t, s_0, \mathcal{Y}^{t-1}; \gamma\right) > 0,$$

for all $\gamma \in \Upsilon$, s_0, w_1^t, and t.

Assumptions 1 to 3 are necessary and sufficient conditions for the model not to be stochastically singular.

We get now to the core of this section. If assumptions 1 to 3 hold, conditional on having N draws of $\left\{\left\{\widehat{w}_1^{t,i}, \widehat{s}_0^i\right\}_{i=1}^N\right\}_{t=1}^T$ from the sequence of densities $\left\{p\left(W_1^t, S_0|\mathcal{Y}^{t-1}; \gamma\right)\right\}_{t=1}^T$ (note that the hats and the superindex i on the variables denote a draw), the likelihood function (4) is approximated by:

$$p\left(\mathcal{Y}^T; \gamma\right) \simeq \prod_{t=1}^T \frac{1}{N} \sum_{i=1}^N p\left(\mathcal{Y}_t|\widehat{w}_1^{t,i}, \widehat{s}_0^i, \mathcal{Y}^{t-1}; \gamma\right)$$

because of a law of large numbers. This shows that the problem of evaluating the likelihood of a dynamic model is equivalent to the problem of drawing from $\left\{p\left(W_1^t, S_0|\mathcal{Y}^{t-1}; \gamma\right)\right\}_{t=1}^T$. In the next section, we propose a particle filter to accomplish this objective.

2.2. A Particle Filter

We saw in the previous expression how the evaluation of the likelihood function is equivalent to the problem of drawing from $\left\{p\left(W_1^t, S_0|\mathcal{Y}^{t-1}; \gamma\right)\right\}_{t=1}^T$. Why is it difficult to draw from $\left\{p\left(W_1^t, S_0|\mathcal{Y}^{t-1}; \gamma\right)\right\}_{t=1}^T$? Because, as we mentioned before, when the model is nonlinear and/or non-normal, this conditional density is a complicated function of $\mathcal{Y}^{t-1}$. The goal of the particle filter is to draw efficiently from $\left\{p\left(W_1^t, S_0|\mathcal{Y}^{t-1}; \gamma\right)\right\}_{t=1}^T$.

Before introducing the particle filter, we fix some additional notation. Let $\left\{w_1^{t-1,i}, s_0^{t-1,i}\right\}_{i=1}^N$ be a sequence of N i.i.d. draws from $p\left(W_1^{t-1}, S_0|\mathcal{Y}^{t-1}; \gamma\right)$. Let $\left\{w_1^{t|t-1,i}, s_0^{t|t-1,i}\right\}_{i=1}^N$ be a sequence of N i.i.d. draws from $p\left(W_1^t, S_0|\mathcal{Y}^{t-1}; \gamma\right)$. We call each draw $\left(w_1^{t,i}, s_0^{t,i}\right)$ a *particle* and the sequence $\left\{w_1^{t,i}, s_0^{t,i}\right\}_{i=1}^N$ a *swarm of particles*. Also, let $h\left(S_t\right)$ be any measurable function for which the expectation

$$E_{p\left(W_1^t,S_0|\mathcal{Y}^t;\gamma\right)}\left(h\left(W_1^t,S_0\right)\right)=\int h\left(W_1^t,S_0\right)p\left(W_1^t,S_0|\mathcal{Y}^t;\gamma\right)dW_1^t dS_0$$

exists and is finite.

The following proposition, a simple and well-known application of importance sampling (e.g., Geweke, 1989, Theorem 1), is key for further results.

Proposition 4. *Let* $\left\{w_1^{t|t-1,i},s_0^{t|t-1,i}\right\}_{i=1}^N$ *be a draw from* $p\left(W_1^t,S_0|\mathcal{Y}^{t-1};\gamma\right)$ *and the weights:*

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

Then:

$$E_{p\left(W_1^t,S_0|\mathcal{Y}^t;\gamma\right)}\left(h\left(W_1^t,S_0\right)\right)\simeq\sum_{i=1}^N q_t^i h\left(w_1^{t|t-1,i},s_0^{t|t-1,i}\right).$$

Proof. By Bayes' theorem:

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

Therefore, if we take $p\left(W_1^t,S_0|\mathcal{Y}^{t-1};\gamma\right)$ as an importance sampling function to draw from the density $p\left(W_1^t,S_0|\mathcal{Y}^t;\gamma\right)$, the result is a direct consequence of the law of large numbers. ■

Rubin (1988) proposed to combine proposition 4 and $p\left(W_1^t,S_0|\mathcal{Y}^{t-1};\gamma\right)$ to draw from $p\left(W_1^t,S_0|\mathcal{Y}^t;\gamma\right)$ in the following way:

Corollary 5. *Let* $\left\{w_1^{t|t-1,i},s_0^{t|t-1,i}\right\}_{i=1}^N$ *be a draw from* $p\left(W_1^t,S_0|\mathcal{Y}^{t-1};\gamma\right)$. *Let the sequence* $\{\widetilde{w}_1^i,\widetilde{s}_0^i\}_{i=1}^N$ *be a draw with replacement from* $\left\{w_1^{t|t-1,i},s_0^{t|t-1,i}\right\}_{i=1}^N$ *where* q_t^i *is the probability of* $\left(w_1^{t|t-1,i},s_0^{t|t-1,i}\right)$ *being drawn* $\forall i$. *Then* $\{\widetilde{w}_1^i,\widetilde{s}_0^i\}_{i=1}^N$ *is a draw from* $p\left(W_1^t,S_0|\mathcal{Y}^t;\gamma\right)$.

Corollary 5 shows how a draw $\left\{w_1^{t|t-1,i},s_0^{t|t-1,i}\right\}_{i=1}^N$ from $p\left(W_1^t,S_0|\mathcal{Y}^{t-1};\gamma\right)$ can be used to get a draw $\left\{w_1^{t,i},s_0^{t,i}\right\}_{i=1}^N$ from $p\left(W_1^t,S_0|\mathcal{Y}^t;\gamma\right)$. How do we get the swarm $\left\{w_1^{t|t-1,i},s_0^{t|t-1,i}\right\}_{i=1}^N$? By taking a swarm $\left\{w_1^{t-1,i},s_0^{t-1,i}\right\}_{i=1}^N$ from $p\left(W_1^{t-1},S_0|\mathcal{Y}^{t-1};\gamma\right)$ and augmenting it with draws from $p\left(W_{1,t};\gamma\right)$ since $p\left(W_1^t,S_0|\mathcal{Y}^{t-1};\gamma\right)=p\left(W_{1,t};\gamma\right)p\left(W_1^{t-1},S_0|\mathcal{Y}^{t-1};\gamma\right)$. Note that $w_1^{t|t-1,i}$ is a growing object with t (it has the additional component of the draw from $p\left(W_{1,t};\gamma\right)$), while $s_0^{t|t-1,i}$ is not. Corollary 5 is crucial for the implementation of the particle filter. We discussed before how, when the model is nonlinear and/or non-normal, the particle filter keeps track of a set of draws from $p\left(W_1^t, S_0|\mathcal{Y}^{t-1};\gamma\right)$ that are updated as new information is available. Corollary 5 shows how importance resampling solves the problem of updating the draws in such a way that we keep the right conditioning.

This recursive structure is summarized in the following pseudo-code for the particle filter:

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

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

Step 2, Filtering: `Assign to each draw` $\left(w_1^{t|t-1,i}, s_0^{t|t-1,i}\right)$ `the weight` q_t^i `defined in proposition 4.`

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

With the output of the algorithm, $\left\{\left\{w_1^{t|t-1,i}, s_0^{t|t-1,i}\right\}_{i=1}^{N}\right\}_{t=1}^{T}$, we compute the likelihood function as:

$$p\left(\mathcal{Y}^T;\gamma\right) \simeq \frac{1}{N}\left(\prod_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N} p\left(\mathcal{Y}_t|w_1^{t|t-1,i}, s_0^{t|t-1,i}, \mathcal{Y}^{t-1};\gamma\right)\right). \tag{6}$$

Since the particle filter does not require any assumption on the distribution of the shocks except the ability to evaluate $p\left(\mathcal{Y}_t|W_1^t, S_0, \mathcal{Y}^{t-1};\gamma\right)$, either analytically or numerically, the algorithm works effortlessly with non-normal innovations. Del Moral and Jacod (2002) and Künsch (2005) provide weak conditions under which the right-hand side of (6) is a consistent estimator of $p\left(\mathcal{Y}^T;\gamma\right)$ and a Central Limit Theorem applies.

The algorithm presented above belongs to the class of particle filters described by Doucet, de Freitas, and Gordon (2001). We modify existing procedures to deal with more general classes of state space representations than the ones addressed in the literature. In particular, through our partition of W_t, we handle those cases, common in macroeconomics, where $\dim\left(V_t\right) < \dim\left(\mathbb{Y}_t\right)$. We consider this more general applicability of our procedure an important advance.

Our partition of the shocks raises a question: Do the identities of $\{W_{1,t}\}$ and $\{W_{2,t}\}$ matter for the results presented in this section? The short answer is no. If, for example, $\dim\left(W_t\right) = 2$ and $\dim\left(W_{1,t}\right) = \dim\left(W_{2,t}\right) = 1$, we can exchange the identities of $\{W_{1,t}\}$ and $\{W_{2,t}\}$ without affecting the theoretical results. Of course, the identities of $\{W_{1,t}\}$ and $\{W_{2,t}\}$

will affect the results for any finite number of particles, but as the number of particles grows, this problem vanishes. Luckily, as is the case with our application below, often there is a natural choice of $\{W_{2,t}\}$ and, therefore, of $\{W_{1,t}\}$.

The intuition of the algorithm is as follows. Given a swarm of particles up to period $t-1$, $\left\{w_1^{t-1,i}, s_0^{t-1,i}\right\}_{i=1}^N$, distributed according to $p\left(W_1^{t-1}, S_0|\mathcal{Y}^{t-1};\gamma\right)$, the **Prediction Step** generates draws $\left\{w_1^{t|t-1,i}, s_0^{t|t-1,i}\right\}_{i=1}^N$ from $p\left(W_1^t, S_0|\mathcal{Y}^{t-1};\gamma\right)$. In the case where $\dim\left(W_{1,t}\right) = 0$, the algorithm skips this step. The **Sampling Step** takes advantage of corollary 5 and resamples from $\left\{w_1^{t|t-1,i}, s_0^{t|t-1,i}\right\}_{i=1}^N$ with the weights $\{q_t^i\}_{i=1}^N$ to draw a new swarm of particles up to period t, $\left\{w_1^{t,i}, s_0^{t,i}\right\}_{i=1}^N$distributed according to $p\left(W_1^t, S_0|\mathcal{Y}^t;\gamma\right)$. The best procedure for resampling in terms of minimizing Monte Carlo variance (and the one we implement in our application below) is known as systematic resampling (Kitagawa, 1996). This procedure matches the weights of each proposed particle with the number of times each particle is accepted. Finally, applying again the Prediction Step, we generate draws $\left\{w_1^{t+1|t,i}, s_0^{t+1|t,i}\right\}_{i=1}^N$ from $p\left(W_1^{t+1}, S_0|\mathcal{Y}^t;\gamma\right)$ and close the algorithm.

The **Sampling Step** is the heart of the algorithm. If we avoid this step and just weight each draw in $\left\{w_1^{t|t-1,i}, s_0^{t|t-1,i}\right\}_{i=1}^N$ by $\{Nq_t^i\}_{i=1}^N$, we have the so-called Sequential Importance Sampling (SIS). The problem with SIS is that $q_t^i \to 0$ for all i but one particular i' as $t \to \infty$ if $\dim\left(W_{1,t}\right) > 0$ (Arulampalam *et al.*, 2002, pp. 178-179 and references there). The reason is that all the sequences become arbitrarily far away from the true sequence of states, which is a zero measure set. The sequence that happens to be closer dominates all the remaining ones in weight. In practice, after a few steps, the distribution of importance weights becomes heavily skewed, and after a moderate number of steps, only one sequence has a nonzero weight. For example, Fernández-Villaverde and Rubio-Ramírez (2006) find that the degeneracy appears after only 20 periods. Since samples in macroeconomics are relatively long (200 observations or so), the degeneracy of SIS is a serious problem.

Several points deserve further discussion. First, we can exploit the last value of the particle swarm $\left\{w_1^{T,i}, s_0^{T,i}\right\}_{i=1}^N$ for forecasting, i.e., to make probability statements about future values of the observables. To do so, we draw from $\left\{w_1^{T,i}, s_0^{T,i}\right\}_{i=1}^N$ and by simulating $p\left(W_{1,T+1};\gamma\right)$, $p\left(W_{2,T+1};\gamma\right)$, and $p\left(V_{T+1};\gamma\right)$, we build $p\left(\mathbb{Y}_{T+1}|\mathcal{Y}^T;\gamma\right)$. Second, we need to explain how to draw from $p\left(S_0;\gamma\right)$ in the **Initialization Step**. In general, since we cannot evaluate $p\left(S_0;\gamma\right)$, it is not possible to draw from it with an McMc. Santos and Peralta-Alva (2005) solve this problem by showing how to sample from $p\left(S_0;\gamma\right)$ using the transition and measurement equations (1) and (2). Finally, we emphasize that we are presenting here only a basic particle filter and that the literature has presented several refinements to improve efficiency, taking advantage of some of the particular characteristics of the estimation at hand. See, for example, the Auxiliary Particle Filter of Pitt and Shephard (1999).

2.3. Estimation Algorithms

We now explain how to employ the approximated likelihood function (6) to perform likelihood-based estimations from both a classical and a Bayesian perspective. First, we describe the classical approach, then the Bayesian one.

On the classical side, the main inference tool is the likelihood function and its global maximum. Once the likelihood is approximated by (6), we can maximize it as follows:

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` $L\left(\mathcal{Y}^T;\gamma_i\right)$ `using (6) and get` γ_{i+1} `from a maximization routine.`

Step 3, Stopping Rule: `If` $\left\|L\left(\mathcal{Y}^T;\gamma_i\right)-L\left(\mathcal{Y}^T;\gamma_{i+1}\right)\right\|>\xi$, `where` $\xi>0$ `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 (MLE), with asymptotic variance-covariance matrix $var(\widehat{\gamma}_{MLE})=-\left(\frac{\partial^2 L\left(\mathcal{Y}^T;\widehat{\gamma}_{MLE}\right)}{\partial\gamma\partial\gamma'}\right)^{-1}$. Since, in general, we cannot directly evaluate this second derivative, we will approximate it with standard numerical procedures. The value of the likelihood function at its maximum is also an input when we build likelihood ratios for model comparison.

However, for the MLE to be an unbiased estimator of the (pseudo-)true parameter values, the likelihood $L\left(\mathcal{Y}^T;\gamma\right)$ has to be differentiable with respect to γ. Furthermore, for the asymptotic variance-covariance matrix $var(\widehat{\gamma}_{MLE})$ to equal $-\left(\frac{\partial^2 L\left(\mathcal{Y}^T;\widehat{\gamma}_{MLE}\right)}{\partial\gamma\partial\gamma'}\right)^{-1}$, $L\left(\mathcal{Y}^T;\gamma\right)$ has to be twice differentiable with respect to γ. Remember that the likelihood can be written as:

$$L\left(\mathcal{Y}^T;\gamma\right)=\prod_{t=1}^{T}p\left(\mathcal{Y}_t|\mathcal{Y}^{t-1};\gamma\right)=\int\left(\int\prod_{t=1}^{T}p\left(W_{1,t};\gamma\right)p\left(\mathcal{Y}_t|W_1^t,S_0,\mathcal{Y}^{t-1};\gamma\right)dW_1^t\right)\mu^*\left(dS_0;\gamma\right),$$

where $\mu^*(S;\gamma)$ is the invariant distribution on S of the dynamic model. Thus, to prove that $L\left(\mathcal{Y}^T;\gamma\right)$ is twice differentiable with respect to γ, we need $p\left(W_{1,t};\gamma\right)$, $p\left(\mathcal{Y}_t|W_1^t,S_0,\mathcal{Y}^{t-1};\gamma\right)$, and $\mu^*(S;\gamma)$ to be twice differentiable with respect to γ.

Under standard regularity conditions, we can prove that both $p\left(\mathcal{Y}_t|W_1^t, S_0, \mathcal{Y}^{t-1}; \gamma\right)$ and $p\left(W_{1,t}; \gamma\right)$ are twice differentiable (Fernández-Villaverde, Rubio-Ramírez, and Santos, 2006). The differentiability of $\mu^*\left(dS_0; \gamma\right)$ is a more complicated issue. Except for special cases (Stokey, Lucas, and Prescott, 1989, Theorem 12.13, and Stenflo, 2001), we cannot even show that $\mu^*\left(dS_0; \gamma\right)$ is continuous. Hence, a proof that $\mu^*\left(dS_0; \gamma\right)$ is twice differentiable is a daunting task well beyond the scope of this paper.

The possible lack of twice differentiability of $L\left(\mathcal{Y}^T; \gamma\right)$ creates two problems. First, it may be that the MLE is biased and $var(\widehat{\gamma}_{MLE}) \neq -\left(\frac{\partial^2 L\left(\mathcal{Y}^T; \widehat{\gamma}_{MLE}\right)}{\partial\gamma\partial\gamma'}\right)^{-1}$. Second, Newton's type algorithm may fail to maximize the likelihood function. In our application, we report $-\left(\frac{\partial^2 L\left(\mathcal{Y}^T; \widehat{\gamma}_{MLE}\right)}{\partial\gamma\partial\gamma'}\right)^{-1}$ as the asymptotic variance-covariance matrix, hoping that the true asymptotic variance-covariance matrix is not very different. We avoid the second problem by using a simulated annealing algorithm to maximize the likelihood function.

Even if we were able to prove that $\mu^*\left(dS_0; \gamma\right)$ is twice differentiable and, therefore, the MLE is consistent with the usual variance-covariance matrix, the direct application of the particle filter will not deliver an estimator of the likelihood function that is continuous with respect to the parameters. This is caused by the resampling steps within the particle filter and seems difficult to avoid. Pitt (2002) has developed a promising bootstrap procedure to get an approximating likelihood that is continuous under rather general conditions when the parameter space is unidimensional. Therefore, the next step should be to expand Pitt's (2002) bootstrap method to the multidimensional case.

Another relevant issue is as follows. For the maximum likelihood algorithm to converge, we need to keep the simulated innovations $W_{1,t}$ and the uniform numbers that enter into the resampling decisions constant as we modified the parameter values γ_i. As pointed out by McFadden (1989) and Pakes and Pollard (1989), this is required to achieve stochastic equicontinuity. With this property, the pointwise convergence of the likelihood (6) to the exact likelihood is strengthen to uniform convergence. Then, we can swap the argmax and the lim operators (i.e., as the number of simulated particles converges to infinity, the MLE also converges). Otherwise, we would suffer numerical instabilities induced by the "chatter" of changing random numbers.

In a Bayesian approach, the main inference tool is the posterior distribution of the parameters given the data $\pi\left(\gamma|\mathcal{Y}^T\right)$. Once the posterior distribution is obtained, we can define a loss function to derive a point estimate. 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, $\pi\left(\gamma\right)$, and to evaluate the likelihood function. The next step in Bayesian inference is to find the parameters' posterior. In general, the posterior does not have a closed form. Thus, we use a Metropolis-Hastings algorithm to draw a chain $\{\gamma_i\}_{i=1}^{M}$ from $\pi\left(\gamma|\mathcal{Y}^T\right)$. The empirical distribution of those draws $\{\gamma_i\}_{i=1}^{M}$ converges to the true posterior distribution $\pi\left(\gamma|\mathcal{Y}^T\right)$. Thus, any moments of interest of the posterior can be computed, as well as the marginal likelihood of the model. The algorithm 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 approximate $L\left(\mathcal{Y}^T;\gamma_i\right)$ with (6). Set $i \rightsquigarrow i+1$.

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

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 $L\left(\mathcal{Y}^T;\gamma_i^*\right)$ using (6).

Step 4, Accept/Reject: Draw $\chi_i \sim U(0,1)$. If $\chi_i \leq \frac{L\left(\mathcal{Y}^T;\gamma_i^*\right)\pi\left(\gamma_i^*\right)}{L\left(\mathcal{Y}^T;\gamma_{i-1}\right)\pi\left(\gamma_{i-1}\right)}$ 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.

2.4. Comparison with Alternative Schemes

The particle filter is not the only procedure to evaluate the likelihood of the data implied by nonlinear and/or non-normal dynamic macroeconomic models. Our previous discussion highlighted how computing the likelihood amounts to solving a nonlinear filtering problem, i.e., generating estimates of the values of W_1^t and S_0 conditional on $\mathcal{Y}^{t-1}$ to evaluate the integral in (4). 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 inaccuracy incurred by the fact that the posterior estimates of the states are non-normal. As the sample size grows, those problems accumulate and the filter diverges. Even refinements such as the Iterated Extended Kalman filter, the quadratic Kalman filter (which carries the second order term of the transition and measurement equations), and the unscented Kalman filter (which considers a set of points instead of just the conditional mean of the state, see Julier and Uhlmann, 1996) cannot fully solve these problems.

A second approach in deterministic filtering is the Gaussian-sum filter (Alspach and Sorenson, 1972), which approximates the densities required to compute the likelihood with a mixture of normals. Under regularity conditions, as the number of normals increases, we will represent 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 rely on the Extended Kalman filter to track the evolution of those different components.

A third alternative in deterministic filtering is grid filters, which use quadrature integration to compute the different integrals of the problem (Bucy and Senne, 1971). Unfortunately, grid filters are difficult to implement, since they require a constant readjustment to small changes in the model or its parameter values. Also, they are too computationally expensive to be of any practical benefit beyond very low dimensions. A final shortcoming of grid filters is that the grid points are fixed ex ante and the results are very dependent on that choice. In comparison, a particle filter can be interpreted as a grid filter where the grid points are chosen endogenously over time based on their ability to account for the data.

Tanizaki (1996) investigates the performance of deterministic filters (Extended Kalman filter, Gaussian Sum approximations, and grid filters). His Monte Carlo evidence documents that all those filters deliver poor performance in economic applications.

A second strategy is to think of the functions f and g as a change in 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, however, this approach is cumbersome and problematic to implement.

Monte Carlo techniques are a third line of research on filtering. The use of simulation techniques for non-linear filtering can be traced back at least to Handschin and Mayne (1969). Beyond the class of particle filters reviewed by Doucet, de Freitas, and Gordon (2001), other simulation techniques are as follows. Keane (1994) develops a recursive importance sampling simulator to estimate multinomial probit models with panel data. However, it is difficult to extend his algorithm to models with continuous observables. Mariano and Tanizaki (1995) propose rejection sampling. This method depends on finding an appropriate density for the rejection test. This search is time-consuming and requires substantial work for each particular model. Geweke and Tanizaki (1999) evaluate the whole joint likelihood through draws from the distribution of the whole set of states over the sample with an McMc algorithm. This approach increases notably the dimensionality of the problem, especially for the sample size used in macroeconomics. Consequently, the resulting McMc may be too slowly mixing to achieve convergence in a reasonable timeframe. Also, it requires good proposal densities and a good initialization of the chain that may be difficult to construct.

In a separate paper by the authors, Fernández-Villaverde and Rubio-Ramírez (2006) compare many of the previous approaches to filtering in a Monte Carlo experiment. We show how the particle filter outperforms the alternative filters in terms of approximating the distribution of states and minimizing the root mean squared error between the computed and exact states. We direct the interested reader to that paper for further information.

2.5. Smoothing

The particle filter allows us to draw from the filtering distribution $p\left(W_1^t, S_0|\mathcal{Y}^{t-1};\gamma\right)$ and compute the likelihood $p\left(\mathcal{Y}^T;\gamma\right)$. Often, we are also interested in the density $p\left(S^T|\mathcal{Y}^T;\gamma\right)$, i.e., the density of states conditional on the whole set of observations. Among other things, these smoothed estimates are convenient for assessing the fit of the model and running counterfactuals. We describe how to use the distribution $p\left(S^T|\mathcal{Y}^T;\gamma\right)$ for these two tasks.

First, we analyze how to assess the fit of the model. Given a value for γ, the sequence of observables implied by the model is a random variable that depends on the history of states and the history of the exogenous perturbations that affect the observables but not the states, $\mathbb{Y}^T\left(S^T, V^T;\gamma\right)$. Thus, for any γ, we compute the mean of the observables implied by the model and the realization of observables, $\mathcal{Y}^T$:

$$\overline{\mathbb{Y}}^T\left(V^T;\gamma\right) = \int \mathbb{Y}^T\left(S^T, V^T;\gamma\right) p\left(S^T|\mathcal{Y}^T;\gamma\right) dS^T. \tag{7}$$

If V^T are measurement errors, comparing $\overline{\mathbb{Y}}^T\left(V^T=0;\gamma\right)$ versus $\mathcal{Y}^T$ is a good measure of the fit of the model.

Second, we study how to run a counterfactual. Given a value for γ, what would have been the expected value of the observables if a particular state had been fixed at value from a given moment in time? We answer that question by computing:

$$\overline{\mathbb{Y}}^T_{S_k^{t:T}=S_{k,t}}\left(V^T;\gamma\right) = \int \mathbb{Y}^T\left(S^T_{-k}, S_k^{t:T}=S_{k,t}, V^T;\gamma\right) p\left(S^T|\mathcal{Y}^T;\gamma\right) dS^T, \tag{8}$$

where $S_{-k,t} = \left(S_{1,t},\ldots,S_{k-1,t},S_{k+1,t}\ldots,S_{\dim(S_t),t}\right)$ and $S_{-k}^{t:T} = \{S_{-k,m}\}_{m=t}^T$. If V^T are measurement errors, $\overline{\mathbb{Y}}^T_{S_k^{t:T}=S_{k,t}}\left(V^T=0;\gamma\right)$ represents the expected value for the whole history of observables when the state k is fixed to its value at t from that moment onward. A counterfactual exercise compares $\overline{\mathbb{Y}}^T\left(V^T=0;\gamma\right)$ and $\overline{\mathbb{Y}}^T_{S_k^{t:T}=S_{k,t}}\left(V^T=0;\gamma\right)$ for different values of k and t.

The two examples share a common theme. To compute integrals like (7) or (8), which will appear in our application below, we need be able to draw from $p\left(S^T|\mathcal{Y}^T;\gamma\right)$. To see this, let $\{s^{t,i}\}_{i=1}^{N}$ be a draw from $p\left(S^T|\mathcal{Y}^T;\gamma\right)$. Then (7) and (8) are approximated by:

$$\overline{\mathbb{Y}}^T\left(V^T=0;\gamma\right)\simeq\frac{1}{N}\sum_{i=1}^{N}\mathbb{Y}^T\left(s^{t,i},0;\gamma\right)$$

and

$$\overline{\mathbb{Y}}^T_{S_k^{t:T}=S_{k,t}}\left(V^T=0;\gamma\right)\simeq\frac{1}{N}\sum_{i=1}^{N}\mathbb{Y}^T\left(s_{-k}^{t,i},S_k^{t:T}=s_{k,t}^{i},0;\gamma\right).$$

Hence, the problem of computing integrals like (7) and (8) is equivalent to the problem of drawing from $p\left(S^T|\mathcal{Y}^T;\gamma\right)$. We now propose a smoothing algorithm to accomplish this objective.

An advantage of particle filtering is that smoothing can be implemented with the simulated filtered distribution from our previous exercise. We do so following the suggestion of Godsill, Doucet, and West (2004). We factorize the density $p\left(S^T|\mathcal{Y}^T;\gamma\right)$ as:

$$p\left(S^{t:T}|\mathcal{Y}^T;\gamma\right)=p\left(S_t|S^{t+1:T},\mathcal{Y}^T;\gamma\right)p\left(S^{t+1:T}|\mathcal{Y}^T;\gamma\right)\tag{9}$$

for all t, where $S^{t+1:T}=\{S_m\}_{m=t+1}^{T}$ is the sequence of states from period $t+1$ to period T. Therefore, from (9) it should be clear that to draw from $p\left(S^{t:T}|\mathcal{Y}^T;\gamma\right)$, we need to draw from $p\left(S_t|s^{t+1:T},\mathcal{Y}^T;\gamma\right)$, where $s^{t+1:T}\left(=\{s_m\}_{m=t+1}^{T}\right)$ is a draw from $p\left(S^{t+1:T}|\mathcal{Y}^T;\gamma\right)$. We describe a recursive procedure to do so.

Because of the Markovian structure of the shocks, we can derive

$$\begin{aligned}
p\left(S_t|s^{t+1:T},\mathcal{Y}^T;\gamma\right) &= p\left(S_t|s_{t+1},\mathcal{Y}^t;\gamma\right)\\
&= \frac{p\left(S_t|\mathcal{Y}^t;\gamma\right)p\left(s_{t+1}|S_t,\mathcal{Y}^t;\gamma\right)}{p\left(s_{t+1}|\mathcal{Y}^t;\gamma\right)}\\
&\propto p\left(S_t|\mathcal{Y}^t;\gamma\right)p\left(s_{t+1}|S_t,\mathcal{Y}^t;\gamma\right)\\
&= p\left(S_t|\mathcal{Y}^t;\gamma\right)p\left(s_{t+1}|S_t;\gamma\right)
\end{aligned}$$

Then, following an argument similar to the one in proposition 4, we show that $p\left(S_t|\mathcal{Y}^t;\gamma\right)$ is an importance sampling function to draw from the density $p\left(S_t|s^{t+1:T},\mathcal{Y}^T;\gamma\right)$. This statement is proved in the following proposition:

Proposition 6. *Let $s^{t+1:T}$ be a draw from $p\left(S^{t+1:T}|\mathcal{Y}^T;\gamma\right)$ and let $\left\{s_t^{t,i}\right\}_{i=1}^{N}$ be a draw from $p\left(S_t|\mathcal{Y}^t;\gamma\right)$. Also, let the weights:*

$$q_{t|t+1}^{i}\left(s_{t+1}\right)=\frac{p\left(s_{t+1}|s_t^{t,i};\gamma\right)}{\sum_{i=1}^{N}p\left(s_{t+1}|s_t^{t,i};\gamma\right)}.$$

Let the sequence $\left\{\widetilde{s}_t^{t,i}\right\}_{i=1}^N$ *be a draw with replacement from* $\left\{s_t^{t,i}\right\}_{i=1}^N$ *where* $q_{t|t+1}^i\left(s_{t+1}\right)$ *is the probability of* $\left\{s_t^{t,i}\right\}_{i=1}^N$ *being drawn* $\forall i$ *. Then* $\left\{\widetilde{s}_t^{t,i}\right\}_{i=1}^N$ *is a draw from* $p\left(S_t|s^{t+1:T},\mathcal{Y}^T;\gamma\right)$.

Proof. The proof uses the same strategy as the proof for proposition 4. ■

The crucial step in proposition 6 is to draw from $p\left(S_t|\mathcal{Y}^t;\gamma\right)$. We accomplish this with the output from the particle filter described in section 2.2. First, we know that $p\left(S_t|\mathcal{Y}^t;\gamma\right)=p\left(W_2^t,W_1^t,S_0|\mathcal{Y}^t;\gamma\right)$. Second, we also know that

$$p\left(W_2^t,W_1^t,S_0|\mathcal{Y}^t;\gamma\right)=p\left(W_2^t|W_1^t,S_0,\mathcal{Y}^t;\gamma\right)p\left(W_1^t,S_0|\mathcal{Y}^t;\gamma\right)$$

and $p\left(W_2^t|W_1^t,S_0,\mathcal{Y}^t;\gamma\right)=\chi_{\{w_2^t(W_1^t,S_0,\mathcal{Y}^t;\gamma)\}}\left(W_2^t\right)$, where $w_2^t\left(W_1^t,S_0,\mathcal{Y}^t;\gamma\right)$ is the function described in assumption 2 and χ is the indicator function. Thus, if $\left\{w_1^{t,i},s_0^{t,i}\right\}_{i=1}^N$ is a draw from $p\left(W_1^t,S_0|\mathcal{Y}^t;\gamma\right)$ obtained using the particle filter, then $\left\{w_2^t\left(w_1^{t,i},s_0^{t,i},\mathcal{Y}^t;\gamma\right),w_1^{t,i},s_0^{t,i}\right\}_{i=1}^N$ is a draw from $p\left(W_2^t,W_1^t,S_0|\mathcal{Y}^t;\gamma\right)$. Clearly, using $\left\{w_2^t\left(w_1^{t,i},s_0^{t,i},\mathcal{Y}^t;\gamma\right),w_1^{t,i},s_0^{t,i}\right\}_{i=1}^N$, we can build $\left\{s^{T,i}\right\}_{i=1}^N$, a draw from $p\left(S_t|\mathcal{Y}^t;\gamma\right)$.

From proposition 6 and the explanation above, the smoother algorithm is:

Step 0, Initialization: `Draw` N `particles` $\left\{w_1^{T,i},s_0^{T,i}\right\}_{i=1}^N$ `from` $p\left(W_1^{T,i},S_0|\mathcal{Y}^T;\gamma\right)$ `using a particle filter. Let` $w_2^{T,i}=w_2^T\left(w_1^{T,i},s_0^{T,i},\mathcal{Y}^t;\gamma\right)$ `and use` $\left\{w_2^{T,i},w_1^{T,i},s_0^{T,i}\right\}_{i=1}^N$ `to build` $\left\{s^{T,i}\right\}_{i=1}^N$.

Step 1, Proposal I: `Set` $i=1$.

Step 2, Proposal II: `Draw states` $\left\{s_{T-1,i}^{T-1,j}\right\}_{j=1}^M$ `from` $p\left(S_{T-1}|\mathcal{Y}^{T-1};\gamma\right)$ `and compute` $\left\{q_{T-1|T}^j\left(s_T^i\right)\right\}_{j=1}^M$.

Step 3, Resampling: `Sample once from` $\left\{s_{T-1,i}^{T-1,j}\right\}_{j=1}^M$ `with probabilities` $q_{T-1|T}^i\left(s_T^i\right)$. `Call the draw` $s^{T-1,i}$. `If` $i<N$ `set` $i\rightsquigarrow i+1$ `and go to step 2. If` $T>1$ `set` $T\rightsquigarrow T-1$ `and go to step 1. Otherwise stop.`

The algorithm works as follows. Starting from $\left\{w_1^{T,i},s_0^{T,i}\right\}_{i=1}^N$ from $p\left(W_1^{T,i},S_0|\mathcal{Y}^T;\gamma\right)$ and taking advantage of corollary 6, steps 1 to 3 generate draws $p\left(S^{t:T}|\mathcal{Y}^T;\gamma\right)$ in a recursive way. The outcome of the algorithm, $\left\{s^{T,i}\right\}_{i=1}^N$, is a draw from $p\left(S^T|\mathcal{Y}^T;\gamma\right)$. As the number of particles goes to infinity, the simulated conditional distribution of states converges to the unknown true conditional density.

3. An Application: A Business Cycle Model

In this section, we present an application of particle filtering. We estimate a business cycle model with investment-specific technological change, preference shocks, and stochastic volatility. Several reasons justify our choice. First, the business cycle model is a canonical example of a dynamic macroeconomic model. Hence, our choice demonstrates how to apply the procedure to many popular economies. Second, the model is relatively simple, a fact that facilitates the illustration of the different parts of our procedure. Third, the presence of stochastic volatility helps us to contribute to one important current discussion: the study of changes in the volatility of aggregate time series.

Even if the first work on time-varying volatility is Engle (1982), who picked as his application of the ARCH model the process for United Kingdom inflation, it is not until recently that research on volatility has acquired a crucial importance in macroeconomics. Kim and Nelson (1999) have used a Markov-Switching model to document a decline in the variance of shocks to output growth and a narrowing gap between growth rates during booms and recessions. They find that the posterior mode of the break is the first quarter of 1984. A similar result appears in McConnell and Pérez-Quirós (2000), who run a battery of different structural tests to characterize the size and timing of the reduction in output volatility. This evidence, reviewed and reinforced by Stock and Watson (2002), begets the question of what has caused the change in volatility.

One possible explanation is that the shocks hitting the economy have been very different in the 1990s than in the 1970s (Primiceri, 2005, and Sims and Zha, 2005). However, this explanation has faced the problem of how to document that, in fact, the structural shocks are now less volatile than in the past. The main obstacle has been the difficulty in evaluating the likelihood of a dynamic equilibrium model with changing volatility. Consequently, the above cited papers have estimated Structural Vector Autoregressions (SVARs). Despite their flexibility, SVARs may, though, uncover evidence that is difficult to interpret from the perspective of a dynamic equilibrium model (Fernández-Villaverde, Rubio-Ramírez, and Sargent, 2005).

The particle filter is perfectly suited for the analyzing dynamic equilibrium models with stochastic volatility. In comparison, the Kalman filter and linearization are totally useless. First, the presence of stochastic volatility induces fat tails on the distribution of observed variables. Fat tails preclude, by themselves, the application of the Kalman filter. Second, the law of motion for the states of the economy is inherently nonlinear. A linearization will drop the volatility terms, and hence, it will prevent the study of time-varying volatility.

We search for evidence of stochastic volatility on technology and on preference shocks. Loosely speaking, the preference shocks can be interpreted as proxying for demand shocks such as changes to monetary and fiscal policy that we do not model explicitly. The technology shocks can be interpreted as supply shocks. However, we are cautious regarding these interpretations, and we appreciate the need for more detailed business cycle models with time-varying volatility.

Our concurrent research applies particle filtering to more general models, and our application should be assessed as an example of the type of exercises that can be undertaken. In related work (Fernández-Villaverde and Rubio-Ramírez, 2005), we estimate the neoclassical growth model with the particle filter and the Kalman filter on a linearized version of the model. We document surprisingly big differences on the parameter estimates, on the level of the likelihood, and on the moments implied by the model.

Also, after the first version of this paper was circulated, several authors have shown the flexibility and good performance of particle filtering for the estimation of dynamic macroeconomic models. An (2006) investigates New Keynesian models. He finds that particle filtering allows him to identify more structural parameters, to fit the data better, and to obtain more accurate estimates of the welfare effects of monetary policies. King (2006) estimates the neoclassical growth model with time-varying parameters. He finds that this version of the model improves the fit to the data. Winschel (2005) applies the Smolyak operator to accelerate the numerical performance of the algorithm. These papers increase our confidence in the value of particle filtering for macroeconomists.

We divide the rest of this section into three parts. First, we present our model. Second, we describe how we solve the model numerically. Third, we explain how to evaluate the likelihood function.

3.1. The Model

We work with a business cycle model with investment-specific technological change, preference shocks, and stochastic volatility. Greenwood, Herkowitz, and Krusell (1997 and 2000) have vigorously defended the importance of technological change specific to new investment goods for understanding postwar U.S. growth and aggregate fluctuations. We follow their lead and estimate a version of their model inspired by Fisher (2004) and modified to include two unit roots, cointegration relations derived from the balanced growth path properties of the model, a preference shock, and stochastic volatility on the economic shocks that drive the dynamics of the model.

There is a representative household in the economy, whose preferences over stochastic sequences of consumption C_t and leisure $1 - L_t$ are representable by the utility function

$$E_0 \sum_{t=0}^{\infty} \beta^t \left(e^{d_t} \log C_t + \psi \log (1 - L_t) \right)$$

where $\beta \in (0, 1)$ is the discount factor, d_t is a preference shock, with law of motion

$$d_t = \rho d_{t-1} + \sigma_{dt} \varepsilon_{dt}, \text{ where } \varepsilon_{dt} \sim \mathcal{N}(0, 1),$$

ψ controls labor supply, and E_0 is the conditional expectation operator. We explain below the law of motion for σ_{dt}.

The economy produces one final good with a Cobb-Douglas production function given by:

$$C_t + X_t = A_t K_t^{\alpha} L_t^{1-\alpha},$$

and the law of motion for capital is:

$$K_{t+1} = (1 - \delta) K_t + V_t X_t,$$

where the technologies evolve as a random walk with drifts:

$$\log A_t = \gamma + \log A_{t-1} + \sigma_{at} \varepsilon_{at}, \ \gamma \geq 0 \text{ and } \varepsilon_{at} \sim \mathcal{N}(0, \sigma_a) \tag{10}$$

$$\log V_t = \upsilon + \log V_{t-1} + \sigma_{\upsilon t} \varepsilon_{\upsilon t}, \ \upsilon \geq 0 \text{ and } \varepsilon_{\upsilon t} \sim \mathcal{N}(0, \sigma_\upsilon) \tag{11}$$

Note that we have two unit roots, one in each of the two technological processes.

The process for the volatility of the shocks is given by (see Shepard, 2005, for a review of the different forms of stochastic volatility in the literature):

$$\log \sigma_{at} = (1 - \lambda_a) \log \overline{\sigma}_a + \lambda_a \log \sigma_{at-1} + \tau_a \eta_{at} \text{ and } \eta_{at} \sim \mathcal{N}(0, 1) \tag{12}$$

$$\log \sigma_{\upsilon t} = (1 - \lambda_\upsilon) \log \overline{\sigma}_\upsilon + \lambda_\upsilon \log \sigma_{\upsilon t-1} + \tau_\upsilon \eta_{\upsilon t} \text{ and } \eta_{\upsilon t} \sim \mathcal{N}(0, 1) \tag{13}$$

$$\log \sigma_{dt} = (1 - \lambda_d) \log \overline{\sigma}_d + \lambda_d \log \sigma_{dt-1} + \tau_d \eta_{dt} \text{ and } \eta_{dt} \sim \mathcal{N}(0, 1) \tag{14}$$

Thus, the matrix of unconditional variances-covariances Ω of the shocks is a diagonal matrix with entries $\{\log \overline{\sigma}_a, \log \overline{\sigma}_\upsilon, \log \overline{\sigma}_d, \tau_a, \tau_\upsilon, \tau_d\}$.

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 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 for shocks, and some initial conditions for capital and technology.

Since the presence of two unit roots makes the model non-stationary, we rescale the variables by $\widetilde{Y}_t = \frac{Y_t}{Z_t}$, $\widetilde{C}_t = \frac{C_t}{Z_t}$, $\widetilde{X}_t = \frac{X_t}{Z_t}$, and $\widetilde{K}_t = \frac{K_t}{Z_t V_{t-1}}$ where $Z_t = A_{t-1}^{\frac{1}{1-\alpha}} V_{t-1}^{\frac{\alpha}{1-\alpha}}$. Then, dividing the resource constraint and the law of motion for capital, we get:

$$\widetilde{C}_t + \widetilde{X}_t = e^{\gamma+\sigma_{at}\varepsilon_{at}} \widetilde{K}_t^{\alpha} L_t^{1-\alpha}$$
$$e^{\frac{\gamma+\alpha\upsilon+\sigma_{at}\varepsilon_{at}+\alpha\sigma_{\upsilon t}\varepsilon_{\upsilon t}}{1-\alpha}} \widetilde{K}_{t+1} = (1-\delta)\, e^{-\upsilon-\sigma_{\upsilon t}\varepsilon_{\upsilon t}} \widetilde{K}_t + \widetilde{X}_t$$

or, summing both expressions:

$$\widetilde{C}_t + e^{\frac{\gamma+\alpha\upsilon+\sigma_{at}\varepsilon_{at}+\alpha\sigma_{\upsilon t}\varepsilon_{\upsilon t}}{1-\alpha}} \widetilde{K}_{t+1} = e^{\gamma+\sigma_{at}\varepsilon_{at}} \widetilde{K}_t^{\alpha} L_t^{1-\alpha} + (1-\delta)\, e^{-\upsilon-\sigma_{\upsilon t}\varepsilon_{\upsilon t}} \widetilde{K}_t$$

Consequently, we rewrite the utility function as

$$E_0 \sum_{t=0}^{\infty} \beta^t \left(e^{d_t} \log \widetilde{C}_t + \psi \log(1-L_t) \right).$$

The intuition for these two expressions is as follows. In the resource constraint, we need to modify the term associated with $\widetilde{K}_{t+1}$ to compensate for the fact that the value of the transformed capital goes down when technology improves. A similar argument holds for the term in front of the undepreciated capital. In the utility function, we just exploit its additive log form to write it in terms of $\widetilde{C}_t$.

Then, the first order conditions for the transformed problem include an Euler equation:

$$\frac{e^{\sigma_{dt}d_t} e^{\frac{\gamma+\alpha\upsilon+\sigma_{at}\varepsilon_{at}+\alpha\sigma_{\upsilon t}\varepsilon_{\upsilon t}}{1-\alpha}}}{\widetilde{C}_t} = \beta E_t \frac{e^{\sigma_{dt+1}d_{t+1}}}{\widetilde{C}_{t+1}} \left(\alpha e^{\gamma+\sigma_{at+1}\varepsilon_{at+1}} \widetilde{K}_{t+1}^{\alpha} L_{t+1}^{1-\alpha} + (1-\delta)\, e^{-\upsilon-\sigma_{\upsilon t+1}\varepsilon_{\upsilon t+1}} \right) \tag{15}$$

and a labor supply condition:

$$\psi \frac{e^{\sigma_{dt}d_t} \widetilde{C}_t}{1-L_t} = (1-\alpha)\, e^{\gamma+\sigma_{at}\varepsilon_{at}} \widetilde{K}_t^{\alpha} L_t^{-\alpha}, \tag{16}$$

together with the resource constraint:

$$\widetilde{C}_t + e^{\frac{\gamma+\alpha\upsilon+\sigma_{at}\varepsilon_{at}+\alpha\sigma_{\upsilon t}\varepsilon_{\upsilon t}}{1-\alpha}} \widetilde{K}_{t+1} = e^{\gamma+\sigma_{at}\varepsilon_{at}} \widetilde{K}_t^{\alpha} L_t^{1-\alpha} + (1-\delta)\, e^{-\upsilon-\sigma_{\upsilon t}\varepsilon_{\upsilon t}} \widetilde{K}_t. \tag{17}$$

These equations imply a deterministic steady state around which we will approximate the solution of our model. In that steady state, the investment to output ratio:

$$\frac{\widetilde{X}_{ss}}{\widetilde{Y}_{ss}} = \alpha \frac{\left(e^{\frac{\gamma+\alpha\upsilon}{1-\alpha}} - (1-\delta)\, e^{-\upsilon}\right)}{\frac{\exp\left(\frac{\gamma+\alpha\upsilon}{1-\alpha}\right)}{\beta} - (1-\delta)\exp(-\upsilon)}$$

is a constant. This expression shows how investment and output are cointegrated either in nominal or in real terms (when we properly deflate investment by the consumption price index). The approximated solution of our model will respect this cointegration relation, and hence, it will show up (implicitly) in our estimation.

3.2. Solving the Model

To solve the model, we find the policy functions for hours worked, L_t, rescaled investment, $\widetilde{X}_t$, and rescaled capital $\widetilde{K}_{t+1}$, such that the system formed by (15)-(17) holds. This system of equilibrium equations does not have a known analytical solution, and we solve it with a numerical method. We select a second order perturbation to do so. Aruoba, Fernández-Villaverde, and Rubio-Ramírez (2006) document that perturbation methods deliver a highly accurate and fast solution in a model similar to the one considered here. We emphasize, however, that nothing in the particle filter stops us from opting for any other nonlinear solution method as projection methods or value function iteration. The appropriate choice of solution method should be dictated by the details of the particular model to be estimated.

As a first step, we parameterize the matrix of variances-covariances of the shocks as $\overline{\Omega}(\chi) = \chi\Omega$, where clearly $\overline{\Omega}(1) = \Omega$. Then, we take a perturbation solution around $\chi = 0$, i.e., around the deterministic steady state implied by the equilibrium conditions of the model.

The states of the model are given by:

$$\widetilde{s}_t = \left(1, \log \widetilde{K}_t, \sigma_{at}\varepsilon_{at}, \sigma_{\upsilon t}\varepsilon_{\upsilon t}, d_t, \log \sigma_{at}, \log \sigma_{\upsilon t}, \log \sigma_{dt}\right)'$$

The volatilities of the shocks are state variables of the model and the households keep track of them when making optimal decisions. Thus, a second order approximation to the policy function for capital is given by:

$$\log \widetilde{K}_{t+1} = \Psi_{k1}\widetilde{s}_t + \frac{1}{2}\widetilde{s}_t'\Psi_{k2}\widetilde{s}_t \tag{18}$$

Note that Ψ_{k1} is a 1×8 vector and Ψ_{k2} is a 8×8 matrix. However, Ψ_{k2} only has 36 distinct elements because it is symmetric. The term $\Psi_{k1}\widetilde{s}_t$ constitutes the linear solution of the model, except for a constant added by the second order approximation that corrects for precautionary behavior. Similarly, the policy functions for rescaled investment and labor are given by:

$$\begin{aligned}
\log \widetilde{X}_t &= \Psi_{x1}\widetilde{s}_t + \frac{1}{2}\widetilde{s}_t'\Psi_{x2}\widetilde{s}_t \\
\log L_t &= \Psi_{l1}\widetilde{s}_t + \frac{1}{2}\widetilde{s}_t'\Psi_{l2}\widetilde{s}_t
\end{aligned}$$

The policy for rescaled output is obtained by noting that since

$$\log \widetilde{Y}_t = \gamma + \sigma_{at}\varepsilon_{at} + \alpha \log \widetilde{K}_t + (1-\alpha)\log L_t$$

we have that

$$\begin{aligned}
\log \widetilde{Y}_t &= \gamma + \sigma_{at}\varepsilon_{at} + \alpha \log \widetilde{K}_t + (1-\alpha)\,\Psi_{l1}\widetilde{s}_t + (1-\alpha)\,\frac{1}{2}\widetilde{s}_t'\Psi_{l2}\widetilde{s}_t \\
&= ((\gamma, \alpha, 1, 0, 0, 0, 0, 0) + (1-\alpha)\,\Psi_{l1})\,\widetilde{s}_t + \frac{1}{2}\widetilde{s}_t'\,(1-\alpha)\,\Psi_{l2}\widetilde{s}_t \\
&= \Psi_{y1}\widetilde{s}_t + \frac{1}{2}\widetilde{s}_t'\Psi_{y2}\widetilde{s}_t
\end{aligned}$$

where $\Psi_{y1} = (\gamma, \alpha, 1, 0, 0, 0, 0, 0) + (1-\alpha)\,\Psi_{l1}$ and $\Psi_{y2} = (1-\alpha)\,\Psi_{l2}$.

4. A State Space Representation

Now we present a state representation for the variables. We discuss first the transition equation and later the measurement equation.

4.1. The Transition Equation

Given our model, we have a vector of structural parameters

$$\gamma \equiv (\alpha, \delta, \rho, \beta, \psi, \upsilon, \zeta, \tau_a, \tau_\upsilon, \tau_d, \overline{\sigma}_a, \overline{\sigma}_\upsilon, \overline{\sigma}_d, \lambda_a, \lambda_\upsilon, \lambda_d, \sigma_{1\epsilon}, \sigma_{2\epsilon}, \sigma_{3\epsilon}) \in \Upsilon \subset R^{19}$$

where $\sigma_{1\epsilon}, \sigma_{2\epsilon}$, and $\sigma_{3\epsilon}$ are the standard deviation of three measurement errors to be introduced below.

We combine the laws of motion for the volatility (12)-(14) and the policy function of capital (18) to build:

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

where $S_{t-1} = (s_{t-1}, s_{t-2})$,

$$s_{t-1} = \left(1, \log \widetilde{K}_t, \sigma_{at-1}\varepsilon_{at-1}, \sigma_{\upsilon t-1}\varepsilon_{\upsilon t-1}, d_{t-1}, \log \sigma_{at-1}, \log \sigma_{\upsilon t-1}, \log \sigma_{dt-1}\right)',$$

and $W_t = (\varepsilon_{at}, \varepsilon_{\upsilon t}, \varepsilon_{dt}, \eta_{at}, \eta_{\upsilon t}, \eta_{dt})'$. We keep track of the past states, s_{t-2}, because some of the observables in the measurement equation below will appear in first differences. If we denote by $f_i(S_{t-1}, W_t)$ the $i-th$ dimension of f, we have

$$
\begin{aligned}
f_1(S_{t-1}, W_t; \gamma) &= 1 \\
f_2(S_{t-1}, W_t; \gamma) &= \Psi_{k1}\widetilde{s}_t + \frac{1}{2}\widetilde{s}_t'\Psi_{k2}\widetilde{s}_t \\
f_3(S_{t-1}, W_t; \gamma) &= e^{(1-\lambda_a)\log\overline{\sigma}_a + \lambda_a \log\sigma_{at-1} + \tau_a\eta_{at}}\varepsilon_{at} \\
f_4(S_{t-1}, W_t; \gamma) &= e^{(1-\lambda_\upsilon)\log\overline{\sigma}_\upsilon + \lambda_\upsilon \log\sigma_{\upsilon t-1} + \tau_\upsilon\eta_{\upsilon t}}\varepsilon_{\upsilon t} \\
f_5(S_{t-1}, W_t; \gamma) &= \rho d_{t-1} + e^{(1-\lambda_d)\log\overline{\sigma}_d + \lambda_d \log\sigma_{dt-1} + \tau_d\eta_{dt}}\varepsilon_{dt} \\
f_6(S_{t-1}, W_t; \gamma) &= (1-\lambda_a)\log\overline{\sigma}_a + \lambda_a\log\sigma_{at-1} + \tau_a\eta_{at} \\
f_7(S_{t-1}, W_t; \gamma) &= (1-\lambda_\upsilon)\log\overline{\sigma}_\upsilon + \lambda_\upsilon\log\sigma_{\upsilon t-1} + \tau_\upsilon\eta_{\upsilon t} \\
f_8(S_{t-1}, W_t; \gamma) &= (1-\lambda_d)\log\overline{\sigma}_d + \lambda_d\log\sigma_{dt-1} + \tau_d\eta_{dt} \\
f_{9-16}(S_{t-1}, W_t; \gamma) &= s_t
\end{aligned}
$$

where $\widetilde{s}_t = \widetilde{s}(S_{t-1}, W_t; \gamma)$ is a function of S_{t-1} and W_t.

4.2. The Measurement Equation

We assume that we observe the following time series $\mathcal{Y}_t = (\Delta \log p_t, \Delta \log y_t, \Delta \log x_t, \log l_t)$, the change in the relative price of investment, the observed real output per capita growth, the observed real gross investment per capita growth, and observed hours worked per capita. We make this assumption out of pure convenience. On one hand, we want to capture some of the main empirical predictions of the model. On the other hand, and for illustration purposes, we want to keep the dimensionality of the problem low. However, the empirical analysis could be performed with very different combinations of data. Thus, our choice should be understood as an example of how to estimate the likelihood function associated with a vector of observations.

In equilibrium the change in that relative price of investment equals the negative log difference of V_t:

$$-\Delta \log V_t = -\upsilon - \sigma_{\upsilon t}\varepsilon_{\upsilon t}$$

This allows us to read $\sigma_{\upsilon t}\varepsilon_{\upsilon t}$ directly from the data conditional on an estimate of υ.

To build the measurement equation for real output per capita growth, we remember that from (10) and (11), we have :

$$\Delta \log A_t = \gamma + \sigma_{at}\varepsilon_{at}, \ \gamma \geq 0 \text{ and } \varepsilon_{at} \sim \mathcal{N}(0, \sigma_a)$$

$$\Delta \log V_t = \upsilon + \sigma_{\upsilon t}\varepsilon_{\upsilon t},\ \upsilon \geq 0 \text{ and } \varepsilon_{\upsilon t} \sim \mathcal{N}(0, \sigma_\upsilon)$$

Also

$$\widetilde{Y}_t = \frac{Y_t}{Z_t} = \frac{Y_t}{A_{t-1}^{\frac{1}{1-\alpha}} V_{t-1}^{\frac{\alpha}{1-\alpha}}},$$

therefore:

$$\log Y_t = \log \widetilde{Y}_t + \frac{1}{1-\alpha}\log A_{t-1} + \frac{\alpha}{1-\alpha}\log V_{t-1}.$$

Then:

$$\begin{aligned}
\Delta \log Y_t &= \Delta \log \widetilde{Y}_t + \frac{1}{1-\alpha}\left(\Delta \log A_{t-1} + \alpha \Delta \log V_{t-1}\right) \\
&= \Delta \log \widetilde{Y}_t + \frac{1}{1-\alpha}\left(\gamma + \alpha\upsilon + \sigma_{at-1}\varepsilon_{at-1} + \alpha\sigma_{\upsilon t-1}\varepsilon_{\upsilon t-1}\right)
\end{aligned}$$

Since

$$\log \widetilde{Y}_t = \Psi_{y1}\widetilde{s}_t + \frac{1}{2}\widetilde{s}_t'\Psi_{y2}\widetilde{s}_t$$

we have

$$\Delta \log Y_t = \Psi_{y1}\widetilde{s}_t + \frac{1}{2}\widetilde{s}_t'\Psi_{y2}\widetilde{s}_t - \Psi_{y1}\widetilde{s}_{t-1} - \frac{1}{2}\widetilde{s}_{t-1}'\Psi_{y2}\widetilde{s}_{t-1} + \frac{1}{1-\alpha}\left(\gamma + \alpha\upsilon + \sigma_{at-1}\varepsilon_{at-1} + \alpha\sigma_{\upsilon t-1}\varepsilon_{\upsilon t-1}\right)$$

Similarly, for real gross investment per capita growth:

$$\Delta \log X_t = \Psi_{x1}\widetilde{s}_t - \Psi_{x1}\widetilde{s}_{t-1} + \frac{1}{2}\widetilde{s}_t'\Psi_{x2}\widetilde{s}_t - \frac{1}{2}\widetilde{s}_{t-1}'\Psi_{x2}\widetilde{s}_{t-1} + \frac{1}{1-\alpha}\left(\gamma + \alpha\upsilon + \sigma_{at-1}\varepsilon_{at-1} + \alpha\sigma_{\upsilon t-1}\varepsilon_{\upsilon t-1}\right)$$

Finally, we introduce measurement errors in the real output per capita growth, real gross investment per capita growth, and hours worked per capita as the easiest way to avoid stochastic singularity (see our assumptions 1 to 3). 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 and preferences, we would have shocks to depreciation and to the discount factor. This alternative might be more empirically relevant, but it would make the solution of the model much more involved. Since our goal here is to illustrate how to apply our particle filtering to estimate the likelihood of the model in a simple example, we prefer to specify measurement errors. We will have three different measurement errors: one in the real output per capita growth, ϵ_{1t}, one on the real gross investment per capita growth, ϵ_{2t}, and one on hours worked per capita, ϵ_{3t}. We do not have a measurement error in the relative price of investment because it will not be possible to separately identify it from $\sigma_{\upsilon t}\epsilon_{\upsilon t}$. The three shocks are an i.i.d. process with distribution $\mathcal{N}(0, \Sigma_\epsilon)$. The matrix of variances-covariances Σ_ϵ is a diagonal matrix with entries $\{\sigma_{1\epsilon}, \sigma_{2\epsilon}, \sigma_{3\epsilon}\}$. In our notation of section 2, $V_t = (\epsilon_{1t}, \epsilon_{2t}, \epsilon_{3t})'$. The measurement errors imply a difference between the value for the variables implied by the model and the observables. Thus:

$$
\begin{aligned}
\Delta \log p_t &= -\Delta \log V_t \\
\Delta \log y_t &= \Delta \log Y_t + \epsilon_{1t} \\
\Delta \log x_t &= \Delta \log X_t + \epsilon_{2t} \\
\log l_t &= \log L_t + \epsilon_{3t}
\end{aligned}
$$

Putting the different pieces together, we have the measurement equation:

$$
\begin{pmatrix} \Delta \log p_t \\ \Delta \log y_t \\ \Delta \log x_t \\ \log l_t \end{pmatrix} = \begin{pmatrix} -\upsilon \\ \frac{\gamma+\alpha\upsilon}{1-\alpha} \\ \frac{\gamma+\alpha\upsilon}{1-\alpha} \\ \log L_{ss} + \Psi_{l3} \end{pmatrix} +
$$

$$
\begin{pmatrix} -\sigma_{\upsilon t}\varepsilon_{\upsilon t} \\ \Psi_{y1}\widetilde{s}_t + \frac{1}{2}\widetilde{s}_t'\Psi_{y2}\widetilde{s}_t - \Psi_{y1}\widetilde{s}_{t-1} - \frac{1}{2}\widetilde{s}_{t-1}'\Psi_{y2}\widetilde{s}_{t-1} + \frac{1}{1-\alpha}\left(\sigma_{at-1}\varepsilon_{at-1} + \alpha\sigma_{\upsilon t-1}\varepsilon_{\upsilon t-1}\right) \\ \Psi_{x1}\widetilde{s}_t - \Psi_{x1}\widetilde{s}_{t-1} + \frac{1}{2}\widetilde{s}_t'\Psi_{x2}\widetilde{s}_t - \frac{1}{2}\widetilde{s}_{t-1}'\Psi_{x2}\widetilde{s}_{t-1} + \frac{1}{1-\alpha}\left(\sigma_{at-1}\varepsilon_{at-1} + \alpha\sigma_{\upsilon t-1}\varepsilon_{\upsilon t-1}\right) \\ \Psi_{l1}\widetilde{s}_t + \frac{1}{2}\widetilde{s}_t'\Psi_{l2}\widetilde{s}_t \end{pmatrix} + \begin{pmatrix} 0 \\ \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{pmatrix}
$$

5. The Likelihood Function

Given that we have six structural shocks in the model and $\dim(V_t) < \dim(\mathbb{Y}_t)$, we set $W_{2,t} = \varepsilon_{\upsilon t}$ and $W_{1,t} = (\varepsilon_{at}, \varepsilon_{dt}, \eta_{dt}, \eta_{at}, \eta_{\upsilon t})'$. Then, in the notation of section 2, the prediction errors are:

$$
\begin{aligned}
\omega_{2,t} &= \omega_{2,t}(W_1^t, S_0, \mathcal{Y}^t; \gamma) = -\frac{\Delta \log p_t + \upsilon}{\sigma_{\upsilon t}}, \\
v_{1,t} &= v_{1,t}(W_1^t, S_0, \mathcal{Y}^t; \gamma) = \Delta \log y_t - \Delta \log Y_t, \\
v_{2,t} &= v_{2,t}(W_1^t, S_0, \mathcal{Y}^t; \gamma) = \Delta \log x_t - \Delta \log X_t, \text{ and} \\
v_{3,t} &= v_{3,t}(W_1^t, S_0, \mathcal{Y}^t; \gamma) = \log l_t - \log L_t.
\end{aligned}
$$

We let $\omega_t = (\omega_{2,t}, v_{1t}, v_{2t}, v_{3t})'$ and define $\widetilde{\Sigma}_t = \widetilde{\Sigma}(W_1^t, S_0, \mathcal{Y}^{t-1}; \gamma)$ be a diagonal matrix with entries $\{\sigma_{\upsilon t}, \sigma_{1\epsilon}, \sigma_{2\epsilon}, \sigma_{3\epsilon}\}$. Hence, we have that:

$$
p\left(\mathcal{Y}_t | W_1^t, S_0, \mathcal{Y}^{t-1}; \gamma\right) = (2\pi)^{-\frac{4}{2}} \left|\widetilde{\Sigma}_t\right|^{-\frac{1}{2}} \exp^{-\frac{1}{2}\omega_t'\widetilde{\Sigma}_t\omega_t}
$$

and we rewrite (4) as:

$$L\left(\mathcal{Y}^T;\gamma\right)=(2\pi)^{-\frac{4T}{2}}\prod_{t=1}^{T}\int\int\left|\widetilde{\Sigma}_t\right|^{-\frac{1}{2}}\exp^{-\frac{1}{2}\omega_t'\widetilde{\Sigma}_t\omega_t}p\left(W_1^t,S_0|\mathcal{Y}^{t-1};\gamma\right)dW_1^t dS_0.$$

This last expression is simple to handle. Given particles $\left\{\left\{w_1^{t|t-1,i},s_0^{t|t-1,i}\right\}_{i=1}^{N}\right\}_{t=1}^{T}$, we build the prediction errors $\left\{\{\omega_t^i\}_{i=1}^N\right\}_{t=1}^T$ and the matrices $\left\{\left\{\widetilde{\Sigma}_t^i\right\}_{i=1}^N\right\}_{t=1}^T$ implied by them. Therefore, the likelihood function is approximated by:

$$L\left(\mathcal{Y}^T;\gamma\right)\simeq(2\pi)^{-\frac{4T}{2}}\prod_{t=1}^{T}\sum_{i=1}^{N}\left|\widetilde{\Sigma}_t\right|^{-\frac{1}{2}}\exp^{-\frac{1}{2}\omega_t^{i\prime}\widetilde{\Sigma}_t^i\omega_t^i}. \tag{19}$$

Equation (19) is nearly identical to the likelihood function implied by the Kalman filter (see equation 3.4.5 in Harvey, 1989) when applied to a linear model. The difference is that in the Kalman filter, the prediction errors ω_t come directly from the output of the Riccati equation, while in our filter they come from the output of the simulation.

6. Findings

In this section we conduct likelihood-based inference on our model with U.S. data. This exercise proves how the particle filter can be brought to real-life applications and how it delivers new results concerning the business cycle dynamics of the U.S. economy.

We estimate the model using the relative price of investment with respect to the price of consumption, real output per capita growth, real gross investment per capita growth, and hours worked per capita in the U.S. from 1955:Q1 to 2000:Q4. Our sample length is limited by the availability of good data on the relative price of investment that account for quality change in the ways dictated by theory (see the description in Fisher, 2004). To match our model predictions with the observed data, we need to be careful when constructing our observed series. In particular, to make the observed series compatible with the model implied series, we have to compute both real output and real gross investment in consumption units.

As the relative price of investment we use the ratio of an investment deflator and a deflator for consumption. The consumption deflator is constructed from the deflators of nondurable goods and services reported in NIPA. Since the NIPA investment deflators are poorly measured, we use the investment deflator constructed by Fisher (2004). For the real output per capita series, we first define nominal output as nominal consumption plus nominal gross investment. We define nominal consumption as the sum of personal consumption expenditures on nondurable goods and services, national defense consumption expenditures, federal nondefense consumption expenditures, and state and local government consumption expenditures. We define nominal gross investment as the sum of personal consumption expenditures on durable goods, national defense gross investment, federal government nondefense gross investment, state and local government gross investment, private nonresidential fixed investment, and private nonresidential residential fixed investment. Per capita nominal output is defined as the ratio between our nominal output series and the civilian noninstitutional population between 16 and 65. Since we need to measure real output per capita in consumption units, we deflate the series by the consumption deflator. For the real gross investment per capita series, we divide our above mentioned nominal gross investment series by the civilian noninstitutional population between 16 and 65 and the consumption deflator. Finally, the hours worked per capita series is constructed with the index of total number of hours worked in the business sector and the civilian noninstitutional population between 16 and 65. Since our model implied series for hours worked per capita is between 0 and 1, we normalize the observed series of hours worked per capita such that it is, on average, 0.33.

We perform our estimation exercises from a classical perspective and from a Bayesian one. For the classical perspective, we maximize the likelihood of the model with respect to the parameters. For the Bayesian approach, we specify prior distributions over the parameters, evaluate the likelihood using the particle filter, and draw from the posterior using a Metropolis-Hastings algorithm. The results from both approaches are very similar. In the interest of space, we report only our classical findings. Bayesian inference would be feasible (although extremely slow) using an McMc, while maximum likelihood estimation is unsolved for this problem. Our Bayesian findings are available upon request.

6.1. Point Estimates

Before taking the model to the data, we fix two parameters to improve the quality of the estimation. We set $\alpha = 0.33$ and $\delta = 0.0132$ to match capital income share and the capital-output ratio in the U.S. economy. Also, we constrain the value of β to be less than one (so the utility of the consumer is well-defined) and the autoregressive coefficients $\{\rho, \lambda_a, \lambda_\upsilon, \lambda_d\}$ to be between 0 and 1 to maintain stationarity.

Table 6.1 reports the MLE for the other 17 parameters of the model and their standard errors. The point estimates are close to the ones coming from a standard calibration exercise, suggesting a good performance of the estimation. More important, the standard errors of the estimates are low, indicating tight estimates. We interpret our finding as an endorsement of the ability of the procedure to uncover sensible values for the parameters of dynamic macroeconomic models.

[Table 6.1 Here]

The autoregressive component of the preference shock level, ρ, reveals a high persistence of this demand component. The discount factor, β, nearly equal to 1, is a common finding when estimating dynamic models. The parameter that governs labor supply, ψ, is closely estimated around 2.343 to capture the level of hours worked per capita. The two drifts in the technology processes, υ and ζ, imply an average growth rate of 0.45 percent quarter to quarter (which corresponds to 1.8 percent per year, the historical long-run growth rate of U.S. real output per capita since the Civil War), out of which 98.8 percent is accounted for by investment-specific technological change. This result highlights the importance of modelling this biased technological change for understanding growth and fluctuations. The autoregressive components of stochastic volatility of the shocks $\{\lambda_a, \lambda_\upsilon, \lambda_d\}$ are more difficult to estimate, ranging from a low number, λ_a, to a persistence close to one, λ_υ. Our results hint that modelling volatility as a random walk to economize on parameters may be misleading, since most of the mass of the likelihood is below 1.

The three estimated measurement error variances are such that the structural model accounts for the bulk of the variation in the data. A formal way to back up our statement and assess the performance of the model is to compare the average of the paths of the observed variables predicted by the smoothed states at the MLE without the measurement errors against the real data. In the language of section 2.5, we need to compare $\overline{\mathbb{Y}}^T\left(V^T = 0; \widehat{\gamma}_{MLE}\right)$ and $\mathcal{Y}^T$. In the four panels of figure 6.1, we plot the average predicted (discontinuous line) and observed (continuous line) paths of real output per capita growth, real gross investment per capita growth, hours worked per capita, and relative price of investment.

The top left panel displays how closely the model captures the dynamics of the real output per capita growth, including the recessions of the 1970s and the expansions of the 1980s and 1990s. We assess the fit of the model using three measures. First, the correlation between the model average predicted and observed real output per capita growth is 72 percent. Second, the model average predicted output accounts for 72 percent of the standard deviation of real output per capita growth. Third, the mean squared error between the model average predicted output and the data is 1.029e-5. We judge that these three measures demonstrate the model's notable performance in accounting for fluctuations in real output per capita growth.

The top right panel shows the fit between the model average predicted and the observed real gross investment per capita growth. Now, the model average predicted real gross investment per capita growth has a correlation of 79 percent with the data, it accounts for 124 percent of the observed standard deviation, and the mean squared error between the model average predicted and the data is 8.660e-5. The model, hence, also seems to do a good job with real gross investment per capita growth, except for the excessive volatility of the predicted real gross investment per capita. This failure is common in business cycle models without adjustment costs. It is in the bottom left panel, hours worked per capita, where the model shows its best: the correlation between the model average predicted and observed hours worked per capita is 99 percent, the model accounts for 98 percent of the observed standard deviation of hours worked per capita, and the mean squared error is only 4.757e-6. The bottom right panel analyzes the fit of the model with respect to the relative price of investment. Since we assume that we observe this series without measurement error, both the average predicted and the observed relative prices are the same. Our summary assessment of figure 6.1 is that the model is fairly successful at capturing aggregate dynamics.

6.2. Evolution of the Volatility

Given the success of the business cycle model at fitting the data, we have high confidence in our findings regarding the (unobserved) states of the model. In figure 6.2, we plot the mean of the smoothed path for capital in log deviations from the balanced growth, $\log \widetilde{K}_t$ (top left panel), neutral productivity shock, $\sigma_{at}\varepsilon_{at}$ (top right panel), investment specific shock, $\sigma_{\upsilon t}\varepsilon_{\upsilon t}$ (bottom left panel), and preference level, d_t (bottom right panel). We also plot the one standard deviation bands around the mean of the smoothed paths. The bands demonstrate that for all four states, our smoothed paths are tightly estimated. All the smoothed paths reported in this section are computed at the MLE.

The most interesting of the panels is the bottom right one. It shows important negative preference shocks in the late 1970s and 1980s and large positive shocks in the early 1980s and 1990s. Since the preference shock can be loosely interpreted as a demand shock, our empirical results are compatible with those accounts of fluctuations in the U.S. economy that emphasize the role of changes in demand induced by monetary policy that occurred during those years. Since our model lacks that margin, we are careful in our interpretation, and we only point out this result as suggestive for future work.

Figure 6.3 presents the mean of the smoothed paths of the volatility for the three shocks, also with the one standard deviation bands. The top left panel reveals how the volatility of the shock to neutral technology, σ_{at}, has been roughly constant over the sample. In comparison, the top right panel shows how the volatility of the shock to investment-specific technology,

σ_{vt}, has evolved. After a small decline until the mid 1960s, the volatility increases steadily until the mid 1970s. Its peak coincides with the oil shocks of late 1970s. After the peak, it falls back during the next 20 years until it stabilizes by the end of the 1990s. The bottom left panel shows a decrease in volatility of the preference shock, σ_{dt}, from 1955 to the late 1960s. That volatility increased slightly during the 1970s and 1980s and fell again in the 1990s. The evidence in figure 6.3 is consistent with the evidence from time series methods summarized in section 5.4 of Stock and Watson (2002). Figure 6.3 teaches the first important empirical lesson of this paper: there is important evidence of time-varying volatility in the aggregate shocks driving the U.S. economy.

How did the time-varying volatility of shocks affect the volatility of the aggregate time series? Figure 6.4 computes the instantaneous standard deviation of each of the four observables implied by the MLE and the smoothed volatilities. For example, each point in the top left panel represents the standard deviation of real output per capita growth if the volatility of the three shocks had stayed constant forever at the level at which we smoothed for that quarter. Figure 6.4 can be interpreted as the estimate of the realized volatility of the observables. Of course, for each quarter the smoothed volatility is not a point but a distribution. Hence, we draw from this distribution using the algorithm described in section 2.5 and compute the instantaneous standard deviation of the observables for each draw. We report the mean of the instantaneous standard deviation of the observables.

Figure 6.4 shows important reductions in the volatility of real output per capita growth and real gross investment per capita growth, a smaller reduction in the volatility of hours worked per capita, and a roughly constant volatility of the relative price of investment. The top left panel indicates that the reduction in the volatility of real output per capita growth is not the product of an abrupt change in the mid 1980s, as defended by a part of the literature, but more of a gradual change. It started in the late 1950s, was interrupted in the late 1960s and 1970s, and resumed again by the end of the 1970s, continuing until today. The presence of large preference shocks in the 1970s hid that reduction in volatility until the mid 1980s, when the literature dates the fall in real output per capita growth volatility. Moreover, the size of the reduction in instantaneous volatility of real output per capita growth is around 55 percent. This reduction is comparable with the computations of Blanchard and Simon (2001) and Stock and Watson (2002).

Our finding of a steady reduction in the volatility of real output per capita growth coincides with the view of Blanchard and Simon (2001). It is interesting to compare the top left panel in our figure 6.4 with figure 1 in their paper to see how we reach qualitatively similar conclusions (although the size of the reduction in volatility estimated by the rolling standard deviation they propose is bigger). This steady decline in volatility of real output per capita growth since the late 1950s, punctuated by an increase in the 1970s, is the second main finding of our paper.

6.3. What Caused the Fall in Volatility?

Which shocks account for the reduction in the volatility of U.S. aggregate time series? In the context of nonlinear models, it is difficult to perform a decomposition of the variance because the random shocks hitting the model enter in multiplicative ways. Instead, to asses the role of the different shocks, we perform three counterfactual experiments.

In the first one, we fix the volatility of one of the three shocks at its level in 1958 and we let the volatility of the other shocks evolve in the same way as in our smoothed estimates. In figure 6.5, we plot the estimated instantaneous variance (continuous line) and the counterfactual variance (discontinuous line) of real output per capita growth (first row), real gross investment per capita growth (second row), hours worked per capita (third row), and the relative price of investment (fourth row), when we keep it at its value in the first quarter of 1958, the volatility of the neutral technological shock (first column), the volatility of the investment-specific technological shocks (second column), and the volatility of the preference shock (third column). The results from figure 6.5 illustrate that neither changes in the volatility of the neutral technological shock nor the investment-specific technological shock explain the reduction in volatility. Indeed, the lines of the estimated actual volatility and of the counterfactual are nearly on top of each other. The decrease in the volatility of the shock to preferences explains the reduction in volatility. We see how in the counterfactual, the volatility of the aggregate time series is roughly constant at the level of the late 1950s.

The second counterfactual experiment repeats the first experiment, except that now we fix the volatilities at their values in the first quarter of 1978. We plot our findings in figure 6.6. Again, we see the same results as in the first experiment. Nearly all the reduction in volatility is accounted for by reductions in the volatility of the preference shock.

Finally, the third counterfactual experiment fixes the volatilities at the fourth quarter of 1974, the peak of the volatility of the relative price of investment. The results are once more the same as in the first two experiments.

From figures 6.5, 6.6, and 6.7, we conclude that our model points to the crucial role of the changes in the volatility of the preference shocks in accounting for the large reduction in the volatility of real output per capita growth, real gross investment per capita growth, and hours worked per capita.

7. Are Nonlinearities and Stochastic Volatility Important?

Our model has one novel feature, stochastic volatility, and one particularity in its implementation, the nonlinear solution. How important are these two elements? How much do they add to the empirical exercise?

To answer these questions, we estimated three additional versions of the model: one with a linear solution without stochastic volatility on the shocks (the volatility parameters $\{\tau_a, \tau_v, \tau_d, \lambda_a, \lambda_v, \lambda_d\}$ are all set to zero), one with a linear solution with stochastic volatility, and one with a quadratic solution but without stochastic volatility. We name those versions in table 7.1.

Table 7.1: Versions of the Model

Solution\Stochastic Volatility	No	Yes
Linear	Version 1	Version 2
Quadratic	Version 3	Benchmark

We spend some time comparing the results of version 2 of the model and the benchmark. This comparison is relevant because version 2 implies a conditionally normal linear model, which can be estimated with alternative McMc algorithms (Justiniano and Primiceri, 2005). We plot in figures 7.1 and 7.2 the evolution of the smoothed states and volatility that we obtain when we estimate version 2 of the model and the benchmark. These two figures tell us the importance of the quadratic component of the solution. In the top left panel of figure 7.1, we observe how the quadratic solution accumulates more capital (in deviations with respect to the scaled steady state) as a response to the increase in volatility than the linear solution. In the top right and bottom left panels, we appreciate important differences in the level of the neutral technological shock and the preference shock, which fluctuate much more in the linear solution. The explanation is simple. In the absence of quadratic components, the model requires bigger shocks to preferences to fit the data. Moreover, the sign of the shocks is often different. When the linear model is telling us there were large positive technological shocks (for example, in the early 1960s), the benchmark model suggests negative shocks. The investment-specific technological shock is roughly the same in both version 2 and the benchmark, since we are reading it from the relative price of new capital (since all the parameters are jointly estimated and the law of motion for investment-specific technological shock affects the policy functions of the households, we do not get exactly the same process).

Figure 7.2 shows that the level and evolution of the volatilities are also quite different in the linear and in the quadratic solution. The top left panel reports how the volatility of the neutral technological shock is failing in the linearized version of the model, while the same volatility is roughly constant for the benchmark quadratic case. The top right panel tells us how the investment-specific technological shock has a bigger variation for the quadratic case, with a later peak. Finally, the bottom left panel suggests that the volatility of the preference shock of the quadratic model is much smaller, and that it has been declining much less than the volatility of the preference shock of the linearized model.

We summarize our results from figures 7.1 and 7.2. First, the magnitude of the shocks (often even their signs) are sufficiently different in the linear and quadratic version of the model that they draw for us different pictures of the cyclical evolution of the U.S. economy. Second, volatility is lower in the benchmark case for two of the three shocks over the entire sample; the quadratic solution implies that we require smaller shocks on average to fit the data. Third, for all three shocks, the reduction in volatility is much smaller in the benchmark case than in version 2 of the model.

However, if both versions of the model deliver different answers, we need to compare the fit of the two models to assess which of the two answers we trust more. A simple way to check the fit of each version is to look at the value of the likelihood at their MLEs. The benchmark model has a loglikelihood of 2350.6, while version 2 has a loglikelihood of 2230.4, an advantage in favor of the quadratic solution of 120 log points. If we implement Rivers and Vuong's (2002) likelihood ratio test for dynamic, non-nested, misspecified models, we find that this difference significantly favors the benchmark model with a p-value of 10.3. This result is somehow strong evidence in favor of the quadratic solution, especially if we consider that the Kernel estimate of the asymptotic variance of the test uses a (conservative) 4 period window.

The good performance of the quadratic model is remarkable because it comes despite two difficulties. First, three of the series of the model enter in first differences. This reduces the advantage of the quadratic solution in comparison with the linear one, since the mean growth rate, a linear component, is well captured by both solutions. Second, the solution is only of second order and some important nonlinearities may be of higher order. Consequently, our results show that, even in those challenging circumstances, the nonlinear estimation pays off.

In the interest of space, we do not discuss the comparison of benchmark with versions 1 and 3 of the model. We just mention that the presence of stochastic volatility does not improve the fit of the linear model (comparing version 1 with version 2) but it does for the quadratic model (version 3 versus benchmark). We interpret this result as further evidence of the importance of the nonlinear components and the interaction between nonlinearity and non-normality.

8. Computational Issues

An attractive feature of particle filtering is that it can be implemented on a good PC. Nevertheless, the computational requirements of the particle filter are orders of magnitude bigger than those of the Kalman filter. On a Xeon Processor at 3.60 GHz, each evaluation of the likelihood with 80,000 particles takes around 18 seconds. The Kalman filter, applied to a linearized version of the model, takes a fraction of a second. The difference in computing time raises two questions. First, is it worth it? Second, can we apply the particle filter to richer models like those of Smets and Wouters (2003) or Christiano, Eichenbaum, and Evans (2005)?

With respect to the first question, Fernández-Villaverde and Rubio-Ramírez (2005) show that the particle filter improves inference with respect to the Kalman filter. In some contexts, this improvement may justify the extra computational effort. Regarding the second question, we point out that most of the computational time is spent in the **Sampling Step**. If we decompose the 18 seconds that each evaluation of the likelihood requires, we discover that the **Sampling Step** takes over 17 seconds, while the solution of the model takes less than 1 second. In an economy with even more state variables than ours (we already have 7 state variables!), we will only increase the computational time of the solution, while the **Sampling Step** will take roughly the same time. The availability of fast solution methods, like perturbation, implies that we can compute the nonlinear policy functions of a model with dozens of state variables in a few seconds. Consequently, an evaluation of the likelihood in such models would take around 20 seconds. This argument shows that the particle filter has the potential to be extended to the class of models needed for serious policy analysis.

To ensure the numerical accuracy of our results, we perform several numerical tests. First, we checked the number of particles required to achieve stable evaluations of the likelihood function. We found that 80,000 particles were more than enough for that purpose. Second, we computed the effective sample size (Arulampalam *et al.*, 2002, equation (51)) to check that we were not suffering from particle impoverishment due to sample depletion problems. We omit details in the interest of space. Finally, since version 1 of the model in the previous section has a linear state space representation with normal innovations, we can evaluate the likelihood both with the Kalman filter and with the particle filter. Both filters should deliver the same value of the likelihood function (the particle filter has a small-sample bias, but with 80,000 particles such bias is absolutely negligible). We corroborated that, in fact, both filters produce the same number up to numerical accuracy.

All programs were coded in Fortran 95 and compiled in Compaq Visual Fortran 6.6 to run on Windows-based PCs. All the code is available upon request.

9. Conclusions

We have presented a general purpose and asymptotically efficient algorithm to perform likelihood-based inference in nonlinear and/or non-normal dynamic macroeconomic models. We have shown how to undertake parameter estimation and model comparison, either from a classical or Bayesian perspective. The key ingredient has been the use of particle filtering to evaluate the likelihood function of the model. The intuition of the procedure is to simulate different paths for the states of the model and to ensure convergence by resampling with appropriately built weights.

We have applied the particle filter to estimate a business cycle model of the U.S. economy. We found strong evidence for the presence of stochastic volatility in U.S. data, that the decline in aggregate volatility has being occurring since the late 1950s, and changes in the volatility of preference shocks seem to be the main force behind the variation in the volatility U.S. output growth over the last 50 years.

Our current research applies particle filtering to models of nominal rigidities and parameter drifting, to dynamic general equilibrium models in continuous time, and to the estimation of dynamic games in macroeconomics.

References

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

[2] An, S. (2006). "Bayesian Estimation of DSGE Models: Lessons from Second-Order Approximations." *Mimeo,* University of Pennsylvania.

[3] An, S. and F. Schorfheide (2005). "Bayesian Analysis of DSGE Models." *CEPR Discussion Paper* 5207.

[4] Arulampalam, A.S., S. Maskell, N. Gordon, and T. Clapp (2002). "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking." *IEEE Transactions on Signal Processing* 50, 174-188.

[5] Aruoba, S.B., J. Fernández-Villaverde and J. Rubio-Ramírez (2006). "Comparing Solution Methods for Dynamic Equilibrium Economies.". *Journal of Economic Dynamics and Control*, forthcoming.

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

[7] Bernanke, B.S. and I. Mihov (1998). "Measuring Monetary Policy." *Quarterly Journal of Economics* 113, 869-902.

[8] Blanchard, O.J. and J. Simon (2001). "The Long and Large Decline in U.S. Output Volatility." *Brookings Papers on Economic Activity* 2001-1, 187-207.

[9] Bucy, R.S. and K.D. Senne (1971). "Digital Synthesis of Nonlinear Filters." *Automatica* 7, 287-298.

[10] Carrasco, M. and J.-P. Florens (2002). "Simulation Based Method of Moments and Efficiency." *Journal of Business and Economic Statistics* 20, 482-492.

[11] 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.

[12] Christiano, L.J., M. Eichenbaum and C.L. Evans (2005). "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* 113, 1-45.

[13] Cogley, T. and T.J. Sargent (2005). "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S." *Review of Economic Dynamics* 8, 262-302.

[14] 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.

[15] Del Moral P., and J. Jacod (2002), "The Monte-Carlo Method for Filtering with Discrete Time Observations. Central Limit Theorems." in T. J. Lyons and T. S. Salisbury (eds), *Numerical Methods and Stochastics*. The Fields Institute Communications, American Mathematical Society.

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

[17] Engle, R.F. (1982). "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation." *Econometrica* 50, 987-1008.

[18] Fermanian, J.D. and B. Salanié (2004). "A Nonparametric Simulated Maximum Likelihood Estimation Method." *Econometric Theory* 20, 701-734.

[19] Fernández-Villaverde, J. and J. Rubio-Ramírez (2004). "Comparing Dynamic Equilibrium Models to Data: A Bayesian Approach." *Journal of Econometrics* 123*,* 153-187.

[20] Fernández-Villaverde, J. and J. Rubio-Ramírez (2005). "Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood." *Journal of Applied Econometrics,* 20, 891-910.

[21] Fernández-Villaverde, J. and J. Rubio-Ramírez (2006). "A Crash Course on Nonlinear and/or Non-gaussian Filtering." *Mimeo,* University of Pennsylvania.

[22] Fernández-Villaverde, J., J. Rubio-Ramírez, and T.J. Sargent (2005). "A,B,C's (and D)'s for Understanding VARs." *NBER Technical Working Paper* T0308.

[23] Fernández-Villaverde, J., J. Rubio-Ramírez, and M.S. Santos (2006). "Convergence Properties of the Likelihood of Computed Dynamic Models." *Econometrica* 74, 93-119.

[24] Fisher, J. (2004). "Technology Shocks Matter." *Federal Reserve Bank of Chicago Working Paper* 2002-14 (revised).

[25] Geweke, J. (1989). "Bayesian Inference in Econometric Models Using Monte Carlo Integration." *Econometrica* 24, 1037-1399.

[26] 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.

[27] Godsill, S.J., A. Doucet and M. West (2004). "Monte Carlo Smoothing for Nonlinear Time Series." *Journal of the American Statistical Association* 99, 156-168.

[28] Greenwood, J, Z. Herkowitz, and P. Krusell (1997). "Long-Run Implications of Investment-Specific Technological Change." *American Economic Review* 87, 342-362.

[29] Greenwood, J, Z. Herkowitz, and P. Krusell (2000). "The Role of Investment-Specific Technological Change in the Business Cycle." *European Economic Review* 44, 91-115.

[30] Handschin, J. and D. Mayne (1969). "Monte Carlo Techniques to Estimate the Conditional Expectation in Multi-Stage Non-linear Filtering". *International Journal of Control* 9, 547-559.

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

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

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

[34] Julier, S.J. and J.K. Uhlmann (1996). "A New Extention of the Kalman Filter to Nonlinear Systems." In *The Proceedings of AeroSense: The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls*. Orlando, Florida.

[35] Justiniano A. and G.E. Primiceri (2005). "The Time Varying Volatility of Macroeconomic Fluctuations." *Mimeo*, Northwestern.

[36] Keane, M. (1994). "A Computationally Practical Simulation Estimator for Panel Data." *Econometrica* 62, 95-116.

[37] Kim, C.J. and C.R. Nelson (1999). "A Computationally Practical Simulation Estimator for Panel Data." *Review of Economics and Statistics* 81, 608-616.

[38] Kim, S., N. Shephard, and S. Chib (1998). "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models." *Review of Economic Studies* 65, 361-93.

[39] King, T.B. (2006). "Dynamic Equilibrium Models with Time-Varying Structural Parameters." *Mimeo*, Washington University.

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

[41] Künsch, H.R. (2005). "Recursive Monte Carlo Filters: Algorithms and Theoretical Analysis." *Annals of Statistics* 33, 1983-2021.

[42] Laforte, J.-P. (2005) DSGE Models and Heteroskedasticity: A Markov-Switching Approach." *Mimeo*, Board of Governors of the Federal Reserve System.

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

[44] Laroque, G. and B. Salanié (1989). "Estimation of Multimarket Fix-Price Models: an Application of Pseudo-Maximum Likelihood Methods." *Econometrica* 57, 831-860.

[45] Laroque, G. and B. Salanié (1993). "Simulation-based Estimation of Models with Lagged Latent Variables." *Journal of Applied Econometrics* 8, S119-S133.

[46] Laroque, G. and B. Salanié (1994). "Estimating the Canonical Disequilibrium Model: Asymptotic Theory and Finite Sample Properties." *Journal of Econometrics* 62, 165-210.

[47] 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.

[48] McConnell, M.M. and G. Pérez-Quirós (2000). "Output Fluctuations in the United States: What Has Changed Since the Early 1980's?" *American Economic Review* 90, 1464-1476.

[49] McFadden, D.L. (1989). "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration." *Econometrica* 57, 995-1026.

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

[51] 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.

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

[53] Pakes, A. and D. Pollard (1989). "Simulation and the Asymptotics of Optimization Estimators." *Econometrica* 57, 1027-1057.

[54] Pitt, M.K. (2002). "Smooth Particle Filters for Likelihood Evaluation and Maximization". *Mimeo*, University of Warwick.

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

[56] Primiceri, G.E. (2005). "Time Varying Structural Vector Autoregressions and Monetary Policy". *Review of Economic Studies* 72, 821-852.

[57] Rivers, D. and Q. Vuong (2002). "Model Selection Tests for Nonlinear Dynamic Models". *Econometrics Journal* 5, 1-39.

[58] Rubin, D.B. (1988). "Using the SIR Algorithm to Simulate Posterior Distributions". in J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith (eds), *Bayesian Statistics 3*, 395-402, Oxford University Press.

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

[60] Santos, M.S. and A. Peralta-Alva (2005). "Accuracy of Simulations for Stochastic Dynamic Models". *Econometrica* 73, 1939-1976.

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

[62] Shephard, N. (2005). *Stochastic Volatility, Selected Readings*. Oxford University Press.

[63] Sims, C.A. and T. Zha (2005). "Were There Regime Switches in U.S. Monetary Policy?" *American Economic Review*, forthcoming.

[64] Smets, F. and R. Wouters (2003). "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area." *Journal of the European Economic Association* 1, 1123-1175.

[65] Stenflo, O. (2001). "Ergodic Theorems for Markov Chains Represented by Iterated Function Systems." *Bulletin of the Polish Academy of Sciences: Mathematics* 49, 27-43.

[66] Stock, J.H. and M.W. Watson (2002). "Has the Business Cycle Changed, and Why?" *NBER Macroeconomics Annual* 17, 159-218.

[67] Stokey, N.L., R.E. Lucas and E.C. Prescott (1989). *Recursive Methods in Economic Dynamics*. Harvard University Press.

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

[69] Uhlig, H. (1997). "Bayesian Vector Autoregressions with Stochastic Volatility." *Econometrica* 65, 59-73.

[70] Winschel, V. (1995). "Solving, Estimating, Selecting Nonlinear Dynamic Economic Models without the Curse of Dimensionality." *Mimeo*, University of Mannheim.

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

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

Table 5.1: Maximum Likelihood Estimates

Parameter	Point Estimate	Standard Error ($x10^{-3}$)
ρ	0.967	3.743
β	0.999	0.460
ψ	2.343	6.825
υ	8.960E-003	0.828
ζ	3.594E-005	2.254
τ_a	7.120E-002	1.589
τ_υ	7.772E-003	2.940
τ_d	5.653E-002	2.034
$\overline{\sigma}_a$	4.008E-004	0.692
$\overline{\sigma}_\upsilon$	8.523E-003	0.101
$\overline{\sigma}_d$	5.016E-003	2.344
λ_a	4.460E-002	6.788
λ_υ	0.998	8.248
λ_d	0.998	2.302
$\sigma_{1\epsilon}$	1.031E-005	0.424
$\sigma_{2\epsilon}$	1.024E-004	0.495
$\sigma_{3\epsilon}$	1.110E-005	0.082

Figure 6.1: Model versus Data

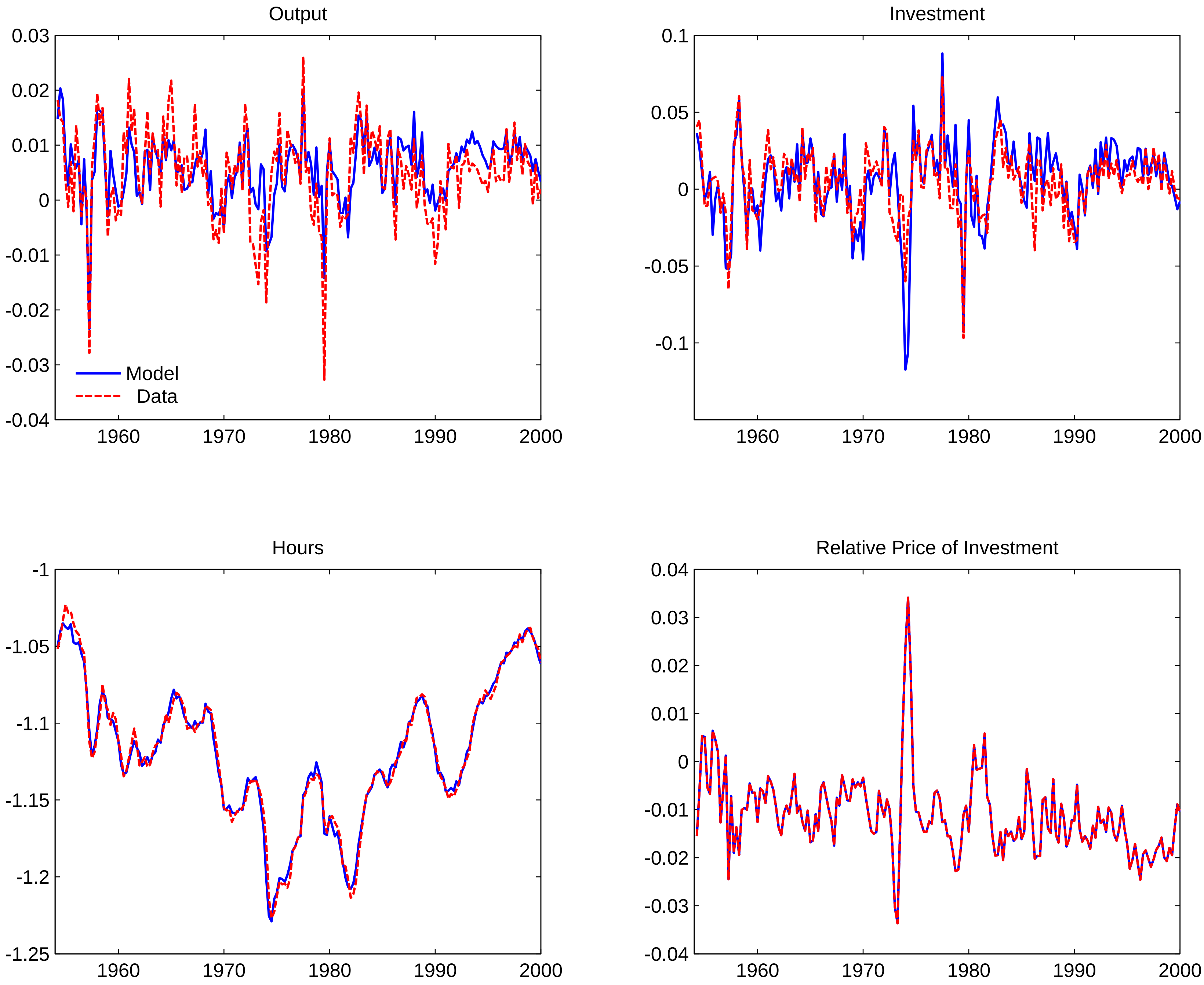

Figure 6.2: Smoothed Capital and Shocks

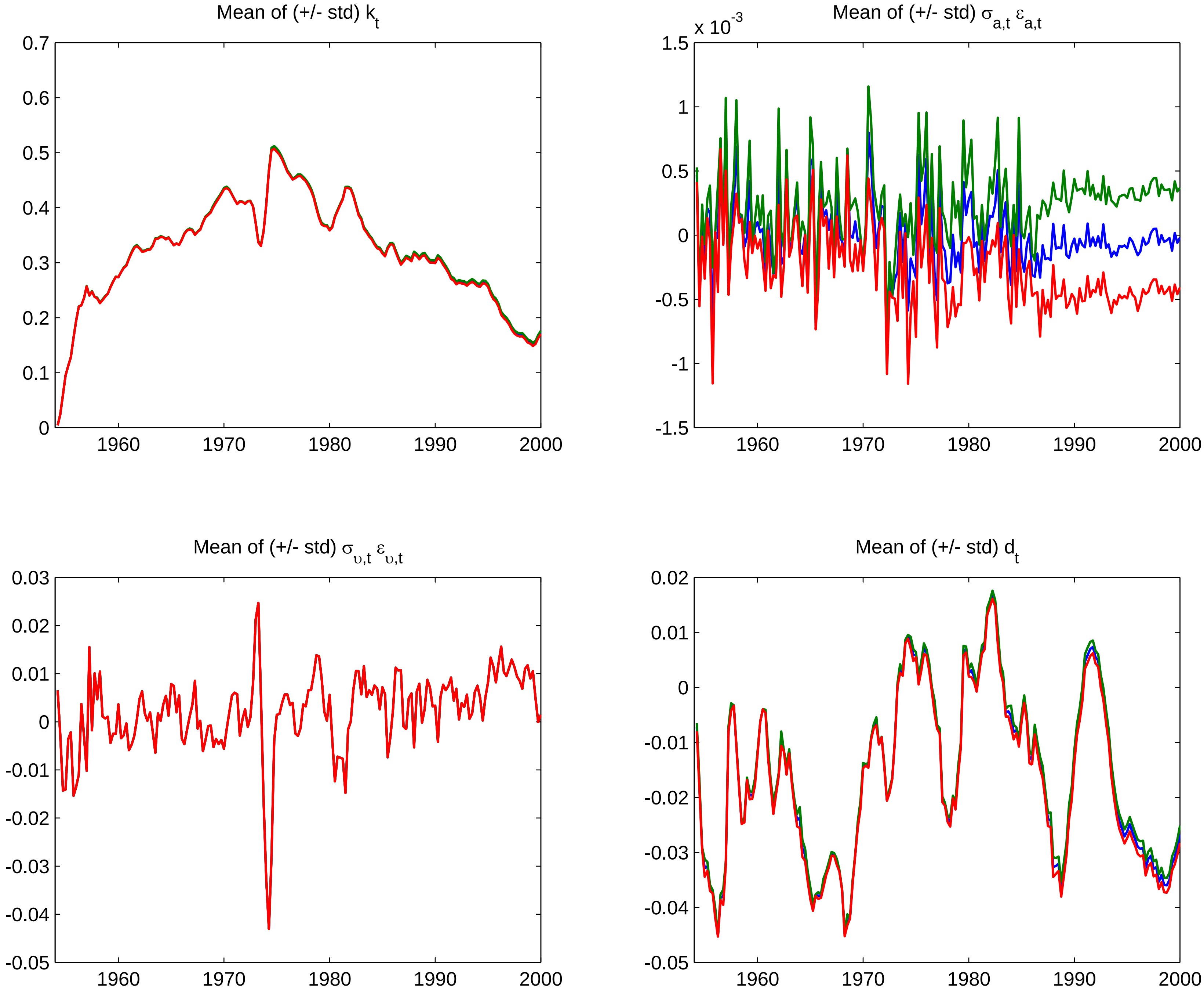

Figure 6.3: Smoothed Volatilities

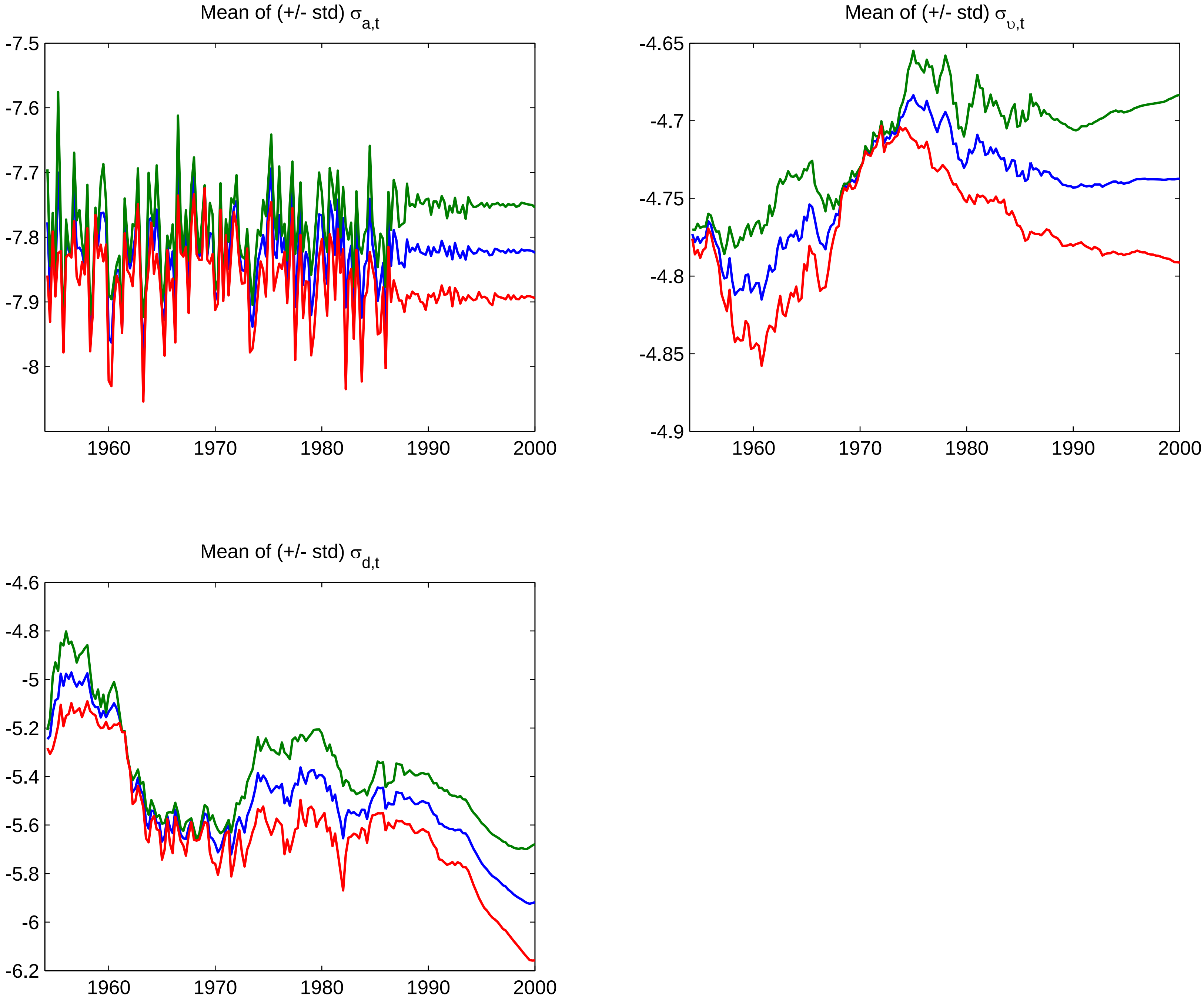

Figure 6.4: Instantaneous Standard Deviation

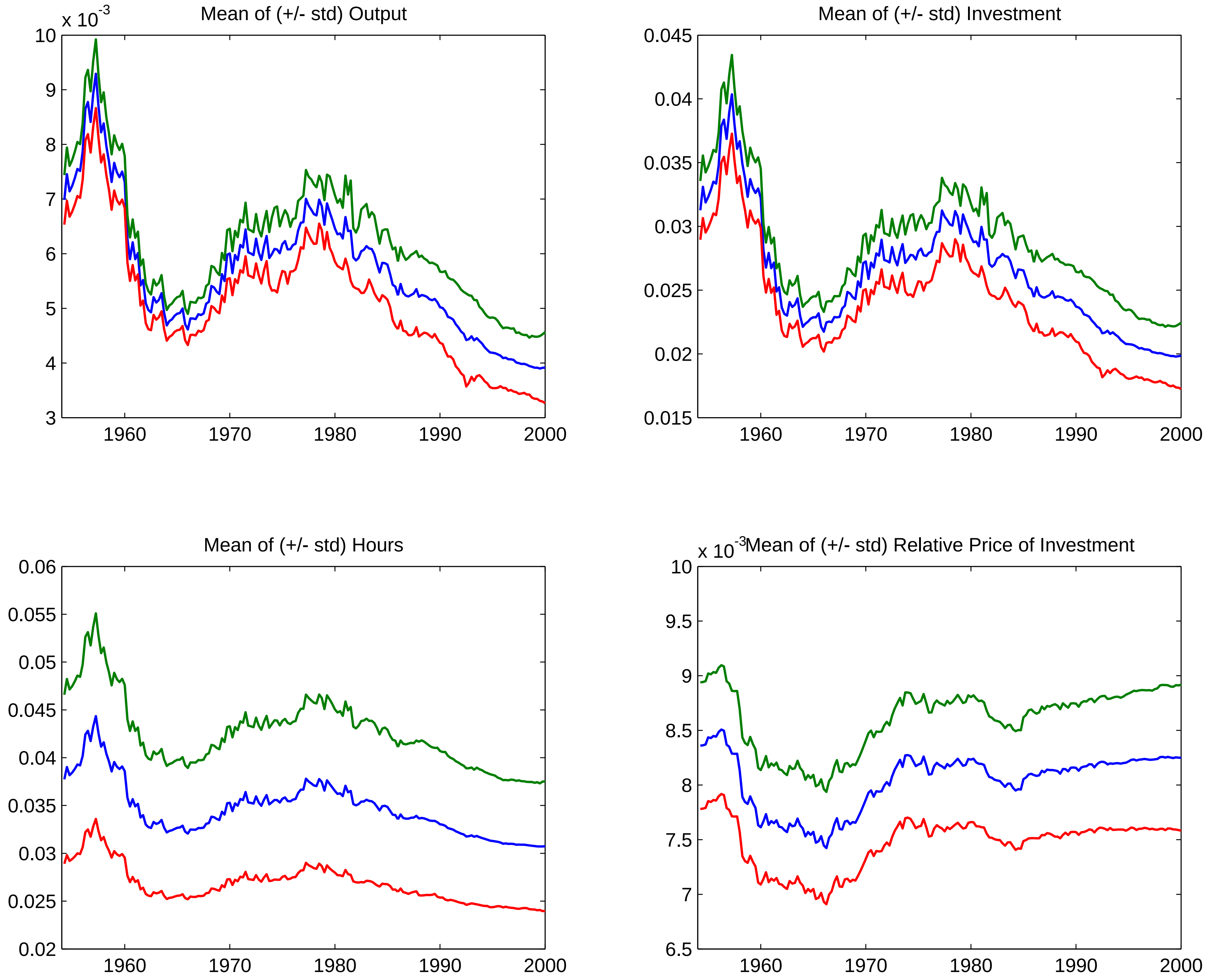

Figure 6.5: Counterfactual Exercise 1

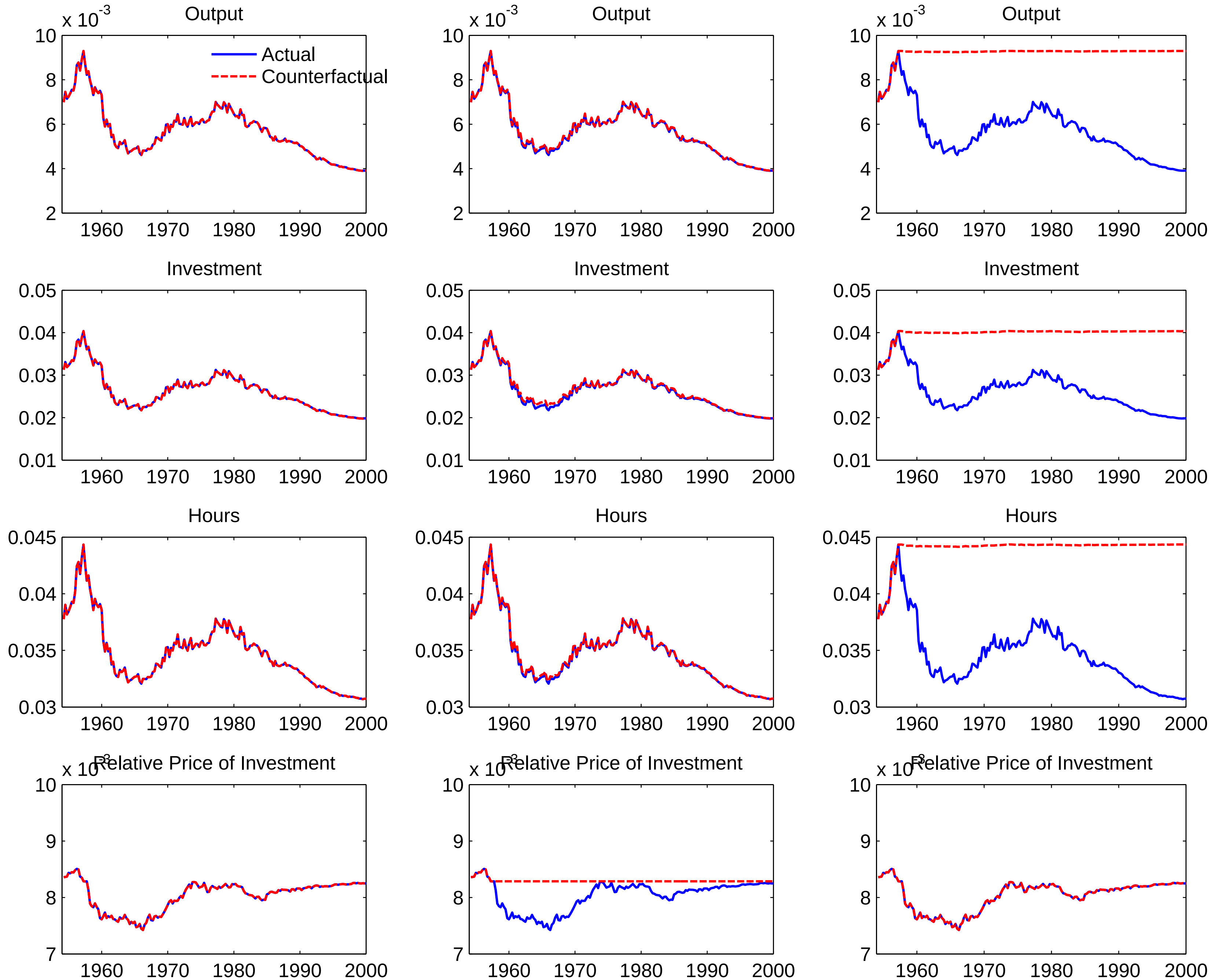

Figure 6.6: Counterfactual Exercise 2

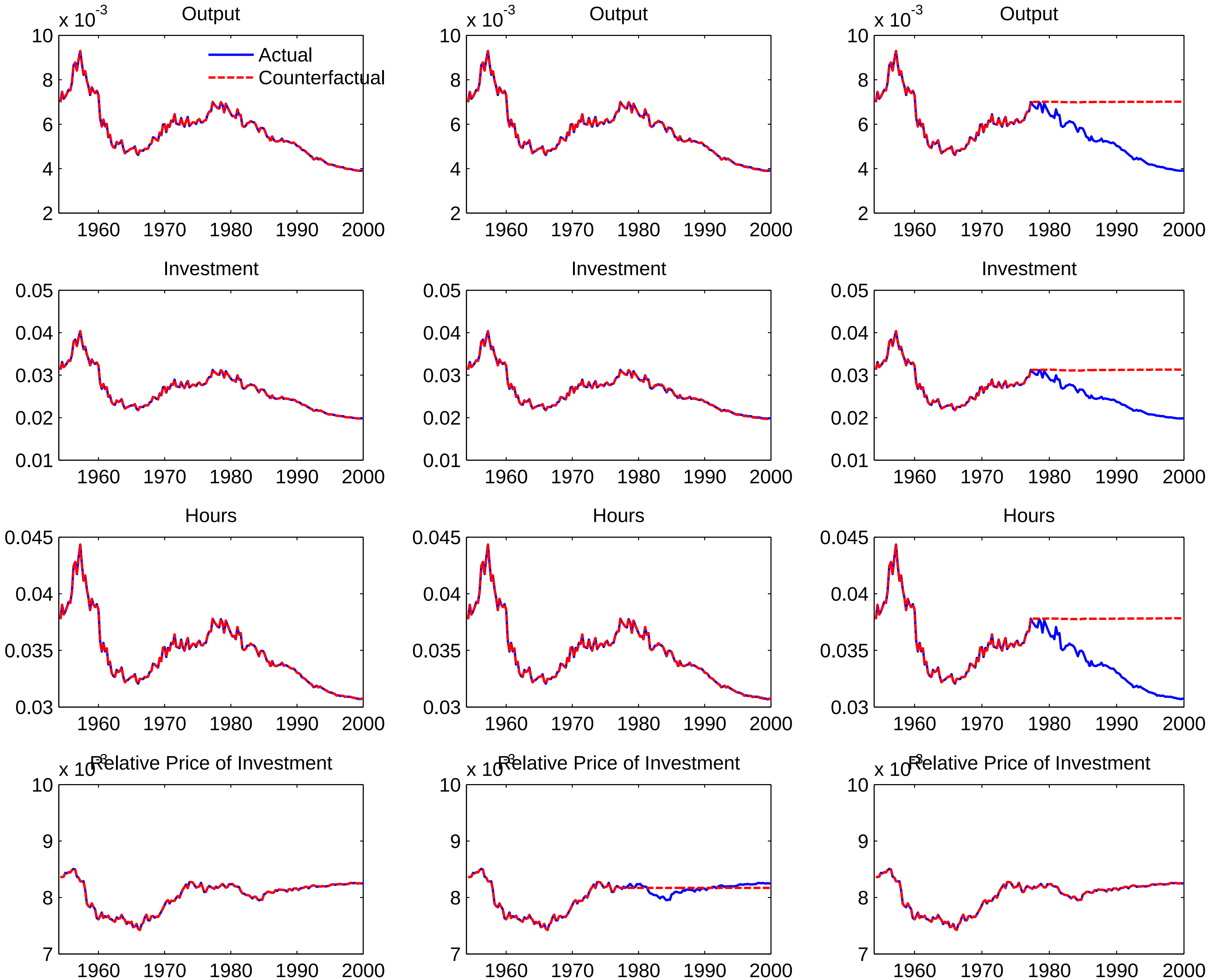

Figure 6.7: Counterfactual Exercise 3

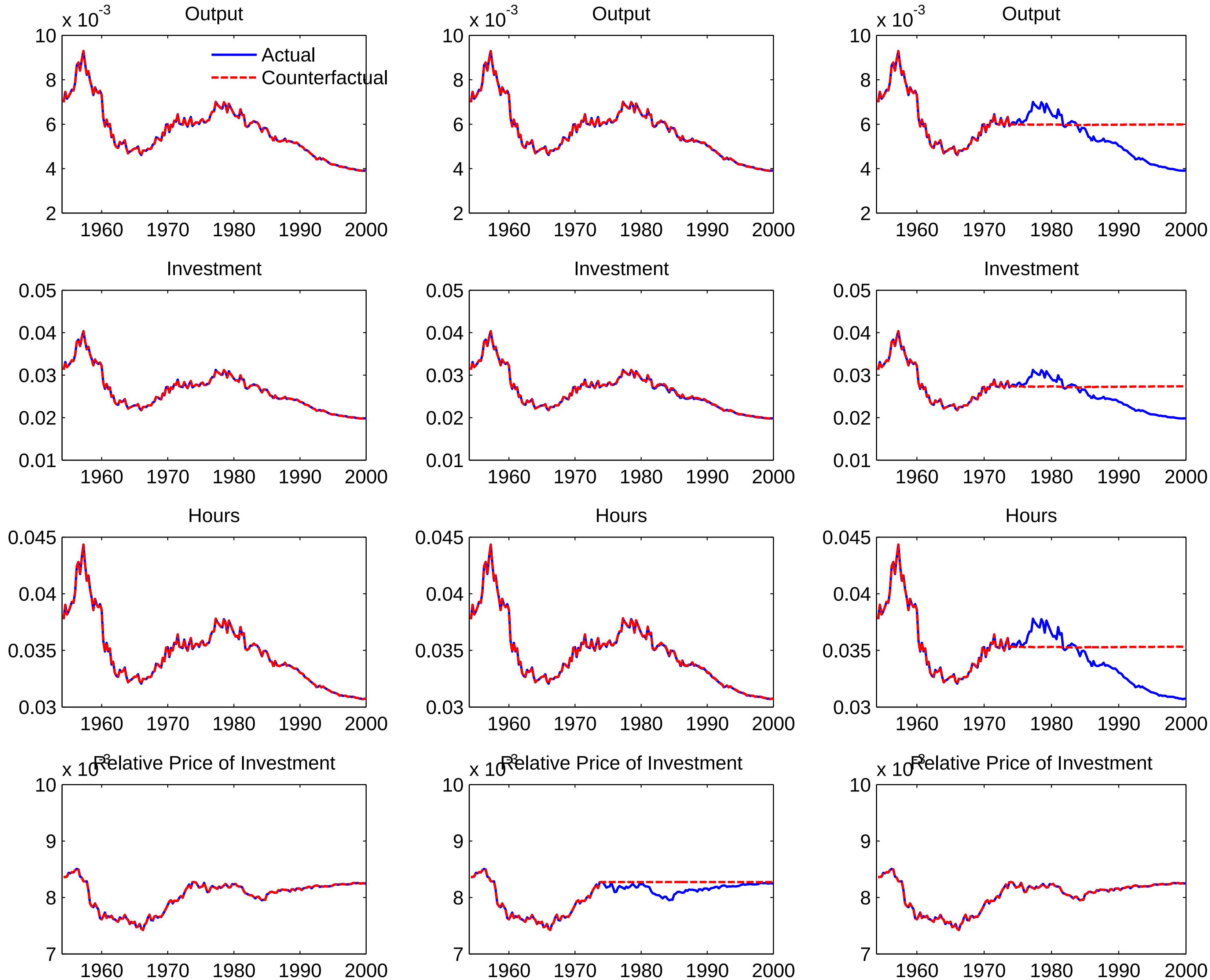

Figure 7.1: Comparison of Smoothed Capital and Shocks

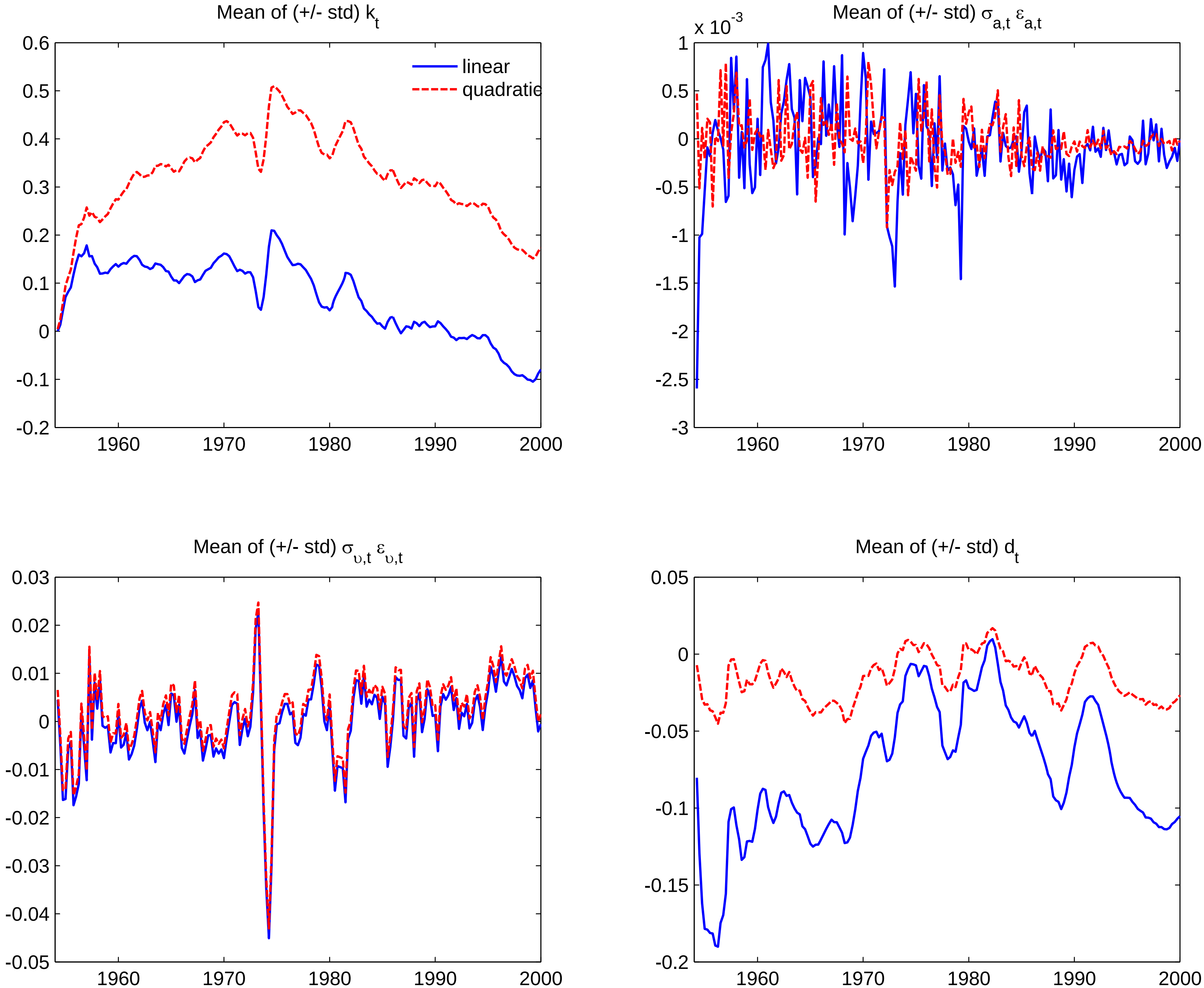

Figure 7.2: Comparison of Smoothed Volatilities