Review of Economic Studies (2007) **74**, 1059–1087 0034-6527/07/00371059$02.00

Estimating Macroeconomic Models: A Likelihood Approach

JESÚS FERNÁNDEZ-VILLAVERDE
University of Pennsylvania, NBER, and CEPR

and

JUAN F. RUBIO-RAMÍREZ
Duke University and Federal Reserve Bank of Atlanta

First version received December 2004; *final version accepted November* 2006 (*Eds.*)

This paper shows how particle filtering facilitates likelihood-based inference in dynamic macroeconomic models. The economies can be non-linear 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.

1. INTRODUCTION

This paper shows how particle filtering facilitates likelihood-based inference in dynamic equilibrium models. The economies can be non-linear 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 highlight three results. First, there is strong evidence of stochastic volatility on U.S. aggregate data. Second, two periods of low and falling aggregate volatility, from the late 1950's to the late 1960's and from the mid 1980's to today, were interrupted by a period of high and rising aggregate volatility from the late 1960's to the early 1980's. Third, variations in the volatility of preferences and investment-specific technological shocks account for most of the variation in the volatility of output growth over the last 50 years.

Likelihood-based inference is a useful tool to take dynamic equilibrium models to the data (An and Schorfheide, 2007). However, most dynamic equilibrium models do not imply a likelihood function that can be evaluated analytically or numerically. To circumvent this problem, the literature has used the approximated likelihood derived from a linearized version of the model, instead of the exact likelihood. But linearization depends on the accurate approximation of the solution of the model by a linear relation and on the shocks to the economy being distributed normally. Both assumptions are problematic.

First, 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. Fernández-Villaverde and Rubio-Ramírez (2005) document how those insights are quantitatively relevant to 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. Second, the assumption of normal shocks precludes investigating many models of interest, such as those with fat tails innovations, time-varying volatility, or autoregressive conditional duration.

To avoid these problems, we need to evaluate the likelihood function of non-linear and/or non-normal macroeconomic models. The particle filter is one procedure that will allow us to do so. We borrow from the literature on sequential Monte Carlo methods (see the book length survey in Doucet, de Freitas and Gordon (2001)). In economics, Pitt and Shephard (1999) and Kim, Shephard and Chib (1998) have pioneered the application of particle filters in financial econometrics. We adapt this know-how to handle the peculiarities of macroeconomic models. In particular, 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 solve for the equilibrium of the model with a non-linear solution method. The researcher can employ the solution algorithm that best fits her needs. With the solution of the model, we construct the state-space representation of the economy. 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 or find posterior distributions of the parameters, possibly by simulation. If we carry out the procedure with several models, we could compare them by building likelihood ratios or Bayes factors.

Particle filtering is a reasonably general purpose and is asymptotically efficient. However, it is not the only procedure we can use to perform likelihood estimation with dynamic models. Kim *et al.* (1998) conduct Bayesian inference on the model parameters and smoothing of states using a Markov chain Monte Carlo method. Fiorentini, Sentanta, and Shephard (2004) obtain maximum likelihood estimators with a simulated expectation-maximization (EM) algorithm. Fermanian and Salanié (2004) implement a nonparametric simulated likelihood method through kernel estimation. The technical appendix to this paper (available at www.econ.upenn.edu/~jesusfv) reviews further alternatives to compute the likelihood of non-linear and/or non-normal dynamic models.

There are also other routes to estimate dynamic models that do not rely on the likelihood function. One procedure is indirect inference (Gallant and Tauchen, 1996), which can, in theory, achieve the same asymptotic efficiency as maximum likelihood by an appropriate choice of auxiliary models. Moreover, simulated methods of moments like indirect inference will yield consistent estimators regardless of the number of simulations, while in our approach consistency can be guaranteed only if the number of simulations goes to infinity. However, the choice of an appropriate auxiliary model may be a challenging task. Alternatively, there is an important literature on simulated likelihood and simulated pseudo-likelihood applied to macroeconomic models (Laroque and Salanié, 1989, 1993). 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.[1]

As an application, we estimate a model of the U.S. business cycle with stochastic volatility on the shocks that drive the economy. Introducing stochastic volatility is convenient for two

1. We thank one referee for his help with the content of the last two paragraphs.

reasons. First, Kim and Nelson (1998), McConnell and Pérez-Quirós (2000), and Stock and Watson (2002) show that time-varying volatility is crucial for understanding U.S. data. This makes the application of interest *per se*. Second, stochastic volatility induces both fundamental non-linearities in the model and non-normal distributions. If we linearized the laws of motion for the shocks, the stochastic volatility terms would drop, killing any possibility of exploring this mechanism. Hence, our application demonstrates 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. We learn about the neutral technology and preference shocks from output, investment, and hours worked. The data reveal three patterns. First, there is compelling evidence for the presence of stochastic volatility in U.S. aggregate time series. Second, the decline in macro-volatility has been a gradual process since the late 1950's, interrupted by the turbulence of the 1970's, and not, as emphasized by the literature, an abrupt change around 1984. Third, the evolution in the volatility of preference and investment-specific technological shocks accounts for the variations in the volatility of output growth over the last 50 years. In addition, we provide evidence of how inference is crucially affected both by the non-linear component of the solution and by stochastic volatility.

We are aware of only one other paper that combines stochastic volatility and a dynamic general equilibrium model: the fascinating contribution of Justiniano and Primiceri (2005). Their innovative paper estimates a rich New Keynesian model of the business cycle with nominal and real rigidities. One difference between our papers is that the particle filter allows us to characterize the non-linear behaviour of the economy induced by stochastic volatility that Justiniano and Primiceri cannot handle. We document how including this non-linear component may be quantitatively important for inference.

We organize the rest of the paper as follows. We describe the particle filter in Section 2. We present our application in Section 3. We discuss computational details in Section 4. We conclude in Section 5.

2. A FRAMEWORK FOR LIKELIHOOD INFERENCE

In this section, we describe a framework to perform likelihood-based inference on a large class of non-linear and/or non-normal dynamic macroeconomic models. Cooley (1995) provides many examples of models in this class. Each of these economies implies a joint probability density function for observables given the model's structural parameters. This density, the likelihood function of the model, is in general difficult to evaluate. Particle filtering is a powerful route (although not the only one) to accomplish this goal.

We structure this section as follows. First, we present the state-space representation of a dynamic macroeconomic model. Second, we define the likelihood function of a dynamic model. Third, we discuss some of our technical assumptions. Fourth, we present a particle filter to evaluate that likelihood. Fifth, we link the filter with the estimation of the structural parameters of the model. We finish by discussing the smoothing of unobserved states.

2.1. *The state-space representation of a dynamic macroeconomic model*

Many 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 transition equation

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

where $\{W_t\}$ is a sequence of 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

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

where $\{V_t\}$ is a sequence of independent random variables that affect the observables but not the states. The sequences $\{W_t\}$ and $\{V_t\}$ are independent of each other. The random variables W_t and V_t are distributed as $p(W_t; \gamma)$ and $p(V_t; \gamma)$. We only require the ability to evaluate these densities. The vector γ also includes 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.

The functions f and g come from the equations that describe the behaviour of the model: policy functions, resource and budget constraints, equilibrium conditions, 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.

2.2. *The likelihood function of the model*

To continue our analysis we need to make three assumptions. Section 2.3 will explain in more detail the role of these assumptions.

Assumption 1. *There exists a partition of* $\{W_t\}$ *into two sequences* $\{W_{1,t}\}$ *and* $\{W_{2,t}\}$, *such that* $W_t = (W_{1,t}, W_{2,t})$ *and* $\dim(W_{2,t}) + \dim(V_t) \geq \dim(\mathbb{Y}_t)$ *for all* t.

The standard particle filter requires a stronger assumption, that is, $\dim(V_t) \geq \dim(\mathbb{Y}_t)$ for all t. This stronger assumption is restrictive in macroeconomics, where most researchers do not want to have a profusion of shocks. To extend the particle filter to a larger class of macroeconomic models, we relax that assumption by extracting from $\{W_t\}$ the sequence $\{W_{2,t}\}$. If $\dim(V_t) \geq \dim(\mathbb{Y}_t)$, we can set $W_{1,t} = W_t \ \forall t$, that is, $\{W_{2,t}\}$ is a zero-dimensional sequence. Conversely, if $\dim(W_t) + \dim(V_t) = \dim(\mathbb{Y}_t)$, we set $W_{2,t} = W_t$ for $\forall t$, that is, $\{W_{1,t}\}$ is a zero-dimensional sequence. We see our partition of the shocks as an important advantage of our approach.

Our partition, however, 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, $\dim(W_t) = 2$ and $\dim(W_{1,t}) = \dim(W_{2,t}) = 1$, 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 $\{W_{1,t}\}$. Our partition of $\{W_t\}$ into two nontrivial sequences $\{W_{1,t}\}$ and $\{W_{2,t}\}$ will allow us to estimate a model with four observables.

As notation for the rest of the paper, 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:

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

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

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

where S_0 is the initial state of the model, the p's represent the relevant densities. In the case that $\{W_{1t}\}$ has zero dimensions

$$\iint p(\mathcal{Y}_t|W_1^t, S_0, \mathcal{Y}^{t-1};\gamma)\, p(W_1^t, S_0|\mathcal{Y}^{t-1};\gamma)\, dW_1^t dS_0 = \int p(\mathcal{Y}_t|S_0, \mathcal{Y}^{t-1};\gamma)\, p(S_0|\mathcal{Y}^{t-1};\gamma)\, dS_0.$$

To save on notation, we assume herein that all the relevant Radon–Nikodym derivatives exist. Extending the exposition to the more general case is straightforward but cumbersome.

Expression (4) shows that the problem of evaluating the likelihood (3) amounts to solving an integral, with conditional densities $p(\mathcal{Y}_t|W_1^t, S_0, \mathcal{Y}^{t-1};\gamma)$ and $p(W_1^t, S_0|\mathcal{Y}^{t-1};\gamma)$ 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 through the Riccati equations. However, when the state representation is non-linear and/or non-normal, the conditional densities are no any longer normal, and we require a more powerful tool than the Kalman filter.

Before continuing, we present two additional assumptions.

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

$$\begin{aligned} S_1 &= f(s_0, (w_{1,1}, W_{2,1});\gamma), \\ \mathcal{Y}_m &= g(S_m, V_m;\gamma) \quad \text{for } m = 1, 2, \ldots t, \\ S_m &= f(S_{m-1}, (w_{1,m}, W_{2,m});\gamma) \quad \text{for } m = 2, 3, \ldots t \end{aligned}$$

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

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

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

for all γ, s_0, w_1^t, and t, where $|dy(v_t, w_{2,t};\gamma)|$ is the absolute value of the determinant of the jacobian of $\mathbb{Y}_t$ with respect to V_t and $W_{2,t}$ evaluated at v_t and $w_{2,t}$. 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 the evaluation if this is convenient.

An implication of (5) is that, to compute $p(\mathcal{Y}_t|w_1^t, s_0, \mathcal{Y}^{t-1};\gamma)$, we only need to solve a system of equations and evaluate the probability of observing the solution to the system,

$p(\upsilon_t, w_{2,t}; \gamma)$, times the absolute value of the determinant of the Jacobian evaluated at the solution, $|dy(\upsilon_t, w_{2,t}; \gamma)|$. The evaluation of $p(\upsilon_t, w_{2,t}; \gamma)$ is generally very simple. How difficult is it to evaluate the Jacobian? Often, this is also a simple task because the Jacobian depends on the derivatives of f and g, which can evaluate, at least, numerically for any given 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.

Finally, 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 $\mathcal{Y}^T$, that is,*

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

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

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

$$p(\mathcal{Y}^T; \gamma) \simeq \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} p(\mathcal{Y}_t | \widehat{w}_1^{t,i}, \widehat{s}_0^i, \mathcal{Y}^{t-1}; \gamma) \tag{6}$$

because of the law of large numbers. Hence, the problem of evaluating the likelihood of a dynamic model is equivalent to the problem of drawing from $\{p(W_1^t, S_0 | \mathcal{Y}^{t-1}; \gamma)\}_{t=1}^T$. This is a challenging problem because this conditional density is a complicated function of $\mathcal{Y}^{t-1}$.

2.3. *Assumptions and stochastic singularity*

Before continuing, we present some further discussion of our assumptions and relate them to stochastic singularity. Assumption 1 is a necessary condition for the model not to be stochastically singular. If Assumption 1 does not hold, that is, $\dim(W_t) + \dim(V_t) < \dim(\mathbb{Y}_t)$, the model cannot assign positive probability to the data because for an arbitrary S_{t-1} it is not possible to find a combination of W_t and V_t such that the model can explain $\mathcal{Y}_t$. This paper does not contribute to the literature on how to solve the problem of stochastic singularity of dynamic macroeconomic models. Therefore, we do not impose any restrictions on how the researcher adjusts the number of shocks and observables to satisfy this assumption. When we complement Assumption 1 with Assumption 3, we get a sufficient condition for the model not to be stochastically singular.

The need to combine Assumptions 1 and 3 also motivates our notation in terms of the history of innovations W_1^t. This notation is slightly more complicated than the standard notation in terms of states, S_t:

$$p(\mathcal{Y}^T; \gamma) = \prod_{t=1}^{T} p(\mathcal{Y}_t | \mathcal{Y}^{t-1}; \gamma) = \prod_{t=1}^{T} \int p(\mathcal{Y}_t | S_t, \mathcal{Y}^{t-1}; \gamma) p(S_t | \mathcal{Y}^{t-1}; \gamma) dS_t. \tag{7}$$

We chose our notation because, as explained above, we want to deal with the common case in macroeconomics where $\dim(V_t) < \dim(\mathbb{Y}_t)$. This case is difficult to handle with the standard notation. If $\dim(V_t) < \dim(\mathbb{Y}_t)$, then $p(\mathcal{Y}_t | S_t, \mathcal{Y}^{t-1}; \gamma) = 0$ almost surely (and hence (7) is also 0) because there is no V_t that is able to generate $\mathcal{Y}_t$ for an arbitrary S_t. In comparison, with the appropriate partition of the shocks, we can evaluate the likelihood given by (4).

Finally, note that Assumption 2 is a technical assumption to compute $p(\mathcal{Y}_t|w_1^t, s_0, \mathcal{Y}^{t-1}; \gamma)$ using the simple formula described in (5). If Assumption 2 does not hold, but Assumptions 1 and 3 do, the model is not stochastically singular and the particle filter can be applied. In that case, we will need to generalize formula (5) to consider cases where the solution to the system of equations is not unique.

2.4. *A particle filter*

The goal of the particle filter is to draw efficiently from $\{p(W_1^t, S_0|\mathcal{Y}^{t-1}; \gamma)\}_{t=1}^T$ to estimate the likelihood function of the model as given by (6).

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

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

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

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

Then

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

Proof. By Bayes' theorem

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

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

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

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

Corollary 1 shows how a draw $\{w_1^{t|t-1,i}, s_0^{t|t-1,i}\}_{i=1}^N$ from $p(W_1^t, S_0|\mathcal{Y}^{t-1}; \gamma)$ can be used to get a draw $\{w_1^{t,i}, s_0^{t,i}\}_{i=1}^N$ from $p(W_1^t, S_0|\mathcal{Y}^t; \gamma)$. How do we get the swarm $\{w_1^{t|t-1,i}, s_0^{t|t-1,i}\}_{i=1}^N$? By taking a swarm $\{w_1^{t-1,i}, s_0^{t-1,i}\}_{i=1}^N$ from $p(W_1^{t-1}, S_0|\mathcal{Y}^{t-1}; \gamma)$ and augmenting it with draws

from $p(W_{1,t};\gamma)$ since $p(W_1^t, S_0|\mathcal{Y}^{t-1};\gamma) = p(W_{1,t};\gamma)p(W_1^{t-1}, S_0|\mathcal{Y}^{t-1};\gamma)$. Note that $w_1^{t|t-1,i}$ is a growing object with t (it has the additional component of the draw from $p(W_{1,t};\gamma)$), while $s_0^{t|t-1,i}$ is not. Corollary 1 is crucial for the implementation of the particle filter. We discussed before how, when the model is non-linear and/or non-normal, the particle filter keeps track of a set of draws from $p(W_1^t, S_0|\mathcal{Y}^{t-1};\gamma)$ that are updated as new information is available. Corollary 1 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 $p(W_1^{t-1}, S_0|\mathcal{Y}^{t-1};\gamma) = p(S_0;\gamma)$.

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 $p(W_1^t, S_0|\mathcal{Y}^{t-1};\gamma) = p(W_{1,t};\gamma)p(W_1^{t-1}, S_0|\mathcal{Y}^{t-1};\gamma)$.

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

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 $(w_1^{t,i}, s_0^{t,i})$. 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

$$p(\mathcal{Y}^T;\gamma) \simeq \frac{1}{N}\left(\prod_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N} p(\mathcal{Y}_t|w_1^{t|t-1,i}, s_0^{t|t-1,i}, \mathcal{Y}^{t-1};\gamma)\right). \tag{8}$$

Since the particle filter does not require any assumption on the distribution of the shocks except the ability to evaluate $p(\mathcal{Y}_t|W_1^t, S_0, \mathcal{Y}^{t-1};\gamma)$, 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 R.H.S. of (8) is a consistent estimator of $p(\mathcal{Y}^T;\gamma)$
and a central limit theorem applies. However, the conditions required to apply those results may not hold in practice and they need to be checked (see, for further details, Koopman and Shephard, 2003).

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 $p(W_1^{t-1}, S_0|\mathcal{Y}^{t-1};\gamma)$, the Prediction Step generates draws $\left\{w_1^{t|t-1,i}, s_0^{t|t-1,i}\right\}_{i=1}^N$ from $p(W_1^t, S_0|\mathcal{Y}^{t-1};\gamma)$. In the case where $\dim(W_{1,t}) = 0$, the algorithm skips this step. The Sampling Step takes advantage of Corollary 1 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 $p(W_1^t, S_0|\mathcal{Y}^t;\gamma)$. A resampling procedure that minimizes the numerical variance (and the one we implement in our application) 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 $p(W_1^{t+1}, S_0|\mathcal{Y}^t;\gamma)$ 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 $\dim(W_{1,t}) > 0$ (Arulampalam, Maskell, Gordon and Clapp, 2002, pp. 178–179). 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. Since samples in macroeconomics are relatively long (200 observations or so), the degeneracy of SIS is a serious problem. However, our filter still may have the problem that a very small number of distinct sequences remain. This phenomenon, known as "sample depletion problem", has been documented in the literature. In Section 4, we describe how to test for this 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, that is, 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 $p(W_{1,T+1}; \gamma)$, $p(W_{2,T+1}; \gamma)$, and $p(V_{T+1}; \gamma)$, we build $p(\mathbb{Y}_{T+1}|\mathcal{Y}^T; \gamma)$. Second, we draw from $p(S_0; \gamma)$ in the Initialization Step using the transition and measurement equations (1) and (2) (see Santos and Peralta-Alva (2005) for details). 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)).

2.5. *Estimation algorithms*

We now explain how to employ the approximated likelihood function (8) to perform likelihood-based estimations from both a classical and a Bayesian perspective. On the classical side, the main inference tool is the likelihood function and its global maximum. Once the likelihood is approximated by (8), 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 $L(\mathcal{Y}^T;\gamma_i)$ using (8) and get γ_{i+1} from a maximization routine.

Step 3, Stopping Rule: If $\|L(\mathcal{Y}^T;\gamma_i) - L(\mathcal{Y}^T;\gamma_{i+1})\| > \xi$, 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, $\widehat{\gamma}_{\text{MLE}}$, is the maximum likelihood point estimate (MLE), with asymptotic variance–covariance matrix $\text{var}(\widehat{\gamma}_{\text{MLE}}) = -\left(\frac{\partial^2 L(\mathcal{Y}^T;\widehat{\gamma}_{\text{MLE}})}{\partial\gamma\,\partial\gamma'}\right)^{-1}$. 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 parameter values, the likelihood $L(\mathcal{Y}^T;\gamma)$ has to be differentiable with respect to γ. Furthermore, for the asymptotic variance–covariance matrix $\text{var}(\widehat{\gamma}_{\text{MLE}})$ to equal $-\left(\frac{\partial^2 L(\mathcal{Y}^T;\widehat{\gamma}_{\text{MLE}})}{\partial\gamma\,\partial\gamma'}\right)^{-1}$, $L(\mathcal{Y}^T;\gamma)$ has to be twice

differentiable with respect to γ. Remember that the likelihood can be written as

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

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

Under standard regularity conditions, we can prove that both $p(\mathcal{Y}_t|W_1^t,S_0,\mathcal{Y}^{t-1};\gamma)$ and $p(W_{1,t};\gamma)$ are twice differentiable (Fernández-Villaverde, Rubio-Ramírez and Santos, 2006). The differentiability of $\mu^*(dS_0;\gamma)$ 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 $\mu^*(dS_0;\gamma)$ is continuous. Hence, a proof that $\mu^*(dS_0;\gamma)$ is twice differentiable is a daunting task well beyond the scope of this paper.

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

Even if we were able to prove that $\mu^*(dS_0;\gamma)$ 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.

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 (8) to the exact likelihood is strengthened 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 $\pi(\gamma|\mathcal{Y}^T)$. 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, $\pi(\gamma)$, 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 $\pi(\gamma|\mathcal{Y}^T)$. The empirical distribution of those draws $\{\gamma_i\}_{i=1}^M$ converges to the true posterior distribution $\pi(\gamma|\mathcal{Y}^T)$. 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 $L(\mathcal{Y}^T;\gamma_i)$ with (8). Set $i \rightsquigarrow i+1$.

Step 1, Proposal draw: Get a draw γ_i^* from a proposal density $q(\gamma_{i-1},\gamma_i^*)$.

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 $L(\mathcal{Y}^T;\gamma_i^*)$ using (8).

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

This algorithm requires us to specify a proposal density $q(\cdot,\cdot)$. The most commonly used proposal density in macroeconomics is the random walk, $\gamma_i^* = \gamma_{i-1} + \kappa_i$, $\kappa_i \sim \mathcal{N}(0, \Sigma_\kappa)$, where Σ_κ is a scaling matrix (An and Schorfheide, 2007). This matrix is chosen to get the appropriate acceptance ratio of proposals (Roberts, Gelman and Gilks, 1997). The random walk proposal is straightforward to implement and the experience accumulated by researchers suggests that it usually works quite efficiently for macroeconomic applications.

Chib and Greenberg (1995) describe many other alternatives for the proposal density, like an independent chain or a pseudo-dominating density. A popular class of proposal densities, those that exploit some property (or approximated property) of the likelihood function, are difficult to implement in macroeconomics, since we know little about the shape of the likelihood.

2.6. *Smoothing*

Often, beyond filtering, we are also interested in $p(S^T|\mathcal{Y}^T;\gamma)$, that is, 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.

First, we analyse how to assess the fit of the model. Given 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 perturbations that affect the observables but not the states, $\mathbb{Y}^T(S^T, V^T;\gamma)$. Define, for any γ, the mean of the observables implied by the model and the data, $\mathcal{Y}^T$:

$$\overline{\mathbb{Y}}^T(V^T;\gamma) = \int \mathbb{Y}^T(S^T, V^T;\gamma)\, p(S^T|\mathcal{Y}^T;\gamma)\, dS^T. \tag{9}$$

If V^T are measurement errors, comparing $\overline{\mathbb{Y}}^T(V^T = 0;\gamma)$ vs. $\mathcal{Y}^T$ 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

$$\overline{\mathbb{Y}}^T_{S_k^{t:T}=S_{k,t}}(V^T;\gamma) = \int \mathbb{Y}^T(S_{-k}^T, S_k^{t:T} = S_{k,t}, V^T;\gamma)\, p(S^T|\mathcal{Y}^T;\gamma)\, dS^T, \tag{10}$$

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

The two examples share a common theme. To compute integrals like (9) or (10), we need to draw from $p(S^T|\mathcal{Y}^T;\gamma)$. To see this, let $\{s^{t,i}\}_{i=1}^N$ be a draw from $p(S^T|\mathcal{Y}^T;\gamma)$. Then (9) and (10) are approximated by

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

and

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

Particle filtering allows us to draw from $p(S^T|\mathcal{Y}^T;\gamma)$ using the simulated filtered distribution (see Godsill, Doucet, and West, 2004). In the interest of space, we omit here the details, but the interested reader can find them in the technical appendix of the paper.

3. AN APPLICATION: A BUSINESS CYCLE MODEL

In this section, we apply particle filtering to 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.

Recently, Kim and Nelson (1998) 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). This evidence begets the question of what has caused the change in volatility.

One possible explanation is that the shocks hitting the economy were very different in the 1990's than in the 1970's (Sims and Zha, 2006). 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. Thus, the literature has estimated structural vector autoregressions (SVARs). Despite their flexibility, SVARs may uncover evidence that is difficult to interpret from the perspective of the theory.

The particle filter is perfectly suited for analysing dynamic equilibrium models with stochastic volatility. In comparison, the Kalman filter and linearization are 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 non-linear. 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 application should be assessed as an example of the type of exercises that can be undertaken with the particle filter. 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 estimated dynamic macroeconomic models with particle filtering. Amisano and Tristani (2005) and An (2006) investigate New Keynesian models. Particle filtering allows them 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. Winschel (2005) applies the Smolyak operator to accelerate the numerical performance of the algorithm.

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, 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 (2006).

There is a representative household with utility function

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \{e^{d_t} \log(C_t) + \psi \log(1 - L_t)\},$$

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

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

ψ controls labour supply, and $\mathbb{E}_0$ 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 $C_t + X_t = A_t K_t^{\alpha} L_t^{1-\alpha}$. The law of motion for capital is $K_{t+1} = (1 - \delta) K_t + U_t X_t$. Technologies evolve as a random walk with drifts

$$\log A_t = \zeta + \log A_{t-1} + \sigma_{at} \varepsilon_{at}, \ \gamma \geq 0 \text{ and } \varepsilon_{at} \sim \mathcal{N}(0, 1), \tag{11}$$

$$\log U_t = \theta + \log U_{t-1} + \sigma_{\upsilon t} \varepsilon_{\upsilon t}, \ \theta \geq 0 \text{ and } \varepsilon_{\upsilon t} \sim \mathcal{N}(0, 1). \tag{12}$$

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 (Shephard, 2005)

$$\log \sigma_{at} = (1 - \lambda_a) \log \overline{\sigma}_a + \lambda_a \log \sigma_{at-1} + \tau_a \eta_{at} \text{ and } \eta_{at} \sim \mathcal{N}(0, 1), \tag{13}$$

$$\log \sigma_{\upsilon t} = (1 - \lambda_\upsilon) \log \overline{\sigma}_\upsilon + \lambda_\upsilon \log \sigma_{\upsilon t-1} + \tau_\upsilon \eta_{\upsilon t} \text{ and } \eta_{\upsilon t} \sim \mathcal{N}(0, 1), \tag{14}$$

$$\log \sigma_{dt} = (1 - \lambda_d) \log \overline{\sigma}_d + \lambda_d \log \sigma_{dt-1} + \tau_d \eta_{dt} \text{ and } \eta_{dt} \sim \mathcal{N}(0, 1). \tag{15}$$

Thus, the matrix of unconditional variances–covariances $\mathbf{\Omega}$ of the shocks is a diagonal matrix with entries $\{\overline{\sigma}_a^2, \overline{\sigma}_\upsilon^2, \overline{\sigma}_d^2, \tau_a^2, \tau_\upsilon^2, \tau_d^2\}$.

Our model is then indexed by a vector of structural parameters:

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

where $\sigma_{1\varepsilon}, \sigma_{2\varepsilon}$, and $\sigma_{3\varepsilon}$ are the S.D. of three measurement errors to be introduced below.

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. Since the model has two unit roots in the technological processes, we rescale the variables by $\widetilde{Y}_t = \frac{Y_t}{Z_t}$, $\widetilde{C}_t = \frac{C_t}{Z_t}$, $\widetilde{X}_t = \frac{X_t}{Z_t}$, $\widetilde{A}_t = \frac{A_t}{A_{t-1}}$, $\widetilde{U}_t = \frac{U_t}{U_{t-1}}$, $\widetilde{Z}_t = \frac{Z_t}{Z_{t-1}}$, and $\widetilde{K}_t = \frac{K_t}{Z_t U_{t-1}}$ where $Z_t = A_{t-1}^{1/(1-\alpha)} U_{t-1}^{\alpha/(1-\alpha)}$.

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

$$\frac{\widetilde{Z}_t}{\widetilde{A}_t \widetilde{K}_t^{\alpha} L_t^{-\alpha} (1 - L_t)} = \mathbb{E}_t \frac{\beta}{\widetilde{A}_{t+1} \widetilde{K}_{t+1}^{\alpha} L_{t+1}^{-\alpha} (1 - L_{t+1})} (\alpha \widetilde{A}_{t+1} \widetilde{K}_{t+1}^{\alpha-1} L_{t+1}^{1-\alpha} + \frac{(1-\delta)}{\widetilde{U}_{t+1}}), \tag{16}$$

the law of motion for capital

$$\widetilde{Z}_{t+1} \widetilde{K}_{t+1} = (1-\delta) \frac{\widetilde{K}_t}{\widetilde{U}_t} + \widetilde{X}_t, \tag{17}$$

and the resource constraint

$$\frac{1-\alpha}{\psi} \widetilde{A}_t \widetilde{K}_t^{\alpha} L_t^{-\alpha} e^{\rho d_{t-1} + \sigma_{dt} \varepsilon_{dt}} (1 - L_t) + \widetilde{X}_t = \widetilde{A}_t \widetilde{K}_t^{\alpha} L_t^{1-\alpha}. \tag{18}$$

These equations imply a deterministic steady state around which we will approximate the solution of our model. Let L_{ss}, $\widetilde{X}_{ss}$, and $\widetilde{K}_{ss}$ be the steady-state values for hours worked, rescaled investment, and rescaled capital.

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 (16)–(18) 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. However, nothing in the particle filter stops us from opting for any other non-linear solution method, such 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 $\overline{\mathbf{\Omega}}(\chi) = \chi \mathbf{\Omega}$, where $\overline{\mathbf{\Omega}}(1) = \mathbf{\Omega}$. Then, we take a perturbation solution around $\chi = 0$, that is, around the deterministic steady state implied by the equilibrium conditions of the model.

For any variable var_t, let $\widehat{\text{var}}_t = \log \frac{\text{var}_t}{\text{var}_{ss}}$. Then, the states of the model are given by

$$\widetilde{s}_t = (\widehat{\widetilde{K}}_t, \varepsilon_{at}, \varepsilon_{\upsilon t}, \varepsilon_{dt}, d_{t-1}, \sigma_{at} - \overline{\sigma}_a, \sigma_{\upsilon t} - \overline{\sigma}_\upsilon, \sigma_{dt} - \overline{\sigma}_d)'.$$

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

$$\widehat{\widetilde{K}}_{t+1} = \mathbf{\Psi}_{k1} \widetilde{s}_t + \frac{1}{2} \widetilde{s}_t' \mathbf{\Psi}_{k2} \widetilde{s}_t + \mathbf{\Psi}_{k3}, \tag{19}$$

where $\mathbf{\Psi}_{k1}$ is a 1×8 vector and $\mathbf{\Psi}_{k2}$ is a 8×8 matrix. The term $\mathbf{\Psi}_{k1}\widetilde{s}_t$ constitutes the linear solution of the model, while Ψ_{k3} is a constant added by the second-order approximation that corrects for precautionary behaviour.

Similarly, we can find the policy functions for rescaled investment and labour:

$$\widehat{\widetilde{X}}_t = \mathbf{\Psi}_{x1}\widetilde{s}_t + \frac{1}{2}\widetilde{s}_t'\mathbf{\Psi}_{x2}\widetilde{s}_t + \Psi_{x3},$$

$$\widehat{\widetilde{L}}_t = \mathbf{\Psi}_{l1}\widetilde{s}_t + \frac{1}{2}\widetilde{s}_t'\mathbf{\Psi}_{l2}\widetilde{s}_t + \Psi_{l3}.$$

Plugging those policy functions into the definition of rescaled output $\widetilde{Y}_t = \widetilde{A}_t\widetilde{K}_t^{\alpha}L_t^{1-\alpha}$, we can get a second-order approximation

$$\widehat{\widetilde{Y}}_t = \mathbf{\Psi}_{y1}\widetilde{s}_t + \frac{1}{2}\widetilde{s}_t'\mathbf{\Psi}_{y2}\widetilde{s}_t + \Psi_{y3} + \sigma_{at}\varepsilon_{at}. \tag{20}$$

3.3. *The transition equation*

We combine the laws of motion for the volatility (13)–(15) and the policy function of capital (19) to build

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

where $S_t = (\widetilde{s}_t, \widetilde{s}_{t-1})$ and $W_t = (\varepsilon_{at}, \varepsilon_{\upsilon t}, \varepsilon_{dt}, \eta_{at}, \eta_{\upsilon t}, \eta_{dt})'$. We keep track of the past states, $\widetilde{s}_{t-1}$, 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

$$\begin{aligned}
f_1(S_{t-1}, W_t; \gamma) &= \mathbf{\Psi}_{k1}\widetilde{s}_{t-1} + \frac{1}{2}\widetilde{s}_{t-1}'\mathbf{\Psi}_{k2}\widetilde{s}_{t-1} + \Psi_{k3}\\
f_2(S_{t-1}, W_t; \gamma) &= \varepsilon_{at}\\
f_3(S_{t-1}, W_t; \gamma) &= \varepsilon_{\upsilon t}\\
f_4(S_{t-1}, W_t; \gamma) &= \varepsilon_{dt}\\
f_5(S_{t-1}, W_t; \gamma) &= \rho d_{t-2} + \sigma_{dt-1}\varepsilon_{dt-1}\\
f_6(S_{t-1}, W_t; \gamma) &= \exp((1-\lambda_a)\log\overline{\sigma}_a + \lambda_a\log\sigma_{at-1} + \tau_a\eta_{at}) - \overline{\sigma}_a\\
f_7(S_{t-1}, W_t; \gamma) &= \exp((1-\lambda_\upsilon)\log\overline{\sigma}_\upsilon + \lambda_\upsilon\log\sigma_{\upsilon t-1} + \tau_\upsilon\eta_{\upsilon t}) - \overline{\sigma}_\upsilon\\
f_8(S_{t-1}, W_t; \gamma) &= \exp((1-\lambda_d)\log\overline{\sigma}_d + \lambda_d\log\sigma_{dt-1} + \tau_d\eta_{dt}) - \overline{\sigma}_d\\
f_{9-16}(S_{t-1}, W_t; \gamma) &= \widetilde{s}_{t-1}.
\end{aligned}$$

3.4. *The measurement equation*

We assume that we observe $\mathcal{Y}_t = (\Delta\log p_t, \Delta\log y_t, \Delta\log x_t, \log l_t)$, that is, 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 the 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 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 U_t:

$$-\Delta \log U_t = -\theta - \sigma_{\upsilon t}\varepsilon_{\upsilon t}.$$

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

To build the measurement equation for real output *per capita* growth, we remember that

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

which implies that

$$\begin{aligned}\Delta \log Y_t &= \Delta \log \widetilde{Y}_t + \frac{1}{1-\alpha}(\Delta \log A_{t-1} + \alpha \Delta \log U_{t-1}) \\ &= \Delta \log \widetilde{Y}_t + \frac{1}{1-\alpha}(\zeta + \alpha\theta + \sigma_{at-1}\varepsilon_{at-1} + \alpha\sigma_{\upsilon t-1}\varepsilon_{\upsilon t-1}).\end{aligned}$$

Using (20), we have that

$$\begin{aligned}\Delta \log Y_t = {} & \mathbf{\Psi}_{y1}\widetilde{s}_t + \frac{1}{2}\widetilde{s}_t'\mathbf{\Psi}_{y2}\widetilde{s}_t - \mathbf{\Psi}_{y1}\widetilde{s}_{t-1} - \frac{1}{2}\widetilde{s}_{t-1}'\mathbf{\Psi}_{y2}\widetilde{s}_{t-1} \\ & + \sigma_{at}\varepsilon_{at} + \frac{1}{1-\alpha}(\zeta + \alpha\theta + \alpha\sigma_{at-1}\varepsilon_{at-1} + \alpha\sigma_{\upsilon t-1}\varepsilon_{\upsilon t-1}).\end{aligned}$$

This equation shows that, on average, the growth rate of the economy will be equal to $\frac{\zeta+\alpha\theta}{1-\alpha}$. Similarly, for real gross investment *per capita* growth, we can find

$$\begin{aligned}\Delta \log X_t = {} & \mathbf{\Psi}_{x1}\widetilde{s}_t - \mathbf{\Psi}_{x1}\widetilde{s}_{t-1} + \frac{1}{2}\widetilde{s}_t'\mathbf{\Psi}_{x2}\widetilde{s}_t - \frac{1}{2}\widetilde{s}_{t-1}'\mathbf{\Psi}_{x2}\widetilde{s}_{t-1} \\ & + \frac{1}{1-\alpha}(\zeta + \alpha\theta + \sigma_{at-1}\varepsilon_{at-1} + \alpha\sigma_{\upsilon t-1}\varepsilon_{\upsilon t-1}).\end{aligned}$$

Finally, we introduce measurement errors in 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 1–3). 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 real output *per capita* growth, ε_{1t}, one in real gross investment *per capita* growth, ε_{2t}, and one in hours worked *per capita*, ε_{3t}. We do not have a measurement error in relative price of investment because it will not be possible to separately identify it from $\sigma_{\upsilon t}\varepsilon_{\upsilon t}$. The three shocks are an i.i.d. process with distribution $\mathcal{N}(0, I_{3\times 3})$. In our notation of Section 2, $V_t = (\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})'$. The measurement errors imply a difference between the value for the variables implied by the model and the observables:

$$\begin{aligned}\Delta \log p_t &= -\Delta \log U_t, \\ \Delta \log y_t &= \Delta \log Y_t + \sigma_{1\varepsilon}\varepsilon_{1t}, \\ \Delta \log x_t &= \Delta \log X_t + \sigma_{2\varepsilon}\varepsilon_{2t}, \\ \log l_t &= \log L_t + \sigma_{3\varepsilon}\varepsilon_{3t}.\end{aligned}$$

Putting the different pieces together, we have the measurement equation

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

$$+ \begin{pmatrix} -\sigma_{\upsilon t}\varepsilon_{\upsilon t} \\ \boldsymbol{\Psi}_{y1}\widetilde{s}_t + \frac{1}{2}\widetilde{s}_t'\boldsymbol{\Psi}_{y2}\widetilde{s}_t - \boldsymbol{\Psi}_{y1}\widetilde{s}_{t-1} - \frac{1}{2}\widetilde{s}_{t-1}'\boldsymbol{\Psi}_{y2}\widetilde{s}_{t-1} + \sigma_{at}\varepsilon_{at} + \frac{\alpha}{1-\alpha}(\sigma_{at-1}\varepsilon_{at-1} + \sigma_{\upsilon t-1}\varepsilon_{\upsilon t-1}) \\ \boldsymbol{\Psi}_{x1}\widetilde{s}_t - \boldsymbol{\Psi}_{x1}\widetilde{s}_{t-1} + \frac{1}{2}\widetilde{s}_t'\boldsymbol{\Psi}_{x2}\widetilde{s}_t - \frac{1}{2}\widetilde{s}_{t-1}'\boldsymbol{\Psi}_{x2}\widetilde{s}_{t-1} + \frac{1}{1-\alpha}(\sigma_{at-1}\varepsilon_{at-1} + \alpha\sigma_{\upsilon t-1}\varepsilon_{\upsilon t-1}) \\ \boldsymbol{\Psi}_{l1}\widetilde{s}_t + \frac{1}{2}\widetilde{s}_t'\boldsymbol{\Psi}_{l2}\widetilde{s}_t \end{pmatrix}$$

$$+ \begin{pmatrix} 0 \\ \sigma_{1\varepsilon}\varepsilon_{1t} \\ \sigma_{3\varepsilon}\varepsilon_{2t} \\ \sigma_{3\varepsilon}\varepsilon_{3t} \end{pmatrix}.$$

3.5. *The likelihood function*

Given that we have six structural shocks in the model and $\dim(V_t) < \dim(\mathbb{Y}_t)$, we partition W_t between $W_{1,t} = (\varepsilon_{at}, \varepsilon_{dt}, \eta_{dt}, \eta_{at}, \eta_{\upsilon t})'$ and $W_{2,t} = \varepsilon_{\upsilon t}$. This partition shows that we can learn much about $W_{2,t}$ directly from the data on the relative price of investment. Then, in the notation of Section 2, the prediction errors are

$$\omega_{2,t} = \omega_{2,t}(W_1^t, S_0, \mathcal{Y}^t; \gamma) = -\frac{\Delta \log p_t + \theta}{\sigma_{\upsilon t}},$$

$$\upsilon_{1,t} = \upsilon_{1,t}(W_1^t, S_0, \mathcal{Y}^t; \gamma) = \frac{\Delta \log y_t - \Delta \log Y_t}{\sigma_{1\varepsilon}},$$

$$\upsilon_{2,t} = \upsilon_{2,t}(W_1^t, S_0, \mathcal{Y}^t; \gamma) = \frac{\Delta \log x_t - \Delta \log X_t}{\sigma_{2\varepsilon}},$$

and

$$\upsilon_{3,t} = \upsilon_{3,t}(W_1^t, S_0, \mathcal{Y}^t; \gamma) = \frac{\log l_t - \log L_t}{\sigma_{3\varepsilon}}.$$

Let $\xi_t = (\omega_{2t}, \upsilon_t)$. Then, we can write

$$p(\mathcal{Y}_t | W_1^t, S_0, \mathcal{Y}^{t-1}; \gamma) = p(\xi_t)|dy(\xi_t)|^{-1},$$

where

$$p(\xi_t) = (2\pi)^{-\frac{4}{2}} \exp\left(-\frac{\xi_t \xi_t'}{2}\right)$$

and the Jacobian of $\mathbb{Y}_t$ with respect to V_t and $W_{2,t}$ evaluated at v_t and $w_{2,t}$ is

$$dy(\xi_t) = \begin{bmatrix} -\sigma_{vt} & 0 & 0 & 0 \\ \frac{\partial \triangle \log Y_t}{\partial \log \varepsilon_{vt}} & \sigma_{1\varepsilon} & 0 & 0 \\ \frac{\partial \triangle \log X_t}{\partial \log \varepsilon_{vt}} & 0 & \sigma_{2\varepsilon} & 0 \\ \frac{\partial \log L_t}{\partial \log \varepsilon_{vt}} & 0 & 0 & \sigma_{3\varepsilon} \end{bmatrix},$$

which implies that $|dy(\xi_t)| = \sigma_{vt}\sigma_{1\varepsilon}\sigma_{2\varepsilon}\sigma_{3\varepsilon}$.

We rewrite (4) as

$$L(\mathcal{Y}^T;\gamma) = (2\pi)^{-\frac{4T}{2}} \prod_{t=1}^{T} \iint p(\xi_t)|dy(\xi_t)|^{-1} p(W_1^t, S_0|\mathcal{Y}^{t-1};\gamma) dW_1^t dS_0.$$

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 $\left\{ \left\{ \xi_t^i \right\}_{i=1}^{N} \right\}_{t=1}^{T}$ implied by them. Therefore, the likelihood function is approximated by

$$L(\mathcal{Y}^T;\gamma) \simeq (2\pi)^{-\frac{4T}{2}} \prod_{t=1}^{T} \sum_{i=1}^{N} p(\xi_t^i)|dy(\xi_t^i)|^{-1}. \tag{21}$$

Equation (21) 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 ξ_t come directly from the output of the Riccati equation, while in our filter they come from the output of the simulation.

3.6. *Findings*

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 data on the relative price of investment. The technical appendix describes in further detail the construction of the different time series for our four observables. Here we only emphasize that, to make the observed series compatible with the model, we compute both real output and real gross investment in consumption units. For 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 the National Income and Product Accounts (NIPA). Since the NIPA investment deflators are poorly measured, we use the investment deflator constructed by Fisher (2006).

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.

3.6.1. Point estimates. Before the estimation, we constrain some parameter values. First, we force the value of β to be less than 1 (so the utility of the consumer is well defined). Second,

TABLE 1

Maximum likelihood estimates

Parameter	Point estimate	S.E. ($\times 10^{-2}$)
ρ	0·9982	0·313
β	0·9992	0·167
ψ	1·9564	3·865
θ	0·0076	0·488
ζ	$0{\cdot}691 \times 10^{-9}$	0·006
α	0·370	2·188
τ_a	0·880	26·727
τ_υ	0·243	19·782
τ_d	0·144	11·876
$\overline{\sigma}_a$	0·005	0·080
$\overline{\sigma}_\upsilon$	0·042	0·872
$\overline{\sigma}_d$	0·014	0·876
λ_a	0·041	14·968
λ_υ	0·998	9·985
λ_d	0·996	10·220
$\sigma_{1\varepsilon}$	$2{\cdot}457 \times 10^{-4}$	0·008
$\sigma_{2\varepsilon}$	$2{\cdot}368 \times 10^{-4}$	0·003
$\sigma_{3\varepsilon}$	$4{\cdot}877 \times 10^{-4}$	0·013

the parameters $\{\tau_a, \tau_\upsilon, \tau_d, \overline{\sigma}_a, \overline{\sigma}_\upsilon, \overline{\sigma}_d\}$ determining S.D. must be positive. Third, the autoregressive coefficients $\{\rho, \lambda_a, \lambda_\upsilon, \lambda_d\}$ will be between 0 and 1 to maintain stationarity. Fourth, we fix the depreciation factor δ to 0·0154 to match the capital-output ratio of the U.S. economy. We do so because the parameter is not well identified: the likelihood function is (nearly) flat along that dimension. This is a common finding in the literature (see Del Negro *et al.*, 2007 and Justiniano and Primiceri, 2005).

Table 1 reports the MLE for the other 18 parameters of the model and their S.E. 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 labour supply, ψ, is closely estimated around 1·956 to capture the level of hours worked *per capita*. The share of capital, α, is 0·37, close to the estimates in the literature. The two drifts in the technology processes, θ and ζ, imply an average growth rate of 0·448% quarter to quarter, which corresponds to the historical long-run growth rate of U.S. real output *per capita* since the Civil War, an annual rate of 1·8%. Nearly all that growth 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 estimates of $\{\tau_a, \tau_\upsilon, \tau_d, \overline{\sigma}_a, \overline{\sigma}_\upsilon, \overline{\sigma}_d\}$ are relatively imprecise. With quarterly data, it is difficult to get accurate estimates of volatility parameters. The autoregressive components of the stochastic volatility of the shocks $\{\lambda_a, \lambda_\upsilon, \lambda_d\}$ are also difficult to estimate, ranging from a low number, λ_a, to a persistence close to 1, 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 2.6, we compare $\overline{\mathbb{Y}}^T(V^T = 0; \widehat{\gamma}_{\text{MLE}})$ and $\mathcal{Y}^T$. In the four panels of Figure 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.

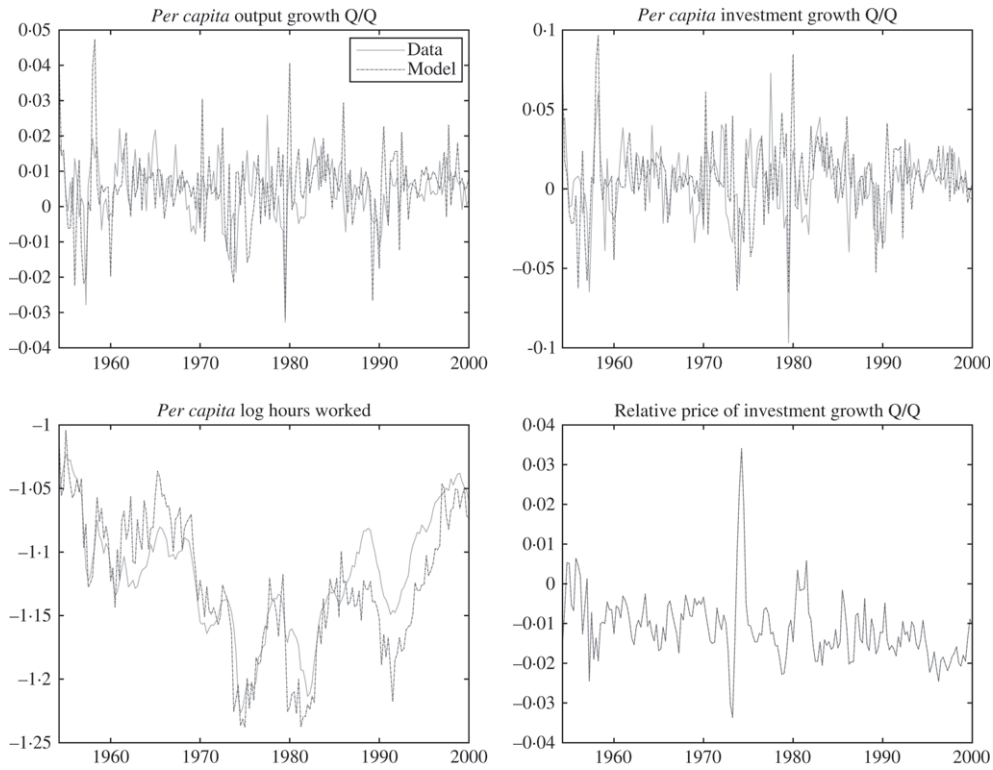

FIGURE 1
Model vs. data

The top left panel displays how the model captures much of the dynamics of the real output *per capita* growth, including the recessions of the 1970's and the expansions of the 1980's and 1990's. The model has a correlation of 35% to the data, and the average predicted output accounts for 63% of the S.D. of 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. The model average predicted real gross investment *per capita* growth has a correlation of 40% to the data, and it accounts for 80% of the observed S.D. 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 85% and the model accounts for 76% of the observed S.D. of hours worked *per capita*. The bottom right panel analyses 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 1 is that the model is fairly successful at capturing aggregate dynamics.

3.6.2. Evolution of the volatility. In Figure 2, we plot the mean of the smoothed paths of the preference level, d_t. This level closely tracks hours worked. Figure 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. In the left column of Figure 3, we plot the smoothed paths of the three normalized shocks of our model (neutral technology shock, investment-specific technology shock, and preference shocks). In the right column, we graph the smoothed volatility

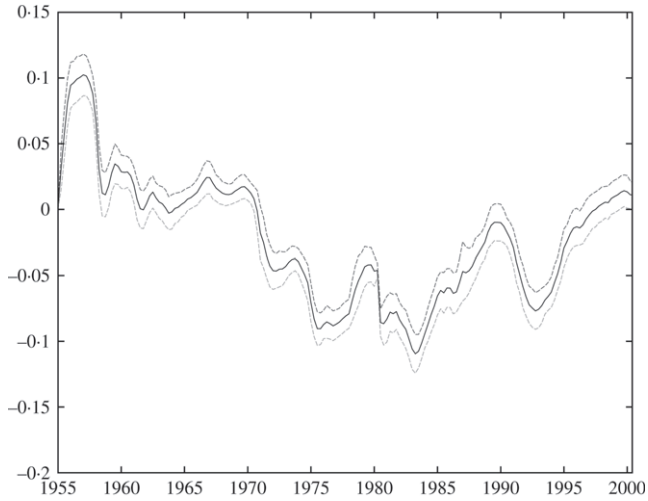

FIGURE 2
Mean (± S.D.) of smoothed d_t

for the corresponding shock. In all graphs, we also plot the 1 S.D. bands around the mean of the smoothed paths. The bands illustrate that our smoothed paths are tightly estimated. All the smoothed paths reported in this section are computed at the MLE.

The first row of Figure 3 shows that little action comes from the evolution of the volatility of the neutral technology shock. This shock has a flat volatility over the sample (note the scale of the graph). More interesting are the second and third rows. The left panel of the second row shows the evolution of the investment-specific technological shocks. The right panel of the second row shows the basic patter of volatility: two periods of low volatility interrupted by a period of high volatility. Volatility increased slightly during the late 1950's, stayed constant during the 1960's, and then notably rose during the 1970's. After its peak in 1981, it decreased during the 1980's and stayed stable again during the 1990's.

The third row documents negative preference shocks in the late 1970's and early 1980's and large positive shocks in the late 1980's. 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. With respect to volatility, it fell during the 1960's. Then, during the 1970's, volatility went back up to its level at the beginning of the sample, where it roughly stayed during the 1980's. The 1990's were times again of falling volatility.

How did the time-varying volatility of shocks affect the volatility of the aggregate time series? Figure 4 plots the log of the instantaneous S.D. 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 S.D. 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 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 mentioned in Section 2.6 and compute the instantaneous S.D. of the observables for each draw. We report the mean of the instantaneous S.D. of the observables. We also plot the 1 S.D. bands around it.

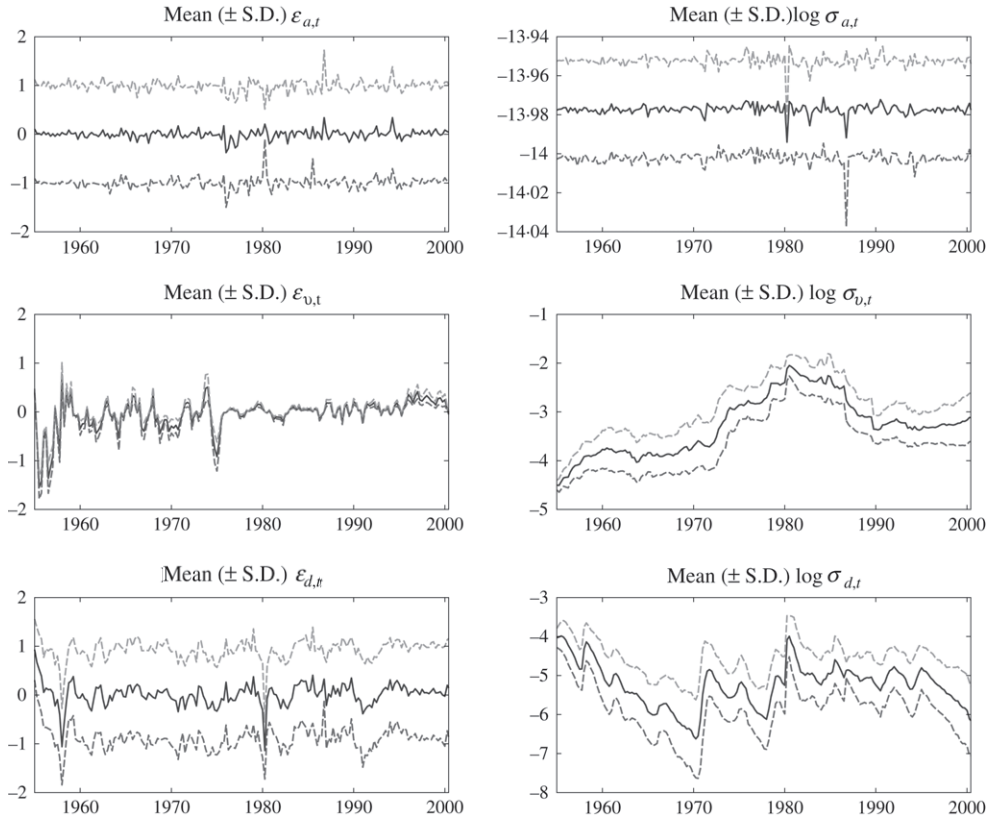

FIGURE 3

Mean (± S.D.) of smoothed shocks and volatilities

All four panels of Figure 4 show the same basic pattern: low and falling volatility in the 1960's (except for the relative price of investment where volatility is low but not falling), a large increase in volatility during the 1970's, with a peak in the early 1980's, and a fall in volatility until the end of the sample.

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 1980's, as defended by a part of the literature, but more of a gradual change. Our findings, in comparison, coincide more with the views of Blanchard and Simon (2001) and Sims and Zha (2006).

3.6.3. 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 non-linear models, it is difficult to perform a decomposition of the variance because the random shocks hitting the model enter in multiplicative ways. Instead, we perform two counterfactual experiments.

In the first experiment, we fix the volatility of one shock at its level at the beginning of the sample in 1955, and we let the volatility of the other two shocks evolve in the same way as in our smoothed estimates. We plot our results in Figure 5, where the reader can see 12 panels in four rows and three columns. Each row corresponds to the evolution of volatility of each of our four observables: real output *per capita* growth (first row), real gross investment *per*

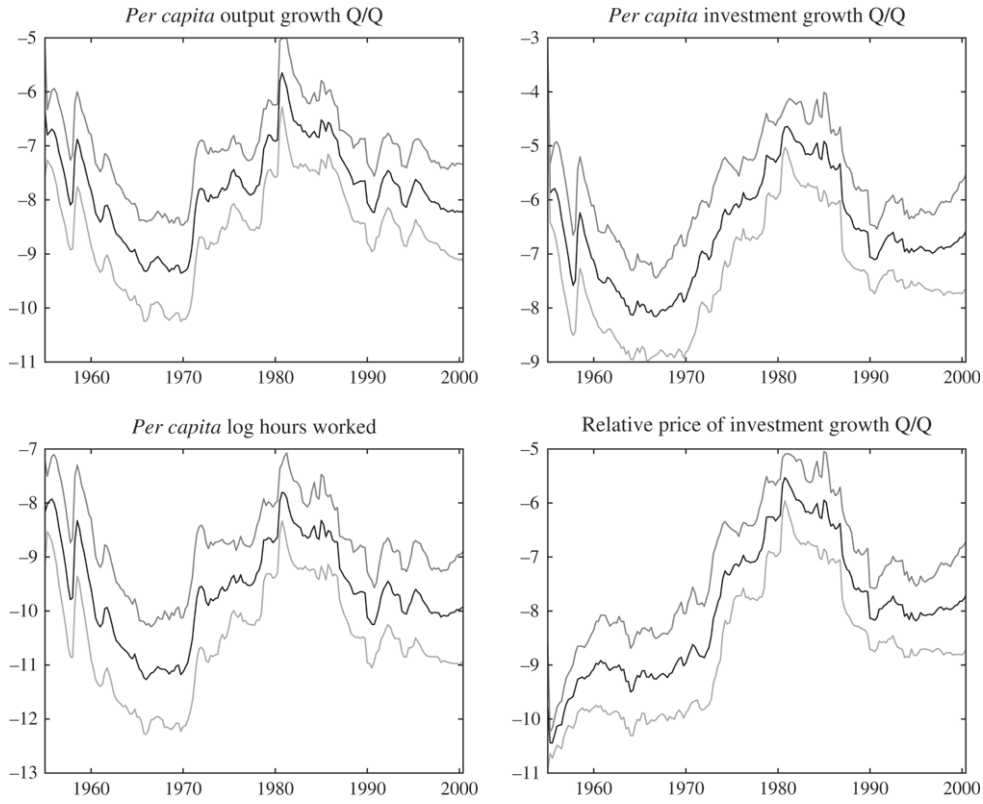

FIGURE 4
Log of the instantaneous S.D.

capita growth (second row), hours worked *per capita* (third row), and relative price of investment (fourth row). Each column corresponds to each of the three different possibilities for fixing one shock: fixing the volatility of the neutral technological shock (first column), fixing the volatility of the investment-specific technological shocks (second column), and fixing the volatility of the preference shock (third column). In each panel, we plot the estimated instantaneous variance of the observable (continuous line) and the counterfactual variance (discontinuous line) implied by keeping the volatility of the shock at its level in 1955.

Figure 5 illustrates that the decrease in the volatility of the shock to preferences explains the reduction in volatility experienced during the 1960's. Neither changes in the volatility of the neutral technological shock nor changes in the volatility of the investment-specific technological shock account for the reduction in volatility. Indeed, without the change in the variance of the investment-specific technological shock, the economy would have been less volatile.

The second counterfactual experiment repeats the first experiment, except that now we fix the volatilities at their values in the first quarter of 1981, the period where instantaneous volatility of output was at its highest. We plot our findings in Figure 6, where we follow the same ordering convention as in Figure 5. Now both the investment-specific technological shock and the preference shock contribute to the reduction in observed volatility. More important (although not reflected in the figure), it is the interaction between the two of them that contributes the most to the reduction in volatility. Again, the change in the volatility neutral technological shock is irrelevant.

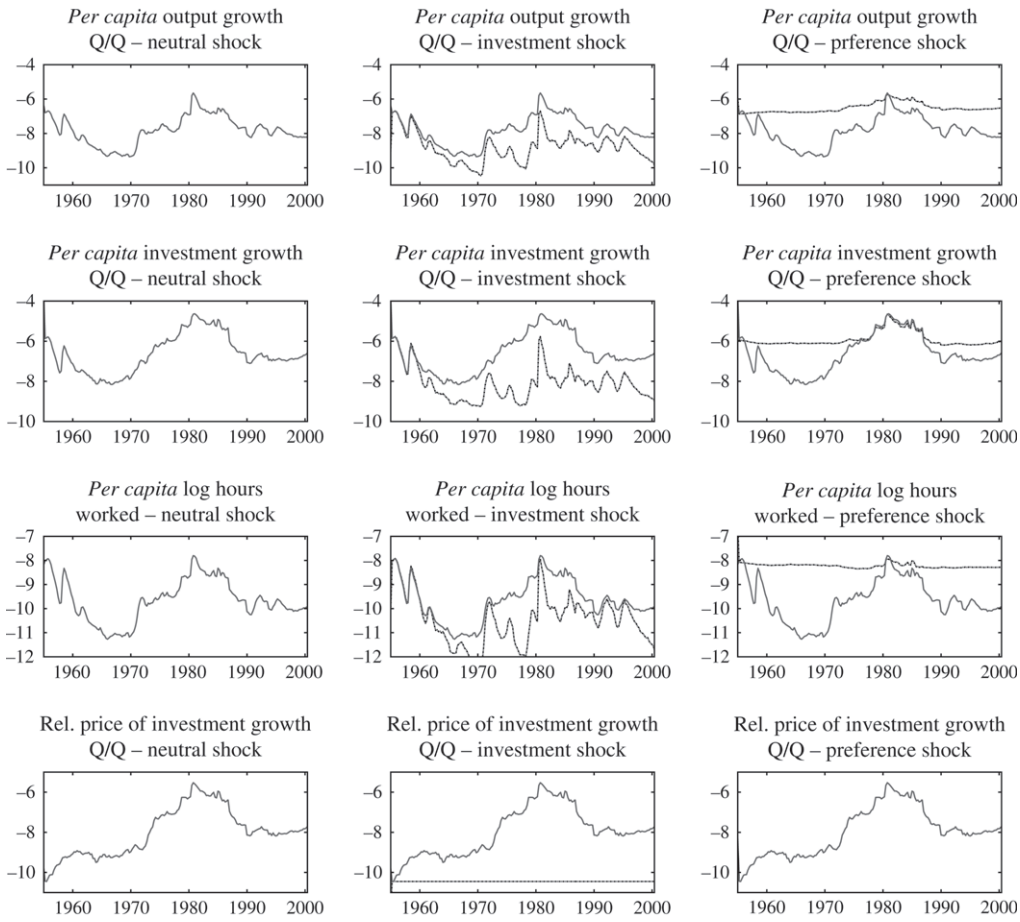

FIGURE 5

Which shocks were responsible for the volatility decrease of the 1960's?

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

To answer these questions, in addition to our benchmark version, we estimated two additional versions of the model: one with a linear solution, in which by construction the stochastic volatility components drop, and one with a quadratic solution but without stochastic volatility.

In Table 2, we report the log likelihoods of each of the three versions of the model at their MLEs. The comparison of these three log likelihoods shows how both the quadratic component and the stochastic volatility improve the performance of the model. The quadratic component adds 116 log points and the stochastic volatility 118 log points. To formally evaluate these numbers, we can undertake likelihood ratio-type tests. For example, since the benchmark model nests the quadratic model without stochastic volatility by setting $\{\tau_a, \tau_v, \tau_d\}$ to 0, we have that

$$LR_{\text{bench,quadratic}} = -2(\log L_{\text{bench}}(\mathcal{Y}^T; \widehat{\gamma}) - \log L_{\text{quad}}(\mathcal{Y}^T; \widehat{\gamma})) \xrightarrow{d} \chi(3).$$

The value of the test rejects the equality of both models at any conventional significance level.

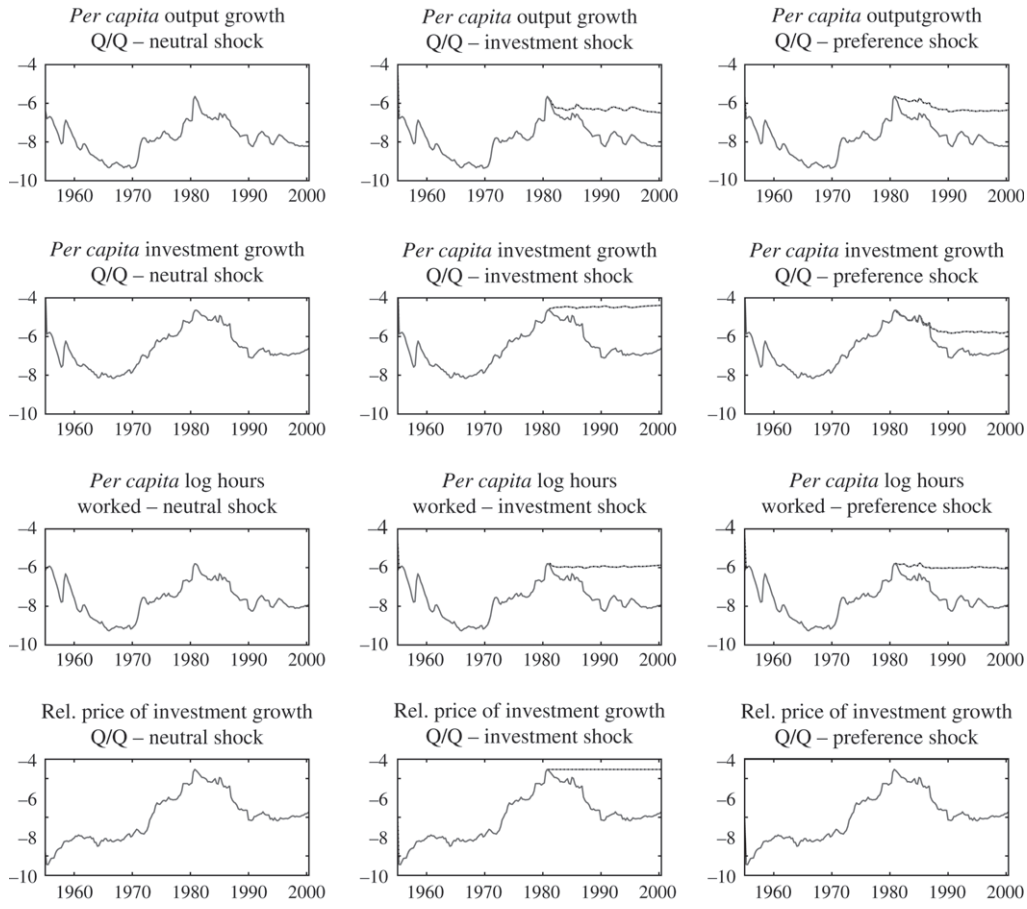

FIGURE 6

Which shocks were responsible for the volatility increase of the 1970's?

TABLE 2

Versions of the model

Version	Log likelihood
Linear	1313·748
Quadratic, no S.V.	1429·637
Benchmark	1547·285

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. Stochastic volatility adds variance to the shocks during the 1970's, when this is most needed, and reduces it during the 1960's and 1990's.

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 non-linearities may be of higher order. Consequently, our results show that, even in those challenging circumstances, the non-linear estimation pays off.

4. 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 5160 EMT64 at 3·00 GHz with 16 GB of RAM, each evaluation of the likelihood with 80,000 particles takes around 12 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?

With respect to the first question, we show in the previous section that the particle filter improves inference with respect to the Kalman filter. Similar results are documented by Amisano and Tristani (2005), Fernández-Villaverde and Rubio-Ramírez (2005), An (2006), and An and Schorfheide (2007). 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 Prediction and Filtering Steps. If we decompose the 12 seconds that each evaluation of the likelihood requires, we discover that the Prediction and Filtering Steps take over 10 seconds, while the solution of the model takes less than 0·1 second and the Sampling Step around 1 second. In an economy with even more state variables than ours (we already have eight state variables!), we will only increase the computational time of the solution, while the other steps may take roughly the same time. The availability of fast solution methods, like perturbation, implies that we can compute the non-linear 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 15 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 a good number for that purpose. However, when combined with simulated annealing, we encountered some numerical instabilities, especially when computing the Hessian. A brute force solution for this problem is to increase the number of particles when computing the Hessian, for which we used one million particles. A possibly more efficient solution could be to use more refined algorithms, such as the auxiliary particle filter of Pitt and Shephard (1999). Second, we checked for a depletion of the sample problem. For example, Arulampalam *et al.* (2002, p. 179), propose to evaluate the "effective sample size" N_{eff}:

$$N_{\text{eff}} = \frac{N}{1 + \text{var}\left(\frac{p(s_t^i|\mathcal{Y}^t;\gamma)}{p(s_t|s_{t-1}^i;\gamma)}\right)},$$

where N is the number of particles, $p(s_t^i|\mathcal{Y}^t;\gamma)$ is the probability of the state s_t^i given observations y^t, and $p(s_t|s_{t-1}^i;\gamma)$ is the proposal density (note that in this definition we follow the more conventional notation in terms of states to facilitate comparison with Arulampalam *et al.*, 2002).

Given that "effective sample size" cannot be evaluated directly, we can obtain an estimate $\widehat{N}_{\text{eff}}$:

$$\widehat{N}_{\text{eff}} = \frac{1}{\sum_{i=1}^{N} (q_t^i)^2},$$

where q_t^i is the normalized weight of each particle.

We computed the effective sample size to check that it was high and stable over the sample. 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 Intel Visual Fortran 9.1 to run on Windows-based PCs. All of the code is available upon request.

5. CONCLUSIONS

We have presented a general purpose and asymptotically efficient algorithm to perform likelihood-based inference in non-linear and/or non-normal dynamic macroeconomic models. We have shown how to undertake parameter estimation and model comparison, from either 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 of the presence of stochastic volatility in U.S. data and of the role of changes in the volatility of preference shocks as a main force behind the variation in the volatility of U.S. output growth over the last 50 years.

Future research on particle filtering could try to extend Pitt's (2002) results to find estimates of the likelihood function that are continuous with respect to the parameters, to reduce the numerical variance of the procedure, and to increase the speed of the algorithm. Also, it would be useful to have more comparisons of particle filtering with alternative procedures and further explorations of the advantages and disadvantages of the different types of sequential Monte Carlos. With all of these new developments, we should be able to estimate richer and more interesting macroeconomic models.

Acknowledgements. We thank Manuel Arellano, James Heckman, John Geweke, Marco Del Negro, Will Roberds, Eric Renault, Barbara Rossi, Stephanie Schmitt-Grohé, Enrique Sentana, Frank Schorfheide, Chris Sims, Ellis Tallman, Tao Zha, several referees, and participants at several seminars for comments. Jonas Fisher provided us with his investment deflator. Mark Fisher's help with coding was priceless. 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. Finally, we also thank the NSF for financial support.

REFERENCES

AMISANO, G. and TRISTANI, O. (2005), "Euro Area Inflation Persistence in an Estimated Nonlinear DSGE Model" (Mimeo, ECB).

AN, S. (2006), "Bayesian Estimation of DSGE Models: Lessons from Second-Order Approximations" (Mimeo, University of Pennsylvania).

AN, S. and SCHORFHEIDE, F. (2007), "Bayesian Analysis of DSGE Models", *Econometric Reviews* (forthcoming).

ARULAMPALAM, A. S., MASKELL, S., GORDON, N. and CLAPP, T. (2002), "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking", *IEEE Transactions on Signal Processing*, **50**, 174–188.

ARUOBA, S. B., FERNÁNDEZ-VILLAVERDE, J. and RUBIO-RAMÍREZ, J. (2006), "Comparing Solution Methods for Dynamic Equilibrium Economies", *Journal of Economic Dynamics and Control*, **30**, 2477–2508.

BLANCHARD, O. J. and SIMON, J. (2001), "The Long and Large Decline in U.S. Output Volatility", *Brookings Papers on Economic Activity*, 2001-1, 187–207.

CHIB, S. and GREENBERG, E. (1995), "Understanding the Metropolis-Hastings Algorithm", *American Statistician*, **49**, 327–335.

COOLEY, T. F. (ed.) (1995) *Frontiers of Business Cycle Research* (Princeton, NJ: Princeton University Press).
DEL MORAL, P. and JACOD, J. (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* (Providence, RI: The Fields Institute Communications, American Mathematical Society).
DEL NEGRO, M., SCHORFHEIDE, F., SMETS, F. and WOUTERS, R. A. F. (2007), "On the Fit and Forecasting Performance of New Keynesian Models", *Journal of Business and Economic Statistics* (forthcoming).
DOUCET, A., DE FREITAS, N. and GORDON, N. (2001) *Sequential Monte Carlo Methods in Practice* (New York, NY: Springer-Verlag).
FERMANIAN, J. D. and SALANIÉ, B. (2004), "A Nonparametric Simulated Maximum Likelihood Estimation Method", *Econometric Theory*, **20**, 701–734.
FERNÁNDEZ-VILLAVERDE, J. and RUBIO-RAMIŕEZ, J. (2005), "Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood", *Journal of Applied Econometrics*, **20**, 891–910.
FERNÁNDEZ-VILLAVERDE, J., RUBIO-RAMÍREZ, J. and SANTOS, M. S. (2006), "Convergence Properties of the Likelihood of Computed Dynamic Models", *Econometrica*, **74**, 93–119.
FIORENTINI, G., SENTANTA, E. and SHEPHARD, N. (2004), "Likelihood Estimation of Latent Generalised ARCH Structures", *Econometrica*, **72**, 1481–1517.
FISHER, J. (2006), "The Dynamic Effects of Neutral and Investment-Specific Technology Shocks", *Journal of Political Economy*, **114**, 413–451.
GALLANT, A. R. and TAUCHEN, G. (1996), "Which Moments to Match?" *Econometric Theory*, **12**, 657–681.
GEWEKE, J. (1989), "Bayesian Inference in Econometric Models Using Monte Carlo Integration", *Econometrica*, **24**, 1037–1399.
GODSILL, S. J., DOUCET, A. and WEST, M. (2004), "Monte Carlo Smoothing for Nonlinear Time Series", *Journal of the American Statistical Association*, **99**, 156–168.
GREENWOOD, J., HERKOWITZ, Z. and KRUSELL, P. (1997), "Long-Run Implications of Investment-Specific Technological Change", *American Economic Review*, **87**, 342–362.
GREENWOOD, J., HERKOWITZ, Z. and KRUSELL, P. (2000), "The Role of Investment-Specific Technological Change in the Business Cycle", *European Economic Review*, **44**, 91–115.
HARVEY, A. C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter* (Cambridge, UK: Cambridge University Press).
JUSTINIANO, A. and PRIMICERI, G. E. (2005), "The Time Varying Volatility of Macroeconomic Fluctuations" (Mimeo, Northwestern).
KIM, C. and NELSON, C. R. (1998), "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle", *Review of Economics and Statistics*, **81**, 608–616.
KIM, S., SHEPHARD, N. and CHIB, S. (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models", *Review of Economic Studies*, **65**, 361–393.
KING, T. B. (2006), "Dynamic Equilibrium Models with Time-Varying Structural Parameters" (Mimeo, Washington University).
KITAGAWA, G. (1996), "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models", *Journal of Computational and Graphical Statistics*, **5**, 1–25.
KOOPMAN, S. J. and SHEPHARD, N. (2003), "Testing the Assumptions Behind the Use of Importance Sampling" (Mimeo, Oxford University).
KÜNSCH, H. R. (2005), "Recursive Monte Carlo Filters: Algorithms and Theoretical Analysis", *Annals of Statistics*, **33**, 1983–2021.
LAROQUE, G. and SALANIÉ, B. (1989), "Estimation of Multimarket Fix-Price Models: An Application of Pseudo-Maximum Likelihood Methods", *Econometrica*, **57**, 831–860.
LAROQUE, G. and SALANIÉ, B. (1993), "Simulation-Based Estimation of Models with Lagged Latent Variables", *Journal of Applied Econometrics*, **8**, S119–S133.
MCCONNELL, M. M. and PÉREZ-QUIRÓS, G. (2000), "Output Fluctuations in the United States: What Has Changed Since the Early 1980's?" *American Economic Review*, **90**, 1464–1476.
MCFADDEN, D. L. (1989) "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration", *Econometrica*, **57**, 995–1026.
PAKES, A. and POLLARD, D. (1989), "Simulation and the Asymptotics of Optimization Estimators", *Econometrica*, **57**, 1027–1057.
PITT, M. K. (2002) "Smooth Particle Filters for Likelihood Evaluation and Maximization" (Mimeo, University of Warwick).
PITT, M. K. and SHEPHARD, N. (1999), "Filtering via Simulation: Auxiliary Particle Filters", *Journal of the American Statistical Association*, **94**, 590–599.
ROBERTS, G., GELMAN, A. and GILKS, W. (1997), "Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms", *Annals of Applied Probability*, **7**, 110–120.
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* (Oxford: Oxford University Press) 395–402.
SANTOS, M. S. and PERALTA-ALVA, A. (2005), "Accuracy of Simulations for Stochastic Dynamic Models", *Econometrica*, **73**, 1939–1976.

SHEPHARD, N. (2005) *Stochastic Volatility, Selected Readings* (Oxford: Oxford University Press).
SIMS, C. A. and ZHA, T. (2006), "Were There Regime Switches in U.S. Monetary Policy?" *American Economic Review*, **96**, 54–81.
STENFLO, O. (2001), "Ergodic Theorems for Markov Chains Represented by Iterated Function Systems", *Bulletin of the Polish Academy of Sciences: Mathematics*, **49**, 27–43.
STOCK, J. H. and WATSON, M. W. (2002), "Has the Business Cycle Changed, and Why?" *NBER Macroeconomics Annual*, **17**, 159–218.
STOKEY, N. L., LUCAS, R. E. and PRESCOTT, E. C. (1989) *Recursive Methods in Economic Dynamics* (Cambridge, MA: Harvard University Press).
WINSCHEL, V. (2005) "Solving, Estimating, Selecting Nonlinear Dynamic Economic Models without the Curse of Dimensionality" (Mimeo, University of Mannheim).