# Technical Appendix to:

# "Estimating Macroeconomic Models: A Likelihood Approach"

Jesús Fernández-Villaverde

University of Pennsylvania, NBER, and CEPR

Juan F. Rubio-Ramírez

Duke University and Federal Reserve Bank of Atlanta

November 20, 2006

**Abstract**

## 1. Introduction

This technical appendix provides further details on three aspects of the paper. First, we discuss alternative procedures to perform likelihood inference. Second, we discuss smoothing. Finally, we discuss the construction of our data.

## 2. Comparison with Alternative Schemes

The particle filter is not the only procedure for evaluating the likelihood of the data implied by nonlinear and/or non-normal dynamic macroeconomic models. Our discussion in the paper highlighted how computing the likelihood amounts to solving a nonlinear filtering problem, i.e., generating estimates of the values of $W_1^t$ and $S_0$ conditional on $\mathcal{Y}^{t-1}$ to evaluate the integral in the likelihood function. Since this task is of interest in different fields, several alternative schemes have been proposed to handle this problem.

A first line of research has been in deterministic filtering. Historically, the first procedure in this line was the Extended Kalman filter (Jazwinski, 1973), which linearizes the transition and measurement equations and uses the Kalman filter to estimate for the states and the shocks to the system. This approach suffers from the approximation error incurred by linearization and from the inaccuracy incurred by the fact that the posterior estimates of the states are non-normal. As the sample size grows, those problems accumulate and the filter diverges. Even refinements such as the Iterated Extended Kalman filter, the quadratic Kalman filter (which carries the second order term of the transition and measurement equations), and the unscented Kalman filter (which considers a set of points instead of just the conditional mean of the state, see Julier and Uhlmann, 1996) cannot fully solve these problems.

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

A third alternative in deterministic filtering is grid filters, which use quadrature integration to compute the different integrals of the problem (Bucy and Senne, 1971). Unfortunately, grid filters are difficult to implement, since they require a constant readjustment to small changes in the model or its parameter values. Also, they are too computationally expensive to be of any practical benefit beyond very low dimensions. A final shortcoming of grid filters is that the grid points are fixed ex ante and the results are very dependent on that choice.

In comparison, a particle filter can be interpreted as a grid filter where the grid points are chosen endogenously over time based on their ability to account for the data.

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

A second strategy is to think of the functions $f$ and $g$ as a change in variables of the innovations to the model and use the jacobian of the transformation to evaluate the likelihood of the observables (Miranda and Rui, 1997). In general, however, this approach is cumbersome and problematic to implement.

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

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

## 3. Smoothing

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

First, we analyze how to assess the fit of the model. Given a value for $\gamma$, 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, $\mathbb{Y}^T\left(S^T, V^T; \gamma\right)$. Thus, for any $\gamma$, we compute the mean of the observables implied by the model and the realization of observables, $\mathcal{Y}^T$:

$$\overline{\mathbb{Y}}^T\left(V^T; \gamma\right) = \int \mathbb{Y}^T\left(S^T, V^T; \gamma\right) p\left(S^T | \mathcal{Y}^T; \gamma\right) dS^T. \tag{1}$$

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

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

$$\overline{\mathbb{Y}}^T_{S_k^{t:T}=S_{k,t}}\left(V^T; \gamma\right) = \int \mathbb{Y}^T\left(S^T_{-k}, S_k^{t:T} = S_{k,t}, V^T; \gamma\right) p\left(S^T | \mathcal{Y}^T; \gamma\right) dS^T, \tag{2}$$

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

The two examples share a common theme. To compute integrals like (1) or (2), which will appear in our application below, we need to draw from $p\left(S^T | \mathcal{Y}^T; \gamma\right)$. To see this, let $\{s^{t,i}\}_{i=1}^N$ be a draw from $p\left(S^T | \mathcal{Y}^T; \gamma\right)$. Then (1) and (2) are approximated by:

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

and

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

Hence, the problem of computing integrals like (1) and (2) is equivalent to the problem of drawing from $p\left(S^T | \mathcal{Y}^T; \gamma\right)$. We now propose a smoothing algorithm to accomplish this objective.

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

$$p\left(S^{t:T}|\mathcal{Y}^T;\gamma\right) = p\left(S_t|S^{t+1:T},\mathcal{Y}^T;\gamma\right)p\left(S^{t+1:T}|\mathcal{Y}^T;\gamma\right) \tag{3}$$

for all $t$, where $S^{t+1:T} = \{S_m\}_{m=t+1}^T$ is the sequence of states from period $t+1$ to period $T$. Therefore, from (3) it should be clear that to draw from $p\left(S^{t:T}|\mathcal{Y}^T;\gamma\right)$, we need to draw from $p\left(S_t|s^{t+1:T},\mathcal{Y}^T;\gamma\right)$, where $s^{t+1:T}\left(=\{s_m\}_{m=t+1}^T\right)$ is a draw from $p\left(S^{t+1:T}|\mathcal{Y}^T;\gamma\right)$. We describe a recursive procedure to do so.

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

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

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

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

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

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

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

The crucial step in proposition 1 is to draw from $p\left(S_t|\mathcal{Y}^t;\gamma\right)$. We accomplish this with the output from the particle filter. First, we know that $p\left(S_t|\mathcal{Y}^t;\gamma\right) = p\left(W_2^t,W_1^t,S_0|\mathcal{Y}^t;\gamma\right)$. Second, we also know that

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

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

From proposition 4 in the paper and the explanation above, the smoother algorithm is:

---

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

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

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

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

---

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

## 4. Construction of Data

As we mention in the main text, to make the observed series compatible with the model, we compute both real output and real gross investment in consumption units. As the relative price of investment we use the ratio of an investment deflator and a deflator for consumption. The consumption deflator is constructed from the deflators of nondurable goods and services reported in the NIPA. Since the NIPA investment deflators are poorly measured, we use the investment deflator constructed by Fisher (2006). For the real output per capita series, we first define nominal output as nominal consumption plus nominal gross investment. We define nominal consumption as the sum of personal consumption expenditures on nondurable goods and services, national defense consumption expenditures, federal nondefense consumption expenditures, and state and local government consumption expenditures. We define nominal gross investment as the sum of personal consumption expenditures on durable goods, national defense gross investment, federal government nondefense gross investment, state and local government gross investment, private nonresidential fixed investment, and private residential fixed investment. Per capita nominal output is defined as the ratio between our nominal output series and the civilian noninstitutional population between 16 and 65. Since we need to measure real output per capita in consumption units, we deflate the series by the consumption deflator. For the real gross investment per capita series, we divide our above mentioned nominal gross investment series by the civilian noninstitutional population between 16 and 65 and the consumption deflator. Finally, the hours worked per capita series is constructed with the index of total number of hours worked in the business sector and the civilian noninstitutional population between 16 and 65. Since our model implies that hours worked per capita are between 0 and 1, we normalize the observed series of hours worked per capita such that it is, on average, 0.33.

## References


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

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

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

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

[5] Fisher, J. (2006). "The Dynamic Effects of Neutral and Investment-Specific Technology Shocks." *Journal of Political Economy* 114, 413-451.

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

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

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

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

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

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

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

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

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