Econometrica, Vol. 74, No. 1 (January, 2006), 93–119

CONVERGENCE PROPERTIES OF THE LIKELIHOOD OF COMPUTED DYNAMIC MODELS

BY JESÚS FERNÁNDEZ-VILLAVERDE, JUAN F. RUBIO-RAMÍREZ, AND MANUEL S. SANTOS[1]

This paper studies the econometrics of computed dynamic models. Since these models generally lack a closed-form solution, their policy functions are approximated by numerical methods. Hence, the researcher can only evaluate an approximated likelihood associated with the approximated policy function rather than the exact likelihood implied by the exact policy function. What are the consequences for inference of the use of approximated likelihoods? First, we find conditions under which, as the approximated policy function converges to the exact policy, the approximated likelihood also converges to the exact likelihood. Second, we show that second order approximation errors in the policy function, which almost always are ignored by researchers, have first order effects on the likelihood function. Third, we discuss convergence of Bayesian and classical estimates. Finally, we propose to use a likelihood ratio test as a diagnostic device for problems derived from the use of approximated likelihoods.

KEYWORDS: Dynamic economic models, convergence, computation.

1. INTRODUCTION

THIS PAPER STUDIES THE FOLLOWING PROBLEM. Most dynamic economic models do not have a closed-form solution. Instead, the solution is approximated by numerical methods. Hence, when a researcher builds the likelihood function of the model given some data, she is not evaluating the exact likelihood, but only an approximated likelihood given her approximated solution to the model. What are the effects on statistical inference of using an approximated likelihood instead of the exact likelihood function?

In the last 20 years, there has been a remarkable increase in the use of dynamic models with approximated likelihoods and simulation techniques in econometrics. We can find examples in labor economics, industrial organization, health economics, demographics, game theory, development, public finance, auction theory, macroeconomics, open economy, finance, and other fields. Without being exhaustive, we can cite Flinn and Heckman (1982), Wolpin (1984), Pakes (1986), Rust (1987), Sargent (1989), Rosenzweig and Wolpin (1993), Daula and Moffitt (1995), Keane and Wolpin (1997), Rust and Phelan (1997), Gilleskie (1998), Keane and Moffitt (1998), DeJong, Ingram, and Whiteman (2000), Schorfheide (2000), Jofre-Bonet and Pesendorfer (2003), Smets and Wouters (2003), Crawford and Shum (2005), and Rabanal

[1]We thank a co-editor, three referees, Jim Nason, Tom Sargent, Frank Schorfheide, Tao Zha, participants at several seminars, and especially Lee Ohanian for useful comments. Jesús Fernández-Villaverde thanks the NSF for financial support under project SES-0338997. Any views expressed herein are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Atlanta or of the Federal Reserve System.

and Rubio-Ramírez (2005) among dozens of others. Moreover, a growing number of policy-making institutions (the European Central Bank, the Federal Reserve Board, the Riksbank, the International Monetary Fund, the Bank of Canada, the Bank of Spain, and the Bank of Italy) are formulating and estimating dynamic models with approximated likelihoods for policy analysis.

The standard practice of researchers when it comes to using likelihood methods is to conduct inference as if they had the exact solution of the model and to ignore the effects of approximation errors on the solution of the model. Surprisingly enough, hardly anything is known about the implications of these approximation errors for likelihood analysis. Consequently, we understand very little about the possible mistakes that researchers face when conducting inference. How different are the approximated and the exact likelihood functions? Does the approximated likelihood function converge to the exact likelihood as the approximated policy function converges to the exact policy function? If it does, at what speed? How do approximated policy functions affect parameter inference and model comparison?

Our first result is to find technical conditions under which, if the approximated policy function converges to the exact policy function in the sup norm for given parameter values, the approximated likelihood function also converges to the exact likelihood. Why is this result important? Because it is easy to build examples where the violation of our technical conditions implies that the sequence of approximated policy functions converges to the exact policy function but the sequence of approximated likelihoods does not converge. The working paper version of this article (Fernández-Villaverde, Rubio-Ramírez, and Santos (2005)) presents a simple example of nonconvergence. This example illustrates that we cannot generally assume the convergence of the approximated likelihood, and hence it underscores the need to derive conditions under which this convergence is guaranteed.

Our second finding is that *second order approximation errors in the policy function*, which almost always are ignored by researchers, have *first order effects on the likelihood function*. We show that the approximated likelihood function converges at the same rate as the approximated policy function. However, the error in the approximated likelihood function gets compounded with the size of the sample. Therefore, period by period, small errors in the policy function accumulate at the same rate at which the sample size grows. Similarly, third order approximation errors in the policy function will have second order effects on the likelihood function and so on.

This result warns us that there could be strong biases from statistical inference performed with the approximated model instead of the exact one. We present a simple application that shows that this bias is quantitatively relevant in real-life models. Our application also illustrates how to diagnose and correct the biases in inference using a likelihood ratio test. Our theoretical and empirical findings thus have *dramatic* and *unexpected* consequences for the many researchers engaged in likelihood-based inference of approximated models.

Our third result concerns the convergence of estimates. We show that the convergence of Bayesian estimators comes directly from our first result, the pointwise convergence of the likelihood. The case of maximum likelihood estimates is more involved. Pointwise convergence of the likelihood does not allow us to swap the arg max and lim operators, but we can mildly impose more stringent conditions to prove the uniform convergence of the approximated likelihood function to the exact likelihood. Uniform convergence implies the convergence of maximum likelihood point estimates.

Our paper is the first systematic analysis of the implications of approximation error on likelihood inference. We build on the recent work by Santos and Peralta-Alva (2005) and Santos (2004), who have derived some pioneering results on the convergence of the moments generated by a numerically approximated model when the computed policy functions converge to the exact ones. Santos and Peralta-Alva have shown that the moments computed under the numerically approximated policy converge to their exact values as the approximation errors of the computed solution go to zero. We extend this research to the study of the convergence properties of approximated likelihood functions. This extension raises a whole new range of issues not previously explored either in economics or statistics. Also, our theoretical results confirm some of the conjectures in Duffie and Singleton (1993) and provide a foundation for the experiments in Keane and Wolpin (1994).

The rest of the paper is organized as follows. Section 2 fixes an environment to discuss the convergence of the likelihood. Our main result concerning convergence is contained in Section 3. Section 4 narrows down the speed of convergence and its relationship to the sample size, and proposes a likelihood ratio rest to diagnose inference problems. Section 5 studies an application that confirms that the results of the paper hold in practice. Section 6 concludes. An Appendix gathers all the proofs of the results in the paper.

2. THE SETTING

The equilibrium law of motion of a large class of dynamic economies can be specified as a state space system of the form (see Stokey, Lucas, and Prescott (1989))

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

$$Y_t = g(S_t, V_t; \gamma). \tag{2}$$

Here S_t is a vector of state variables that characterize the evolution of the system. The vector S_t belongs to the compact set $S \subset \mathbb{R}^l$. Often, we will use the measurable space $(S, \mathcal{S})$, where $\mathcal{S}$ is the Borel *σ-field*. The variables W_t and V_t are independent and identically distributed shocks with compact supports in subsets of some Euclidean space, with bounded and continuous densities; W_t and V_t are independent of each other. The observables in each period are stacked in a vector Y_t. If we have T periods of observations, we let $Y^T \equiv \{Y_t\}_{t=1}^T$ with $Y^0 = \{\emptyset\}$. We assume that Y^T is distributed according to the probability density function $p_0^T(\cdot)$. Finally, γ, which belongs to the compact set $\Upsilon \subset \mathbb{R}^n$, is the vector of structural parameters, i.e., those describing the preferences, technology, and information sets of the economy. To avoid stochastic singularity, we have that $\dim(W_t) + \dim(V_t) \geq \dim(Y_t)$.

Equation (1) is known as the transition equation, since it governs the evolution of states over time. Equation (2) is called the measurement equation because it relates states and observables. Note that in this framework we can accommodate cases in which the dimensionality of the shocks could be zero or where the shocks have more involved stochastic structures. Also, the states might be part of the observables if g is the identity function along some dimension.

To deal with a larger class of models, we partition $\{W_t\}$ into two sequences $\{W_{1,t}\}$ and $\{W_{2,t}\}$, such that $W_t = (W_{1,t}, W_{2,t})$ and $\dim(W_{2,t}) + \dim(V_t) = \dim(Y_t)$. If $\dim(V_t) = \dim(Y_t)$, we set $W_{1,t} = W_t$ $\forall t$, i.e., $\{W_{2,t}\}$ is a zero-dimensional sequence. If $\dim(W_t) + \dim(V_t) = \dim(Y_t)$, we set $W_{2,t} = W_t$ $\forall t$, i.e., $\{W_{1,t}\}$ is a zero-dimensional sequence. Also, let $W_i^T = \{W_{i,t}\}_{t=1}^T$, $V^T = \{V_t\}_{t=1}^T$, and $S^T = \{S_t\}_{t=0}^T$ $\forall T$. Let y^T be a realization of the random variable Y^T. We define $W_i^0 = \{\emptyset\}$ and $y^0 = \{\emptyset\}$.

Finally, we introduce some additional constructs. Let $C(S)$ be the space of all continuous, $\mathcal{S}$-measurable, real-valued functions on S. We endow $C(S)$ with the norm $\|f\| = \max_{s\in S} |f(s)|$ so that this is a Banach space. For a vector-valued function $f = (\ldots, f^i, \ldots)$, we define $\|f\| = \max_i \|f^i\|$. Convergence of a sequence of functions $\{f_j\}$ should be understood in the metric induced by this norm.

Before getting into our analysis, we make the following assumptions.

ASSUMPTION 1: *For all* γ, *functions* $\varphi(\cdot, \cdot; \gamma)$ *and* $g(\cdot, \cdot; \gamma)$ *are continuously differentiable*, *with bounded partial derivatives*.

Assumption 1 arises naturally in a number of economic models. The continuity of $\varphi(\cdot, \cdot; \gamma)$ often follows from primitive conditions of the economic model (see Theorem 4.8 in Stokey, Lucas, and Prescott (1989)) that ensure the continuity and single-valuedness of the agents' policy functions.

Standard arguments show that there exists an invariant distribution of the dynamic model, $\mu^*(S; \gamma)$. In the next assumption, we state the existence of that invariant distribution, $\mu^*(S; \gamma)$, and that it has a density that we can use in our future derivations. With some extra work (see Brin and Kifer (1987) and Peres and Solomyak (1996)), this assumption could be written directly in terms of the policy functions $\varphi(\cdot, \cdot; \gamma)$ and $g(\cdot, \cdot; \gamma)$.

ASSUMPTION 2: *For all* γ, *there exists a unique invariant distribution for* S, $\mu^*(S; \gamma)$, *that has a Radon–Nikodym derivative with respect to the Lebesgue measure*.

Let us also make the following assumption.

ASSUMPTION 3: *For all* γ *and* t, *the system of equations*

$$S_1 = \varphi(S_0, (W_{1,1}, W_{2,1}); \gamma),$$
$$y_m = g(S_m, V_m; \gamma) \quad \textit{for} \quad m = 1, 2, \ldots, t,$$
$$S_m = \varphi(S_{m-1}, (W_{1,m}, W_{2,m}); \gamma) \quad \textit{for} \quad m = 2, 3, \ldots, t$$

has a unique solution, $(v^t(W_1^t, S_0, y^t; \gamma), s^t(W_1^t, S_0, y^t; \gamma), w_2^t(W_1^t, S_0, y^t; \gamma))$, *and we can evaluate* $p(v^t(W_1^t, S_0, y^t; \gamma); \gamma)$ *and* $p(w_2^t(W_1^t, S_0, y^t; \gamma); \gamma)$ *for all* S_0, W_1^t, *and* t.

Assumptions 1 and 3 imply that, for all γ and t, we can evaluate the conditional densities $p(y_t|W_1^t, S_0, y^{t-1}; \gamma)$ for all S_0 and W_1^t. To simplify the notation, we write (v^t, s^t, w_2^t) instead of the cumbersome expression $(v^t(W_1^t, S_0, y^t; \gamma), s^t(W_1^t, S_0, y^t; \gamma), w_2^t(W_1^t, S_0, y^t; \gamma))$. Hence, for all γ and t, we have $p(y_t|W_1^t, S_0, y^{t-1}; \gamma) = p(v_t; \gamma)p(w_{2,t}; \gamma)|dy(v_t, w_{2,t}; \gamma)|$ for all S_0 and W_1^t, where $|dy(v_t, w_{2,t}; \gamma)|$ stands for the determinant of the Jacobian matrix of y_t with respect to V_t and $W_{2,t}$ evaluated at v_t and $w_{2,t}$.

Using Assumptions 2 and 3, we can define the likelihood of the data as follows. If y^T is a realization of the random variable Y^T, its likelihood conditional on parameter values γ is

$$(3) \qquad L(y^T; \gamma) = \prod_{t=1}^{T} p(y_t|y^{t-1}; \gamma)$$
$$= \prod_{t=1}^{T} \int\int p(y_t|W_1^t, S_0, y^{t-1}; \gamma) p(W_1^t, S_0|y^{t-1}; \gamma)\, dW_1^t\, dS_0.$$

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

ASSUMPTION 4: *For all* γ *and* t, *the model gives some positive probability to the data* y^T, *that is*, $p(y_t|W_1^t, S_0, y^{t-1}; \gamma) > \xi \geq 0$ *for all* S_0 *and* W_1^t.

Assumption 4 allows us to write the likelihood (3) in the recursive form

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

for all γ. This structure will be useful for proving our results in the next sections.

We let $\widehat{\gamma}(y^T) \equiv \arg\max_{\gamma \in \Upsilon} p(y^T; \gamma)$ be the pseudo-maximum likelihood point estimate (PMLE). Note that we do not assume that there exists a value γ^* such that $p(y^T; \gamma^*) = p_0^T(y^T)$, i.e., the model may be misspecified (hence, the term "pseudo").

3. CONVERGENCE OF THE LIKELIHOOD

If the researcher knows the transition and measurement equations, $\varphi(\cdot, \cdot; \gamma)$ and $g(\cdot, \cdot; \gamma)$, evaluation of the likelihood function (3) is conceptually a simple task. However, in most real-life applications, the economist has access only to numerical approximations to the transition and measurement equations, $\varphi_j(\cdot, \cdot; \gamma)$ and $g_j(\cdot, \cdot; \gamma)$. We index the approximations by j to emphasize that, frequently, the approximation to the unknown transition and measurement equations admits refinements that will imply that $\varphi_j(\cdot, \cdot; \gamma)$ and $g_j(\cdot, \cdot; \gamma)$ converge to their exact values as j goes to infinity.

However, the use of $\varphi_j(\cdot, \cdot; \gamma)$ and $g_j(\cdot, \cdot; \gamma)$ raises a fundamental issue. The researcher cannot evaluate the exact likelihood function $L(y^T; \gamma)$, implied by the exact $\varphi(\cdot, \cdot; \gamma)$ and $g(\cdot, \cdot; \gamma)$, because she does not have access to those latter two functions. The researcher can only evaluate the approximated likelihood $L_j(y^T; \gamma)$ implied by the approximated $\varphi_j(\cdot, \cdot; \gamma)$ and $g_j(\cdot, \cdot; \gamma)$. What are the effects on inference of employing $L_j(y^T; \gamma)$ instead of $L(y^T; \gamma)$? Does $L_j(y^T; \gamma)$ converge to $L(y^T; \gamma)$? What about the point estimates?

This section shows that, under some conditions, for any given value of the parameters γ, the approximated likelihood function, $L_j(y^T; \gamma)$, converges to the exact likelihood function, $L(y^T; \gamma)$, as the approximated transition and measurement equations $\varphi_j(\cdot, \cdot; \gamma)$ and $g_j(\cdot, \cdot; \gamma)$ converge to the exact transition and measurement functions. Formally, we prove that for any given γ, $L_j(y^T; \gamma) \to L(y^T; \gamma)$ as $\varphi_j(\cdot, \cdot; \gamma) \to \varphi(\cdot, \cdot; \gamma)$ and $g_j(\cdot, \cdot; \gamma) \to g(\cdot, \cdot; \gamma)$.

First, we establish that for all γ and t, the conditional probability $p(y_t | W_1^t, S_0, y^{t-1}; \gamma)$ is a continuous, real-valued function of S_0 for all W_1^t.

LEMMA 1: *Let $\gamma \in \Upsilon$. Under Assumptions* 1 *and* 3, *for all t*, $p(y_t | W_1^t, S_0, y^{t-1}; \gamma) \in C(W_1^t, S_0)$.

The proof of Lemma 1, as well as the proofs of the other results in this paper, can be found in the Appendix. This lemma implies that $L(y^T; \gamma)$ is bounded, since $p(y_t | W_1^t, S_0, y^{t-1}; \gamma)$ is continuous with bounded support.

We now prove that for all j, γ, and t, the conditional probability $p_j(y_t | W_1^t, S_0, y^{t-1}; \gamma)$ associated with the approximated transition and measurement equations is also a continuous, real-valued function of S_0 for all W_1^t. To do so, we follow a structure parallel to that of our previous section.

First we assume:

ASSUMPTION 5: *For all j and γ, functions $\varphi_j(\cdot,\cdot;\gamma)$ and $g_j(\cdot,\cdot;\gamma)$ are continuous. For all j and γ, $\varphi_j(\cdot,\cdot;\gamma)$ and $g_j(\cdot,\cdot;\gamma)$ are continuously differentiable at all points except at a finite number of points. At the points of differentiability, all partial derivatives are bounded, and the bounds are independent of j.*

Assumption 5 ensures continuity of $\varphi_j(\cdot,\cdot;\gamma)$ and $g_j(\cdot,\cdot;\gamma)$ at all points, while both functions may not be differentiable at a finite number of points. This lack of differentiability allows us to consider solution methods that, by construction, have kinks at a finite number of points. Those include, for example, value function iteration with linear interpolation or the finite elements method with linear basis functions. Under further mild regularity conditions, all our results will hold when these functions fail to be differentiable at a countable number of points.

Let $\mu_j^*(S;\gamma)$ be the invariant distribution associated with the approximated function $\varphi_j(\cdot,\cdot;\gamma)$. Then we assume:

ASSUMPTION 6: *For all γ, the invariant distribution for S, $\mu_j^*(S;\gamma)$, has a Radon–Nikodym derivative with respect to the Lebesgue measure.*

As was the case in Assumption 2, with some extra work this assumption could be written in terms of the policy functions $\varphi_j(\cdot,\cdot;\gamma)$ and $g_j(\cdot,\cdot;\gamma)$.

We also postulate the equivalent of Assumption 3 for the approximated functions.

ASSUMPTION 7: *For all j, γ, and t, the system of equations*

$$\begin{aligned}
&S_1 = \varphi_j(S_0,(W_{1,1},W_{2,1});\gamma),\\
&y_m = g_j(S_m,V_m;\gamma) \quad \textit{for} \quad m=1,2,\dots,t,\\
&S_m = \varphi_j(S_{m-1},(W_{1,m},W_{2,m});\gamma) \quad \textit{for} \quad m=2,3,\dots,t
\end{aligned}$$

has a unique solution, $(v_j^t(W_1^t,S_0,y^t;\gamma), s_j^t(W_1^t,S_0,y^t;\gamma), w_{j,2}^t(W_1^t,S_0,y^t;\gamma))$, and we can evaluate $p(v_j^t(W_1^t,S_0,y^t;\gamma);\gamma)$ and $p(w_{j,2}^t(W_1^t,S_0,y^t;\gamma);\gamma)$ for all S_0 and W_1^t but a finite number of points.

As before, Assumptions 5 and 7 imply that for all j, γ, and t, we can evaluate the conditional densities $p_j(y_t|W_1^t,S_0,y^{t-1};\gamma)$ for all S_0 and W_1^t but a finite number of points. Again, to simplify the notation, we write $(v_j^t,s_j^t,w_{j,2}^t)$ rather than the cumbersome expression $(v_j^t(W_1^t,S_0,y^t;\gamma), s_j^t(W_1^t,S_0,y^t;\gamma), w_{j,2}^t(W_1^t, S_0,y^t;\gamma))$. Since Assumption 5 implies that $dy_j(v_{j,t},w_{j,2,t};\gamma)$ exists for all but a finite set of S_0 and W_1^t, we have that, for all j, γ, and t, $p_j(y_t|W_1^t,S_0,y^{t-1};\gamma) = p(v_{j,t};\gamma)p(w_{j,2,t};\gamma)|dy_j(v_{j,t},w_{j,2,t};\gamma)|$ for all S_0 and W_1^t but a finite number of points. Notice that the Jacobian matrix of y_t with respect to V_t and $W_{2,t}$ in the approximated solution, $dy_j(\cdot,\cdot;\gamma)$, is now a function of j because of its dependency on $\varphi_j(\cdot,\cdot;\gamma)$ and $g_j(\cdot,\cdot;\gamma)$.

We also define the pseudo-maximum likelihood point estimate of the approximated model as $\widehat{\gamma}_j(y^T) \equiv \arg\max_{\gamma \in \Upsilon} p_j(y^T; \gamma)$ and require the approximated model to explain the data even if it does so with arbitrarily low probability.

ASSUMPTION 8: *For all j, γ, and t, the model gives some positive probability to the data y^T, that is, $p_j(y_t|W_1^t, S_0, y^{t-1}; \gamma) \geq \xi > 0$ for all S_0 and W_1^t but a finite number of points.*

Now we can prove the equivalent to Lemma 1 for the approximated functions. As in the case of the exact probability, this lemma ensures that $L_j(y^T; \gamma)$ is bounded.

LEMMA 2: *Let $\gamma \in \Upsilon$. Under Assumptions 5 and 7, for all j and t, $p_j(y_t|W_1^t, S_0, y^{t-1}; \gamma) \in C(W_1^t, S_0)$ but in a finite number of points.*

Under Assumptions 1 and 5, as the approximated transition and measurement functions converge to the exact transition and measurement functions, the invariant distribution generated by those approximations will converge to the invariant distribution implied by exact measurement and transition functions. This result is formally stated in Theorem 2 in Santos and Peralta-Alva (2005), which we reproduce here.

LEMMA 3: *Let $\gamma \in \Upsilon$, $\varphi_j(\cdot, \cdot; \gamma) \to \varphi(\cdot, \cdot; \gamma)$. Then, under Assumptions 1 and 5, every sequence of invariant distributions $\{\mu_j^*(S; \gamma)\}$ converges weakly to the invariant distribution $\mu^*(S; \gamma)$ associated with $\varphi(\cdot, \cdot; \gamma)$.*

3.1. *Main Result: Convergence of the Likelihood Function*

Whereas the densities $p_j(y_t|y^{t-1}; \gamma)$ and $p(y_t|y^{t-1}; \gamma)$ depend on the Jacobians of $\varphi_j(\cdot, \cdot; \gamma)$, $g_j(\cdot, \cdot; \gamma)$, $\varphi(\cdot, \cdot; \gamma)$, and $g(\cdot, \cdot; \gamma)$, to prove convergence of the likelihood function, we need to consider the convergence of such Jacobians as an intermediate step. To show that $d\varphi_j(\cdot, \cdot; \gamma) \to d\varphi(\cdot, \cdot; \gamma)$ and $dg_j(\cdot, \cdot; \gamma) \to dg(\cdot, \cdot; \gamma)$ as $\varphi_j(\cdot, \cdot; \gamma) \to \varphi(\cdot, \cdot; \gamma)$ and $g_j(\cdot, \cdot; \gamma) \to g(\cdot, \cdot; \gamma)$, we first need to assume:

ASSUMPTION 9: *For all j and γ, $\varphi_j(\cdot, \cdot; \gamma)$ and $g_j(\cdot, \cdot; \gamma)$ have bounded second partial derivatives at all points except at a finite number of points. The bounds are independent of j.*

This assumption is satisfied naturally by most solution methods for dynamic economic models, since a common strategy is to find an approximation to the unknown functions using some well-behaved basis, such as polynomials. Our previous examples of value function iteration and the finite elements method fit into this category. Other popular procedures, such as linearization and perturbation methods, do as well.

Our next lemma ensures that wherever the transition and measurement equations are differentiable, we have that $d\varphi_j(\cdot,\cdot;\gamma) \to d\varphi(\cdot,\cdot;\gamma)$ and $dg_j(\cdot,\cdot;\gamma) \to dg(\cdot,\cdot;\gamma)$ as $\varphi_j(\cdot,\cdot;\gamma) \to \varphi(\cdot,\cdot;\gamma)$ and $g_j(\cdot,\cdot;\gamma) \to g(\cdot,\cdot;\gamma)$. This seemingly surprising result follows from the Arzelà–Ascoli theorem.

LEMMA 4: *Let $\gamma \in \Upsilon$. Under Assumption* 9, *if $\varphi_j(\cdot,\cdot;\gamma) \to \varphi(\cdot,\cdot;\gamma)$ and $g_j(\cdot,\cdot;\gamma) \to g(\cdot,\cdot;\gamma)$, then $d\varphi_j(\cdot,\cdot;\gamma) \to d\varphi(\cdot,\cdot;\gamma)$ and $dg_j(\cdot,\cdot;\gamma) \to dg(\cdot,\cdot;\gamma)$.*

Now, we are ready to use Lemmas 1–4 to prove the main result of this section—the convergence of the likelihood function. Formally:

PROPOSITION 1: *Let $\gamma \in \Upsilon$. Under Assumptions* 1–9, *if $\varphi_j(\cdot,\cdot;\gamma) \to \varphi(\cdot,\cdot;\gamma)$ and $g_j(\cdot,\cdot;\gamma) \to g(\cdot,\cdot;\gamma)$, then*

$$\prod_{t=1}^{T} p_j(y_t|y^{t-1};\gamma) \to \prod_{t=1}^{T} p(y_t|y^{t-1};\gamma).$$

The result is key for applied work. It states that for any given γ, as we get better approximations of the policy function in our dynamic economic model, the computed likelihood converges to the exact likelihood. This finding provides a foundation for empirical estimates based on the approximation of policy functions, since it guarantees, at least asymptotically, that we are finding the right object of interest, the likelihood function implied by the economic model.

3.2. *Comments on the Assumptions*

The assumptions in the previous sections are intended to get the convergence of the approximated likelihood function to the exact likelihood, and hence, they seem necessary to carry out statistical inference in approximated dynamic equilibrium models via the likelihood function. In this section, we distinguish between technical and substantive assumptions.

We assume compactness of the support of S_t, V_t, and W_t. This assumption is important for the results shown in the paper because Lemma 3, which we borrow from Santos and Peralta-Alva (2005), requires such compactness. Recent work by Stachurski (2002), who studies the asymptotic behavior of the stochastic neoclassical growth model without compactness of the shocks and states, suggests it could be possible to relax this assumption with extra work.

We also assume that the densities of V_t and W_t are bounded and continuous. The continuity of the density is needed to prove Lemma 1, whereas the boundness of the density is used in the proof of Proposition 1. The assumption of independence of V_t and W_t within and across time is a technical assumption that can be relaxed with heavier notation.

Assumption 1 is substantial for the proofs of Lemmas 1 and 3. Assumption 2 implies that the invariant distribution $\mu^*(S;\gamma)$ is unique and has a Radon–Nikodym derivative with respect to the Lebesgue measure.[2] These two requirements are essential and may seem restrictive in certain contexts. However, if the model has multiple invariant distributions, the likelihood function is not unequivocally defined. Moreover, in the case of multiple invariant distributions, the likelihood function may not be approximated by numerical methods, since some ergodic sets may not be robust to perturbations of the model, i.e., the correspondence of invariant distributions may fail to be lower semicontinuous (see Santos and Peralta-Alva (2005)). Multiple steady states arise in several deterministic and stochastic settings (see Boldrin and Woodford (1990) for models with sunspots and endogenous fluctuations, Kehoe and Levine (1985) for overlapping generations models, and Benhabib and Farmer (1999) for models with taxes and externalities).[3] The existence of a Radon–Nikodym derivative is also important to handle the lack of differentiability of $\varphi_j(\cdot,\cdot;\gamma)$ and $g_j(\cdot,\cdot;\gamma)$ in a finite number of points. Assumption 3 is vital to have a well-defined likelihood. Assumption 4 is technical since it only rules out trivial models that assign zero probability to the data. Similar arguments apply for Assumptions 5–8. Finally, Assumption 9 is fundamental for Lemma 4.

3.3. *Applications of the Main Result*

Proposition 1 has a number of applications. We highlight just two of them. First, pointwise convergence implies that for any given γ and γ', the ratio of likelihood functions converges. This result is useful in all contexts in which likelihood ratios are built, such as in classical hypothesis testing or when implementing the Metropolis–Hastings algorithm for posterior simulation.

COROLLARY 1: *Let $\gamma, \gamma' \in \Upsilon$. If the conditions of Proposition* 1 *are satisfied for γ and γ', it follows that*

$$\frac{L_j(y^T;\gamma')}{L_j(y^T;\gamma)} \to \frac{L(y^T;\gamma')}{L(y^T;\gamma)}.$$

The second application of the result directly affects Bayesian inference. There are two main objects of interest in the Bayesian paradigm: the marginal

[2]The violation of this condition is the reason why, in the example in the working paper of the article, the sequence of approximated likelihoods of the model fails to converge to the exact likelihood.

[3]For some valuable criteria for testing the uniqueness of the invariant distribution, see Futia (1982) and Stokey, Lucas, and Prescott (1989, Chaps. 11 and 12).

likelihood of the model and the posterior of the parameters. The marginal likelihood of the model is defined as $p(y^T) = \int_Y L(y^T; \gamma)\pi(\gamma)\,d\gamma$, while the marginal likelihood of the approximated model is $p_j(y^T) = \int_Y L_j(y^T; \gamma)\pi(\gamma)\,d\gamma$. Marginal likelihoods are important as measures of fit of the model and for building Bayes ratios, a key step in the Bayesian comparison of models (Geweke (1998)).

Under further mild regularity conditionism we can show that $L(y^T; \gamma)$ and $L_j(y^T; \gamma)$ are continuous in γ and uniformly bounded over all j. Then, an application of Arzelà's theorem delivers the convergence of the marginal likelihood:

COROLLARY 2: *If the conditions of Proposition* 1 *are satisfied for all* γ, *it follows that* $p_j(y^T) \to p(y^T)$.

Given some prior distribution of the parameters, $\pi(\gamma)$, the posterior is given by $p(\gamma|y^T) \propto L(y^T; \gamma)\pi(\gamma)$ for the exact likelihood and by $p_j(\gamma|y^T) \propto L_j(y^T; \gamma)\pi(\gamma)$ for the approximated likelihood. Proposition 1 also implies the convergence of the posterior.

COROLLARY 3: *If the conditions of Proposition* 1 *are satisfied for all* γ, *it follows that* $p_j(\gamma|y^T) \to p(\gamma|y^T)$.

The posterior distribution of the parameters of the model—beyond expressing our conditional belief—is also useful for evaluating expectations of the form $E(h(\gamma)|y^T)$, in which $h(\gamma)$ is a function of interest. Examples of such functions include loss functions for point estimation and point prediction, indicator functions for percentile statements, moment conditions, predictive intervals, or turning point probabilities.

Consider the expectation of the exact model

$$E(h(\gamma)|y^T) = \frac{1}{p(y^T)} \int_Y h(\gamma) L(y^T; \gamma)\pi(\gamma)\,d\gamma$$

and the approximated model

$$E_j(h(\gamma)|y^T) = \frac{1}{p_j(y^T)} \int_Y h(\gamma) L_j(y^T; \gamma)\pi(\gamma)\,d\gamma.$$

Then:

COROLLARY 4: *If the conditions of Proposition* 1 *are satisfied for all* γ, *and if* $h(\gamma)L_j(y^T; \gamma)\pi(\gamma)$ *and* $h(\gamma)L(y^T; \gamma)\pi(\gamma)$ *are Riemann-integrable*, *then* $E_j(h(\gamma)|y^T) \to E(h(\gamma)|y^T)$.

It is important to notice that Proposition 1 shows pointwise convergence of the likelihood function. Therefore, we cannot use it to prove convergence of the PMLE, since we cannot swap the arg max and lim operator. The working paper version of this article provides additional assumptions to prove uniform convergence of the likelihood function and, consequently, to prove the convergence of the PMLE $\widehat{\gamma}_j(y^T) \to \widehat{\gamma}(y^T)$.

4. SPEED OF CONVERGENCE OF THE LIKELIHOOD

The goal of this section is to analyze the speed of convergence of the approximated likelihood function, $L_j(y^T;\gamma)$, to the exact likelihood function, $L(y^T;\gamma)$, for a fixed γ. Given a bound for the difference between the approximated and the exact transition, and measurement equations, $\|\varphi_j(\cdot,\cdot;\gamma) - \varphi(\cdot,\cdot;\gamma)\| \le \delta$ and $\|g_j(\cdot,\cdot;\gamma) - g(\cdot,\cdot;\gamma)\| \le \delta$, we will obtain a bound for the difference between the approximated and exact likelihood functions

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

Let us introduce some additional assumptions needed in this section.

ASSUMPTION 10: *For all γ, the densities of W_t and V_t are differentiable, with bounded partial derivatives.*

ASSUMPTION 11: *For all γ, $\varphi(\cdot,\cdot;\gamma)$ and $g(\cdot,\cdot;\gamma)$ are twice continuously differentiable, with bounded second partial derivatives.*

Now we prove:

LEMMA 5: *Let $\gamma \in \Upsilon$. Under Assumptions* 1, 3, 10, *and* 11, $p(y_t|W_1^t, S_0, y^{t-1};\gamma)$ *is continuously differentiable with bounded partial derivatives with respect to S_0 for all t.*

It follows that $p(y_t|W_1^t, S_0, y^{t-1};\gamma)$ is Lipschitz with respect to S_0 for all t, with Lipschitz constant L_p.

Once we have the continuity and differentiability of $p(y_t|W_1^t, S_0, y^{t-1};\gamma)$, the next step is to bound the difference $|p_j(y_t|W_1^t, S_0, y^{t-1};\gamma) - p(y_t|W_1^t, S_0, y^{t-1};\gamma)|$. This difference will be a key component when we evaluate the differences between likelihoods. We then parameterize both $\varphi_j(\cdot,\cdot;\gamma) = \varphi(\cdot,\cdot;\gamma,\theta_j)$, and $g_j(\cdot,\cdot;\gamma) = g(\cdot,\cdot;\gamma,\theta_j)$, where $\theta_j \in \Phi\ \forall j$ for a compact subset $\Phi \in \mathbb{R}^M$, in such a way that they have bounded partial derivatives with respect to θ, as functions of S, W, and V. The bounds are independent of j. This parameterization and Assumptions 10 and 11 will allow us to apply the implicit function theorem to prove Lemma 6. The next result uses Lemma 4 and follows from an application of the implicit theorem in a space of functions.

LEMMA 6: *Let* $\gamma \in \Upsilon$. *Under Assumptions* 1–11, *if* $\|\varphi_j(\cdot,\cdot;\gamma) - \varphi(\cdot,\cdot;\gamma)\| \leq \delta$ *and* $\|g_j(\cdot,\cdot;\gamma) - g(\cdot,\cdot;\gamma)\| \leq \delta$, *then there exists a positive constant* χ *such that, for all* t,

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

for all S_0 *and* W_1^t *but in a finite number of points.*

In the next proposition, we apply Theorem 6 of Santos and Peralta-Alva (2005). We first impose a contractivity condition on φ, which is equivalent to their Condition C.

CONDITION 1: *Let* $\gamma \in \Upsilon$. *There exists some constant* $0 < \alpha < 1$ *such that*

$$\int \|\varphi(S, W;\gamma) - \varphi(S', W;\gamma)\|\, dQ(W;\gamma) \leq \alpha\|S - S'\|$$

for all S, S' *and where* $Q(\cdot;\gamma)$ *is the distribution of* W.

Condition 1 arises naturally in a large class of applications in economics. For example, it appears in the stochastic neoclassical growth model (Schenk-Hoppé and Schmalfuss (2001)), in concave dynamic programs (Foley and Hellwig (1975) and Santos and Peralta-Alva (2005)), in learning models (Schmalensee (1975) and Ellison and Fudenberg (1993)), and in stochastic games (Sanghvi and Sobel (1976)). Also, it is a common condition in the literature on Markov chains (Stenflo (2001)). Santos and Peralta-Alva (2005), in their Examples 5.3 and 5.4, show how this condition can be checked for dynamic models both with and without a closed-form solution.

Now we are ready to prove the main result of this section. Given a bound for the difference between the approximated and exact transition and measurement equations, we can bound the difference between the approximated and exact likelihood functions. Formally:

PROPOSITION 2: *Let* $\gamma \in \Upsilon$. *Assume that Condition* 1 *holds. Under Assumptions* 1–11, *if* $\|\varphi_j(\cdot,\cdot;\gamma) - \varphi(\cdot,\cdot;\gamma)\| \leq \delta$ *and* $\|g_j(\cdot,\cdot;\gamma) - g(\cdot,\cdot;\gamma)\| \leq \delta$, *there are some positive constants* B *and* L *such that, for all* T,

$$\left|\prod_{t=1}^{T} p_j(y_t|y^{t-1};\gamma) - \prod_{t=1}^{T} p(y_t|y^{t-1};\gamma)\right| < \left(TB\chi + \frac{L}{1-\alpha}\right)\delta.$$

Proposition 2 states that the difference between the likelihoods is bounded by a linear function of the length of the sample of observations, T, and the bound on the error in the transition and measurement equation δ.[4]

[4]Santos (2000) shows that for a class of dynamic optimization problems, the approximation error of the policy function δ is of the same order of magnitude as the Euler equation residual.

A number of insights emerge from this result. First, second order approximation errors in the policy function, which almost always are ignored by researchers, have first order effects on the likelihood function. Any given error in the policy function δ gets multiplied by T. The intuition is that small errors in the policy function accumulate at the same rate at which the sample size grows. Similarly, third order approximation errors in the policy function will have second order effects on the likelihood function and so on.

Moreover, in empirical applications, the constants in the bound, B and χ, can be estimated from repeated solutions of the model under different numerical approximations. Consequently, the researcher can bound the difference between the exact and the approximated likelihood, and use Proposition 2 as a guide to determine the accuracy δ that she needs to ask from her solution method.

Second, to guarantee asymptotic convergence in the estimation of dynamic models, the error in the policy function must depend on the length of the sample: the longer the sample, the smaller the policy function error. Otherwise, the bound in the difference between the approximated and the exact likelihood goes to infinity. Our proposition suggests that justifying a solution method based on small errors in the policy function without a reference to the size of the sample may be misleading for estimation purposes.

Third, our result shows that there is an inherent limitation in the use of linearization methods to estimate nonlinear dynamic economies. This point is important because linearization is the most common strategy for computing approximate solutions of dynamic stochastic general equilibrium models, like the ones popular in macroeconomics. Proposition 2 shows that linearization is due to fail as the sample size grows. The reason is that linearization fixes the policy function error and, then, the exact and approximated likelihoods diverge as the T goes to infinity. Thus, Proposition 2 cautions us on the indiscriminate use of linearization.

Finally, Proposition 2 suggests the use of likelihood ratios as a diagnosis device to check for the importance of the errors in the approximated likelihood. The researcher can evaluate the likelihood of the model at PMLE parameter values for different choices of approximation errors and build a likelihood ratio. Suppose that we have two approximations of the transition and measurement equation j and j' with approximation errors δ and δ' such that $\delta < \delta'$. Then, for observations y^T, we can compute

$$LR\big(\widehat{\gamma}_j(y^T), \widehat{\gamma}_{j'}(y^T), j, j'\big) = \sum_{t=1}^{T} \log p_j\big(y_t|y^{t-1}; \widehat{\gamma}_j(y^T)\big) - \sum_{t=1}^{T} \log p_{j'}\big(y_t|y^{t-1}; \widehat{\gamma}_{j'}(y^T)\big),$$

Since Euler errors are easy to estimate, we can replace δ by an Euler error estimate and obtain a bound of the same order of magnitude.

where $\widehat{\gamma}_j(y^T)$ and $\widehat{\gamma}_{j'}(y^T)$ are the PMLEs of the parameters of the model. Note that, in general, $\widehat{\gamma}_j(y^T) \neq \widehat{\gamma}_{j'}(y^T)$.

Vuong (1989) develops the asymptotic behavior of this statistic to compare competing models that are nonnested, overlapping, or nested and to determine whether both, one, or neither is misspecified. Given the flexibility of his findings, we can interpret the two different approximations of the same exact model as two competing models. Consequently, the result of the likelihood ratio test will tell us if the data support one approximation of the model significantly better than the other one.

We suggest that a promising strategy to check the robustness of the inference could be to increase the accuracy of the numerical solution of the model until the value of the likelihood ratio is such that a researcher cannot distinguish between the version of the model implied by the less accurate solution and the version of the model with a more accurate solution. This proposal is easy to implement and might protect against some of the worst forms of incorrect inference that we document in the next section.

Note that, in general, it is dangerous to use a statistical test to evaluate a numerical approximation method, because approximation errors are not random quantities and because it is always possible to generate a sample long enough such that the approximate solution is rejected by the test. In our case, however, the likelihood ratio we propose escapes this criticism because it compares the importance of approximation errors with the sample error. Consequently, longer sample will discriminate better among competing solutions.[5]

A Bayesian version of this procedure will compute the Bayes factor of the two models $p_j(y^T)/p_{j'}(y^T)$ and follow the standard interpretation of the value of such ratio (see, for an example of this approach, Fernández-Villaverde and Rubio-Ramírez (2005)). The strategy will be again to increase the accuracy of the solution until we cannot tell the two approximations of the model apart using the Bayes factor.

5. AN APPLICATION

The working paper version of this article contains an application that illustrates how the first order effect on the likelihood function of second order errors on the solution of the policy function has a crucial impact when we perform statistical inference in real-life models. Our application makes three points. First, it shows how we can reject a correctly specified economic model in favor of an alternative (misspecified) statistical model just because the accuracy of the solution of the economic model is too low. Second, it demonstrates how we can diagnose the problem cleanly and choose a solution accuracy that leads us to reject the statistical model. Third, it documents important biases

[5]We thank a referee for pointing out this observation. In the previous lines, we follow her argument.

in parameter estimates because of approximation errors in the solution of the model and how we can eliminate those biases. Because of space considerations, we provide here only a short summary of our main findings and refer the interested reader to the working paper version of this article.

We use the neoclassical growth model as our underlying theoretical framework. A pragmatic consideration guides this choice. The neoclassical growth model and its variations are the workhorses of modern macroeconomics. For example, the successful new generation of business cycle models initiated by Smets and Wouters (2003) and estimated by likelihood methods is built around the backbone of the neoclassical growth model augmented with real and nominal rigidities. Consequently, any lesson learned with the basic model is likely to be useful in a large class of applications, including those that are highly relevant for policy-making. At the same time, the model is sufficiently simple to allow us to derive analytical results that will be useful for interpreting our findings.

We simulate 200 observations from the neoclassical growth model with full depreciation because, in this case, the model has an exact log-linear solution. Thus, we can compute the exact likelihood as the benchmark result for comparison purposes.

Imagine that we have three researchers trying to estimate parameters of the neoclassical growth model with full depreciation and to evaluate its fit to the data. To do so, we give each of them the same sample of 200 observations that we generated from the exact log-linear model. Since none of the three researchers knows that the model has a closed-form solution, they solve the model using value function iteration and estimate it using maximum likelihood. The only difference is that the first researcher uses 2,000 points in the grid, the second researcher uses 4,000 points, and the third researcher uses 40,000 points. The problem is so well-behaved that the approximation that uses 2,000 points is already excellent in terms of accuracy: the welfare loss from using the approximated policy function instead of the exact one is less than $1/20$th of 1% in terms of consumption. Hence, even the 2,000 grid points approximation delivers a much higher accuracy than the approximations typically used in the estimation literature.[6]

The researchers employ a likelihood ratio test to compare their model against a $\mathrm{VAR}(1)$ (see, again, Vuong (1989)). Note that a $\mathrm{VAR}(1)$ is misspecified, since the dynamics of the neoclassical growth model, from which we have simulated the data, imply a $\mathrm{VAR}(\infty)$.

What are the results of the test? The researcher using 2,000 points rejects the neoclassical growth model in favor of the $\mathrm{VAR}(1)$. The researcher using 4,000 points cannot tell the two models apart. Finally, the researcher using 40,000 points rejects the $\mathrm{VAR}(1)$ in favor of the neoclassical growth model.

[6]Note that we do not log-linearize the equilibrium conditions of the model around the steady state, as is common in macroeconomics, since those will deliver, for our particular case with full depreciation, the exact solution.

This situation is striking: three different outcomes of the test induced just because of the difference in the grid. Note how misleading the inference is in the first two cases: although the data are generated by the model that the researcher is testing, she cannot reject that the VAR(1) fits the data better because of the second order numerical errors in the approximated policy function.

Our choice of 2,000, 4,000, and 40,000 grid points was not random. We applied the likelihood ratio principle, as we previously suggested, to determine the accuracy cutoff given the sample size. When we use less than 20,000 points, the value of the likelihood changes substantially as we add more grid points. Around 20,000 grid points, the likelihood ratios for different levels of accuracy stabilize and we can reject the VAR(1) at higher and higher confidence levels. Moreover, since in our application we can compute the exact likelihood of the model, 40,000 grid points deliver an approximated likelihood that cannot be statistically distinguished from the exact one. Our application shows how likelihood ratio testing saves the researchers from some of the dangers of mistaken inference induced by approximation errors in the solution of the economic model.

Finally, what are the implications of the approximation errors in the policy function for parameter inference? When we use 2,000 points, we get an estimate of the discount factor that implies a steady-state interest rate 140 basis points higher than the exact one. When we use 4,000 points, our estimate is only 10 basis points off, and with 40,000 points, we get an almost perfect point estimate. Those differences show a substantial bias in estimation of parameters. Also, the convergence of the estimates as we increase the number of grid points demonstrates how the application of the likelihood ratio test may alleviate the biases.

Our results show how approximating the policy functions has a nontrivial effect on either model selection or parameter point estimation, and how we can easily diagnose and correct the problem using a likelihood ratio test. Our findings appear in a simple model. Further evidence provided by Fernández-Villaverde and Rubio-Ramírez (2005) suggests that these problems also arise in real data and affect real inferences.

6. CONCLUSIONS

In this paper we have studied the consequences of using approximated likelihood functions instead of the exact likelihood when we estimate computed dynamic economic models. We have offered a positive result—the convergence of the approximated likelihood to the exact likelihood as the approximated policy functions converge to the exact policy functions. We have also shown that the errors in the approximated likelihood function accumulate as the sample size grows and that to guarantee convergence of our estimates, we need to reduce the size of the error in the approximated policy function as we include more data. Finally, we propose to use a likelihood ratio test as a diagnostic device for problems derived from the use of approximated likelihoods.

There are several additional issues that we leave for future analysis. First, it would be important to eliminate the assumption of continuity of the transition and the measurement equations, since a large class of models, especially in microeconomic applications, implies choices with jumps and discontinuities. Second, it seems desirable to establish the error bounds of Proposition 2 under milder assumptions than our contractivity condition. However, it appears that the structure of the bound will remain the same under more general assumptions. Hence, all the implications of our analysis for conducting statistical inference and testing of dynamic equilibrium models are not specific to this contractivity property and should prevail in more general settings. Third, we could show the convergence of standard error estimates to complete the analysis of classical estimation. Finally, we could cover settings with multiplicity of equilibria like those that often appear in game theory (Bajari, Hong, and Ryan (2004)), industrial organization (Pakes, Ostrovsky, and Berry (2004)), and macroeconomics (Lubik and Schorfheide (2004)).

Dept. of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19101, U.S.A.; jesusfv@econ.upenn.edu; www.econ.upenn.edu/˜jesusfv/,

Federal Reserve Bank of Atlanta, Research Department, 1000 Peachtree Street NE, Atlanta, GA 30309-4470, U.S.A.; Juan.Rubio@atl.frb.org; http://www.econ.umn.edu/˜rubio/,

and

Dept. of Economics, W. P. Carey School of Business, Arizona State University, Box 873806, Tempe, AZ 85287-3806, U.S.A.; Manuel.Santos@asu.edu.

Manuscript received May, 2005; final revision received September, 2005.

APPENDIX

PROOF OF LEMMA 1: Let $\gamma \in \Upsilon$. Note that v_t and $w_{2,t}$ are continuous functions of W_1^t and S_0, and that $|dy(v_t, w_{2,t}; \gamma)|$ is a continuous function of v_t and $w_{2,t}$. Therefore, since V_t and $W_{2,t}$ have continuous densities, it is the case that $p(y_t|W_1^t, S_0, y^{t-1}; \gamma) \in C(W_1^t, S_0)$. *Q.E.D.*

PROOF OF LEMMA 2: The proof follows the same steps as the proof of the previous lemma. *Q.E.D.*

In the proof of Lemma 4 we use the following well-known theorems.

THEOREM 1: *Assume $\{a_n\}$ is an infinite sequence in a metric space (X, d). Then $a_n \to a$ if and only if every infinite subsequence $\{a'_n\} \subset \{a_n\}$ has a convergent subsequence $\{a''_n\} \subset \{a'_n\}$ such that $a''_n \to a$ (Proposition* 19 *in DePree and Swartz* (1988, *p*. 31)).

THEOREM 2: *If* $f_n \to f$ *in the* sup norm *and* $f'_n \to g$ *in the* sup norm, *then* $g = f'$ (*Theorem* 8.6.3 *in Dieudonné* (1960, *p*. 157)).

PROOF OF LEMMA 4: Assumption 9 implies that $\{\varphi_j(\cdot,\cdot;\gamma)\}$ and $\{g_j(\cdot,\cdot;\gamma)\}$ have uniformly bounded second derivatives. Hence, $\{d\varphi_j(\cdot,\cdot;\gamma)\}$ and $\{dg_j(\cdot,\cdot;\gamma)\}$ is a family of equicontinuous functions. Moreover, by the Arzelà–Ascoli theorem, every subsequence of $\{\varphi_j(\cdot,\cdot;\gamma)\}$ and $\{g_j(\cdot,\cdot;\gamma)\}$ has a convergent subsequence in the C^1 topology.[7] Since $\{\varphi_j(\cdot,\cdot;\gamma)\}$ and $\{g_j(\cdot,\cdot;\gamma)\}$ converge to $\varphi(\cdot,\cdot;\gamma)$ and $g(\cdot,\cdot;\gamma)$, respectively, every subsequence of $\{\varphi_j(\cdot,\cdot;\gamma)\}$ and $\{g_j(\cdot,\cdot;\gamma)\}$ has a convergent subsequence in the sup norm to $\varphi(\cdot,\cdot;\gamma)$ and $g(\cdot,\cdot;\gamma)$. Moreover, Theorem 2 implies that every subsequence of $\{d\varphi_j(\cdot,\cdot;\gamma)\}$ and $\{dg_j(\cdot,\cdot;\gamma)\}$ has a convergent subsequence in the sup norm to $d\varphi(\cdot,\cdot;\gamma)$ and $dg(\cdot,\cdot;\gamma)$. Therefore, Theorem 1 implies that $d\varphi_j(\cdot,\cdot;\gamma) \to d\varphi(\cdot,\cdot;\gamma)$ and $dg_j(\cdot,\cdot;\gamma) \to dg(\cdot,\cdot;\gamma)$ in the sup norm. *Q.E.D.*

PROOF OF PROPOSITION 1: Let $\gamma \in \Upsilon$. The proof is divided into two steps.

Step 1—*Convergence of* $p_j(y_t|W_1^t, S_0, y^{t-1};\gamma)$: First, remember that Assumption 1 entails that $\varphi(\cdot,\cdot;\gamma)$, $g(\cdot,\cdot;\gamma)$, and their partial derivatives are continuous. Second, note that Assumption 5 states that $\varphi_j(\cdot,\cdot;\gamma)$ and $g_j(\cdot,\cdot;\gamma)$ are continuous, while their partial derivatives are continuous at all but a finite number of points. Third, recall that the densities of V_t and $W_{2,t}$ are continuous. Finally, we have also assumed that $\varphi_j(\cdot,\cdot;\gamma) \to \varphi(\cdot,\cdot;\gamma)$ and $g_j(\cdot,\cdot;\gamma) \to g(\cdot,\cdot;\gamma)$. Thus, by Assumption 9, we have that $|dy_j(\cdot,\cdot;\gamma)| \to |dy(\cdot,\cdot;\gamma)|$ at all but a finite number of points, and we can assert that $p_j(y_t|W_1^t, S_0, y^{t-1};\gamma) \to p(y_t|W_1^t, S_0, y^{t-1};\gamma)$ except in a finite number of points.

Step 2—*Convergence of* $\prod_{t=1}^T p_j(y_t|y^{t-1};\gamma)$: Assumptions 4 and 8 allow us to write

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

$$\prod_{t=1}^T \widetilde{p}_j(y_t|y^{t-1};\gamma) = \int\left(\int \prod_{t=1}^T p(W_{1,t};\gamma)\,p(y_t|W_1^t, S_0, y^{t-1};\gamma)\,dW_1^t\right)\mu_j^*(dS_0;\gamma),$$

[7] The C^1 topology is defined as $\|f\|_{C^1} = \|f\| + \|f'\|$, where $\|\cdot\|$ is the sup norm.

and

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

Let

$$f_T(S_0;\gamma) = \int \prod_{t=1}^{T} p(W_{1,t};\gamma) p(y_t|W_1^t, S_0, y^{t-1};\gamma)\, dW_1^t.$$

Thus the likelihood functions are

$$\prod_{t=1}^{T} p(y_t|y^{t-1};\gamma) = \int f_T(S_0;\gamma)\mu^*(dS_0;\gamma)$$

and

$$\prod_{t=1}^{T} \widetilde{p}_j(y_t|y^{t-1};\gamma) = \int f_T(S_0;\gamma)\mu_j^*(dS_0;\gamma).$$

By Lemma 1, $f_T(S_0;\gamma)$ is continuous. Therefore, we can apply Corollary 3.3 of Santos and Peralta-Alva (2005) to show that

$$\prod_{t=1}^{T} \widetilde{p}_j(y_t|y^{t-1};\gamma) \to \prod_{t=1}^{T} p(y_t|y^{t-1};\gamma). \tag{4}$$

If we define

$$f_{j,T}(S_0;\gamma) = \int \prod_{t=1}^{T} p(W_{1,t};\gamma) p_j(y_t|W_1^t, S_0, y^{t-1};\gamma)\, dW_1^t,$$

it follows that

$$\prod_{t=1}^{T} p_j(y_t|y^{t-1};\gamma) = \int f_{j,T}(S_0;\gamma)\mu_j^*(dS_0;\gamma).$$

Note that $W_{1,t}$ has bounded support and bounded density. Also, Lemma 2 shows that $p_j(y_t|W_1^t, S_0, y^{t-1};\gamma)$ is continuous except in a finite number of points, with bounded support, and hence it is bounded. Consequently,

$f_{j,T}(S_0;\gamma)$ is bounded. In addition, Step 1 shows that $p_j(y_t|W_1^t, S_0, y^{t-1};\gamma) \to p(y_t|W_1^t, S_0, y^{t-1};\gamma)$ except in a finite number of points. Hence, $f_{j,T}(S_0;\gamma) \to f_T(S_0;\gamma)$, but in a finite number of points.

Therefore, for every $\varepsilon > 0$, $\exists N$ such that if $j > N$, $|f_{j,T}(S_0;\gamma) - f_T(S_0;\gamma)| < \varepsilon$, except in a finite number of points. Thus, we can write

$$\left|\prod_{t=1}^{T} p_j(y_t|y^{t-1};\gamma) - \prod_{t=1}^{T} \widetilde{p}_j(y_t|y^{t-1};\gamma)\right| \leq \int |f_{j,T}(S_0;\gamma) - f_T(S_0;\gamma)| \mu_j^*(dS_0;\gamma). \tag{5}$$

Since $\mu_j^*(dS_0;\gamma)$ has a density with respect to the Lebesgue measure (Assumption 6), the right-hand side of (5) is less than ε. We then conclude that

$$\prod_{t=1}^{T} p_j(y_t|y^{t-1};\gamma) \to \prod_{t=1}^{T} \widetilde{p}_j(y_t|y^{t-1};\gamma). \tag{6}$$

To complete the proof, we put together the convergence results (4) and (6).

Q.E.D.

PROOF OF COROLLARY 1: Let $\gamma, \gamma' \in \Upsilon$. Proposition 1 shows that $L_j(y^T;\gamma) \to L(y^T;\gamma)$ and $L_j(y^T;\gamma') \to L(y^T;\gamma')$. Moreover, Assumptions 4 and 8 imply that $L(y^T;\gamma) \geq \xi > 0$ and $L_j(y^T;\gamma) \geq \xi > 0$ for all j. Therefore

$$\frac{L_j(y^T;\gamma')}{L_j(y^T;\gamma)} \to \frac{L(y^T;\gamma')}{L(y^T;\gamma)}.$$ *Q.E.D.*

PROOF OF COROLLARY 2: Let $\gamma \in \Upsilon$. Since the approximated likelihoods $L_j(y^T;\gamma)$ and $L(y^T;\gamma)$ are bounded and Riemann-integrable (because they are continuous in γ), we can apply Arzelà's theorem (see Apostol (1974, Theorem 9.12)) and Proposition 1 to get

$$\int_{\Upsilon} L_j(y^T;\gamma)\pi(\gamma)\,d\gamma \to \int_{\Upsilon} L(y^T;\gamma)\pi(\gamma)\,d\gamma.$$ *Q.E.D.*

PROOF OF COROLLARY 3: Let $\gamma \in \Upsilon$. Proposition 1 shows that $L_j(y^T;\gamma) \to L(y^T;\gamma)$. Then $L_j(y^T;\gamma)\pi(\gamma) \to L(y^T;\gamma)\pi(\gamma)$ and the result follows. *Q.E.D.*

PROOF OF COROLLARY 4: Let $\gamma \in \Upsilon$. Proposition 1 shows that $L_j(y^T;\gamma) \to L(y^T;\gamma)$ and Corollary 2 shows that $p_j(y^T) \to p(y^T)$. The result follows from an application of Arzelà's theorem. *Q.E.D.*

PROOF OF LEMMA 5: Let $\gamma \in \Upsilon$. To prove that $p(y_t|W_1^t, S_0, y^{t-1}; \gamma)$ is continuously differentiable with respect to S_0, we need to show that

$$\frac{\partial p(y_t|W_1^t, S_0, y^{t-1}; \gamma)}{\partial S_{0,i}}$$

exists and is continuous for all i.

Assumption 3 allows us to write

$$p(y_t|W_1^t, S_0, y^{t-1}; \gamma) = p(v_t; \gamma)p(w_{2,t}; \gamma)|dy(v_t, w_{2,t}; \gamma)|$$

for all t. Since, in addition, V_t and $W_{2,t}$ have bounded densities, Assumptions 1, 10, and 11 imply that

$$\frac{\partial p(y_t|W_1^t, S_0, y^{t-1}; \gamma)}{\partial S_{0,i}}$$

exists and is bounded for all t and all i. *Q.E.D.*

PROOF OF LEMMA 6: Let $\gamma \in \Upsilon$. Let (v^t, s^t, w_2^t) be the unique solution to the system of equations

$$S_1 = \varphi(S_0, (W_{1,1}, W_{2,1}); \gamma),$$
$$y_m = g(S_m, V_m; \gamma) \quad \text{for} \quad m = 1, 2, \ldots, t,$$
$$S_m = \varphi(S_{m-1}, (W_{1,m}, W_{2,m}); \gamma) \quad \text{for} \quad m = 2, 3, \ldots, t,$$

and let $(v_j^t, s_j^t, w_{j,2}^t)$ be the unique solution to the approximated system of equations[8]

$$S_1 = \varphi_j(S_0, (W_{1,1}, W_{2,1}); \gamma),$$
$$y_m = g_j(S_m, V_m; \gamma) \quad \text{for} \quad m = 1, 2, \ldots, t,$$
$$S_m = \varphi_j(S_{m-1}, (W_{1,m}, W_{2,m}); \gamma) \quad \text{for} \quad m = 2, 3, \ldots, t.$$

By Assumption 1, functions φ and g are differentiable. In addition, Assumption 4 implies that $|dy(v_t, w_{2,t}; \gamma)| \neq 0$ for all t. Since $\|\varphi_j(\cdot, \cdot; \gamma) - \varphi(\cdot, \cdot; \gamma)\| \leq \delta$ and $\|g_j(\cdot, \cdot; \gamma) - g(\cdot, \cdot; \gamma)\| \leq \delta$, by the implicit function theorem of Schwartz (see Theorem G.2.3 in Mas-Colell (1985, p. 32)), there exists a

[8]Both $(v_j^t, s_j^t, w_{j,2}^t)$ and (v^t, s^t, w_2^t) depend on s_0, and w_1^t, but to simplify notation, we do not make this dependence explicit.

$\lambda(S_0, W_1^t)$ such that

$$\|(v_j^t, s_j^t, w_{j,2}^t) - (v^t, s^t, w_2^t)\| \leq \lambda(S_0, W_1^t)\delta. \tag{7}$$

Since the model is stationary, equation (7) holds for all t.

Notice that $\lambda(S_0, W_1^t)$ depends on the derivatives of $\varphi_j(\cdot, \cdot; \gamma)$ and $g_j(\cdot, \cdot; \gamma)$ with respect to θ_j. These derivatives are bounded independently of j. Therefore $\exists\, \lambda$ such that $\|(v_j^t, s_j^t, w_{j,2}^t) - (v^t, s^t, w_2^t)\| \leq \lambda\delta$ for all S_0 and W_1^t.

Assumption 10 implies that the densities of V_t and W_t are absolutely continuous. Then, $\exists\, \varepsilon$ such that

$$|p(v_{j,t}; \gamma)p(w_{j,2,t}; \gamma) - p(v_t; \gamma)p(w_{2,t}; \gamma)| \leq \varepsilon\delta \tag{8}$$

for all S_0 and W_1^t. As before, note that equation (8) also holds for all t.

By Lemma 5, the determinant of the Jacobian matrix of y_t with respect to $V_t, W_{2,t}$, $|dy(\cdot, \cdot; \gamma)|$ is Lipschitz. Let L_y be the Lipschitz constant. Then

$$\big||dy(v_{j,t}, w_{j,2,t}; \gamma)| - |dy(v_t, w_{2,t}; \gamma)|\big| \leq L_y\lambda\delta \tag{9}$$

for all S_0 and W_1^t.

By Lemma 4 and the fact that $\|\varphi_j(\cdot, \cdot; \gamma) - \varphi(\cdot, \cdot; \gamma)\| \leq \delta$ and $\|g_j(\cdot, \cdot; \gamma) - g(\cdot, \cdot; \gamma)\| \leq \delta$ we have that $\|d\varphi_j(\cdot, \cdot; \gamma) - d\varphi(\cdot, \cdot; \gamma)\| \leq \kappa\delta$ and $\|dg_j(\cdot, \cdot; \gamma) - dg(\cdot, \cdot; \gamma)\| \leq \kappa\delta$ for some constant κ except in a finite number of points. Then, by Assumptions 1 and 5, we know that $\exists\, \Psi_1$ such that

$$\big|dy_j(v_{j,t}, w_{j,2,t}; \gamma)[r, s] - dy(v_{j,t}, w_{j,2,t}; \gamma)[r, s]\big| < \Psi_1\delta$$

for all r and s, and for all s_0 and w_1^t, except in a finite number of points. Here $A[r, s]$ stands for the row r and column s of matrix A.

Note that if A and B are two $n \times n$ matrices such that $|A[i, j] - B[i, j]| < \Psi_1\delta$ and $|A[i, j]|, |B[i, j]| < \Psi_2$, then $|\det(A) - \det(B)| < n!n\Psi_2^{n-1}\Psi_1\delta$. In addition, Assumptions 1 and 5 also imply that φ_j, φ, g_j, and g are Lipschitz. Therefore, $\exists\, \Psi_2$ such that

$$\big|\det(dy_j(v_{j,t}, w_{j,2,t}; \gamma)) - \det(dy(v_{j,t}, w_{j,2,t}; \gamma))\big| \leq n!n\Psi_2^{n-1}\Psi_1\delta \tag{10}$$

for all S_0 and W_1^t except in a finite number of points.

Using equations (9) and (10), we get

$$\big|\det(dy_j(v_{j,t}, w_{j,2,t}; \gamma)) - \det(dy(v_t, w_{2,t}; \gamma))\big| \leq (n!n\Psi_2^{n-1}\Psi_1 + L_y\lambda)\delta \tag{11}$$

for all S_0 and W_1^t.

Now, let $\Psi_3 = (n!n\Psi_2^{n-1}\Psi_1 + L_y\lambda)$. We can put together equations (8) and (11) to find

$$\begin{aligned}
&\big|p(v_{j,t};\gamma)p(w_{j,2,t};\gamma)|dy_j(v_{j,t},w_{j,2,t};\gamma)| \\
&\qquad - p(v_t;\gamma)p(w_{2,t};\gamma)|dy(v,w_{2,t};\gamma)|\big| \\
&\quad \le |p(v_{j,t};\gamma)p(w_{j,2,t};\gamma)|\varepsilon\delta + |dy(v_t,w_{2,t};\gamma)|\Psi_3\delta
\end{aligned}$$

for all S_0 and W_1^t except in a finite number of points.

Note that $p(v;\gamma)$ and $p(w_2;\gamma)$ are bounded functions. By Assumption 1, $|dy(v,w_2;\gamma)|$ is also a bounded function. Let B_1 and B_2 be the bounds to $p(v;\gamma)p(w_2;\gamma)$ and $|dy(v,w_2;\gamma)|$, respectively. Define $B = \max\{B_1, B_2\}$. Then

$$\begin{aligned}
&\big|p(v_{j,t};\gamma)p(w_{j,2,t};\gamma)|dy_j(v_{j,t},w_{j,2,t};\gamma)| \\
&\qquad - p(v_t;\gamma)p(w_{2,t};\gamma)|dy(v,w_{2,t};\gamma)|\big| \\
&\quad \le B\delta(\varepsilon + \Psi_3)
\end{aligned}$$

for all s_0 and w_1^t, but in a finite number of points. For $\chi = B(\varepsilon + \Psi_3)$, the lemma is proved. *Q.E.D.*

PROOF OF PROPOSITION 2: Let $\gamma \in \Upsilon$. Define $f_T(S_0;\gamma)$ as in the proof of Proposition 1 and note that

$$\begin{aligned}
\frac{\partial f_T(S_0;\gamma)}{\partial S_{0,i}} = \int \prod_{t=1}^{T} p(W_{1,t};\gamma) \sum_{t=1}^{T} \frac{\partial p(y_t|W_1^t, S_0, y^{t-1};\gamma)}{\partial S_{0,i}} \\
\times \prod_{s=1, s\neq r}^{T} p(y_s|W_1^s, S_0, y^{s-1};\gamma)\, dW_1^t
\end{aligned}$$

is bounded because Lemma 5 bounds

$$\frac{\partial p(y_t|W_1^t, S_0, y^{t-1};\gamma)}{\partial S_{0,i}}$$

for all t and i, and Lemma 1 bounds $p(y_s|W_1^s, S_0, y^{s-1};\gamma)$ for all s. Therefore, $f_T(S_0;\gamma)$ is Lipschitz for all t with Lipschitz constant L (the Lipschitz constant is different for each t, but since t is finite, we can set a global L).

By Condition 1, we can then apply Theorem 6 of Santos and Peralta-Alva (2005) to $f_T(S_0;\gamma)$ to get

$$\left|\prod_{s=1}^{T} p(y_t|y^{t-1};\gamma) - \prod_{s=1}^{T} \widetilde{p}_j(y_t|y^{t-1};\gamma)\right| \le \frac{L\delta}{1-\alpha}. \tag{12}$$

Note now that using the values for the likelihoods in the proof of Proposition 1, we have

$$
\begin{aligned}
(13)\qquad & \left|\prod_{s=1}^{T} p_j(y_t|y^{t-1};\gamma) - \prod_{s=1}^{T} \widetilde{p}_j(y_t|y^{t-1};\gamma)\right| \\
&= \int\Bigg(\int \prod_{t=1}^{T} p(W_{1,t};\gamma) \\
&\qquad \times \big(p_j(y_t|W_1^t, S_0, y^{t-1};\gamma) - p(y_t|W_1^t, S_0, y^{t-1};\gamma)\big)\, dW_1^t\Bigg) \\
&\qquad \times \mu_j^*(dS_0;\gamma).
\end{aligned}
$$

Lemmas 1 and 2 show that $p(y_t|W_1^t, S_0, y^{t-1};\gamma)$ and $p_j(y_t|W_1^t, S_0, y^{t-1};\gamma)$ are bounded for all t and j. Thus, we can define a constant B such that

$$
\int\Bigg(\int \prod_{t=1}^{T} p(W_{1,t};\gamma) B \sum_{t=1}^{T} \big| p_j(y_t|W_1^t, S_0, y^{t-1};\gamma) - p(y_t|W_1^t, S_0, y^{t-1};\gamma)\big|\, dW_1^t\Bigg)\mu_j^*(dS_0;\gamma)
$$

is an upper bound to (13).

Lemma 6 shows that $|p_j(y_t|W_1^t, S_0, y^{t-1};\gamma) - p(y_t|W_1^t, S_0, y^{t-1};\gamma)| \leq \chi\delta$ for all t, and for all S_0 and W_1^t but for a finite number of points. Therefore,

$$
(14)\qquad \left|\prod_{s=1}^{T} p_j(y_t|y^{t-1};\gamma) - \prod_{s=1}^{T} \widetilde{p}_j(y_t|y^{t-1};\gamma)\right| \leq TB\chi\delta.
$$

Putting together (12) and (14) delivers the result. *Q.E.D.*

REFERENCES

APOSTOL, T. M. (1974): *Mathematical Analysis* (Second Ed.). Reading, MA: Addison–Wesley.

BAJARI, P., H. HONG, AND S. RYAN (2004): "Identification and Estimation of Discrete Games of Complete Information," Mimeo, Michigan University.

BENHABIB, J., AND R. E. A. FARMER (1999): "Indeterminacy and Sunspots in Macroeconomics," in *The Handbook of Macroeconomics*, ed. by J. B. Taylor and M. Woodford. Amsterdam: North-Holland, 387–448.

BOLDRIN, M., AND M. WOODFORD (1990): "Equilibrium Models Displaying Endogenous Fluctuations and Chaos: A Survey," *Journal of Monetary Economics*, 25, 189–222.

BRIN, M., AND Y. KIFER (1987): "Dynamics of Markov Chains and Stable Manifolds for Random Diffeomorphisms," *Ergodic Theory and Dynamical Systems*, 7, 351–374.

CRAWFORD, G. S., AND M. SHUM (2005): "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, 73, 1137–1173.

DAULA, T., AND R. MOFFITT (1995): "Estimating Dynamic Models of Quit Behavior: The Case of Military Reenlistment," *Journal of Labor Economics*, 13, 499–523.

DEJONG, D. N., B. F. INGRAM, AND C. H. WHITEMAN (2000): "A Bayesian Approach to Dynamic Macroeconomics," *Journal of Econometrics*, 98, 203–223.

DEPREE, J. D., AND C. W. SWARTZ (1988): *Introduction to Real Analysis*. New York: John Wiley & Sons.

DIEUDONNÉ, J. (1960): *Treatise on Analysis*, Vol. 1. New York: Academic Press.

DUFFIE, D., AND K. J. SINGLETON (1993): "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929–952.

ELLISON, G., AND D. FUDENBERG (1993): "Rules of Thumb for Social Learning," *Journal of Political Economy*, 101, 612–643.

FERNÁNDEZ-VILLAVERDE, J., AND J. F. RUBIO-RAMÍREZ (2005): "Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood," *Journal of Applied Econometrics*, forthcoming.

FERNÁNDEZ-VILLAVERDE, J., J. F. RUBIO-RAMÍREZ, AND M. S. SANTOS (2005): "Convergence Properties of the Likelihood of Computed Dynamic Models," Working Paper, University of Pennsylvania. Available at *www.econ.upenn.edu/~jesusfv*.

FLINN, C. C., AND J. HECKMAN (1982): "New Methods for Analyzing Structural Models of Labor Force Dynamics," *Journal of Econometrics*, 18, 115–168.

FOLEY, D., AND M. HELLWIG (1975): "Asset Management with Trading Uncertainty," *Review of Economic Studies*, 42, 327–346.

FUTIA, C. (1982): "Invariant Distributions and the Limiting Behavior of Markovian Economic Models," *Econometrica*, 50, 377–408.

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

GILLESKIE, D. B. (1998): "A Dynamic Stochastic Model of Medical Care Use and Work Absence," *Econometrica*, 66, 1–45.

JOFRE-BONET, M., AND M. PESENDORFER (2003): "Estimation of a Dynamic Auction Game," *Econometrica*, 71, 1443–1489.

KEANE, M. P., AND R. MOFFITT (1998): "A Structural Model of Multiple Welfare Program Participation and Labor Supply," *International Economic Review*, 39, 553–589.

KEANE, M. P., AND K. I. WOLPIN (1994): "The Simulation and Estimation of Discrete Choice Programming Models by Simulation and Interpolation: Monte Carlo Evidence," *Review of Economics and Statistics*, 76, 648–672.

——— (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105, 473–522.

KEHOE, T., AND D. K. LEVINE (1985): "Comparative Statics and Perfect Foresight in Infinite Horizon Economies," *Econometrica*, 53, 433–453.

LUBIK, T., AND F. SCHORFHEIDE (2004): "Testing for Indeterminacy: An Application to U.S. Monetary Policy," *American Economic Review*, 94, 190–217.

MAS-COLELL, A. (1985): *The Theory of General Economic Equilibrium: A Differentiable Approach*. Econometric Society Monographs. Cambridge, U.K.: Cambridge University Press.

PAKES, A. (1986): "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 54, 755–785.

PAKES, A., M. OSTROVSKY, AND S. BERRY (2004): "Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Examples)," Mimeo, Harvard University.

PERES, Y., AND B. SOLOMYAK (1996): "Absolute Continuity of Bernoulli Convolutions: A Simple Proof," *Mathematical Research Letters*, 3, 231–239.

RABANAL, P., AND J. F. RUBIO-RAMÍREZ (2005): "Comparing New Keynesian Models of the Business Cycle: A Bayesian Approach," *Journal of Monetary Economics*, forthcoming.

ROSENZWEIG, M. R., AND K. I. WOLPIN (1993): "Credit Market Constraints, Consumption Smoothing, and the Accumulation of Durable Production Assets in Low-Income Countries: Investments in Bullocks in India," *Journal of Political Economy*, 101, 223–244.

RUST, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55, 999–1033.

RUST, J., AND C. PHELAN (1997): "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets," *Econometrica*, 65, 781–831.

SANGHVI, A. P., AND M. J. SOBEL (1976): "Bayesian Games as Stochastic Processes," *International Journal of Game Theory*, 5, 1–22.

SANTOS, M. S. (2000): "Accuracy of Numerical Solutions Using the Euler Equation Residuals," *Econometrica*, 68, 1377–1402.

——— (2004): "Simulation-Based Estimation of Dynamic Models with Continuous Equilibrium Solutions," *Journal of Mathematical Economics*, 40, 465–491.

SANTOS, M. S., AND A. PERALTA-ALVA (2005): "Accuracy of Simulations for Stochastic Dynamic Models," *Econometrica*, forthcoming.

SARGENT, T. J. (1989): "Two Models of Measurements and the Investment Accelerator," *Journal of Political Economy*, 97, 251–287.

SCHENK-HOPPÉ, K. R., AND B. SCHMALFUSS (2001): "Random Fixed-Points in a Stochastic Solow Growth Model," *Journal of Mathematical Economics*, 36, 19–30.

SCHMALENSEE, R. (1975): "Alternative Models of Bandit Selection," *Journal of Economic Theory*, 10, 333–342.

SCHORFHEIDE, F. (2000): "Loss Function-Based Evaluation of DGSE Models," *Journal of Applied Econometrics*, 15, 645–670.

SMETS, F., AND R. WOUTERS (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1, 1123–1175.

STACHURSKI, J. (2002): "Stochastic Optimal Growth with Unbounded Shock," *Journal of Economic Theory*, 106, 45–60.

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

STOKEY, N. L., R. E. LUCAS, AND E. C. PRESCOTT (1989): *Recursive Methods in Economic Dynamics*. Cambridge, MA: Harvard University Press.

VUONG, Q. H. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypothesis," *Econometrica*, 57, 307–333.

WOLPIN, K. I. (1984): "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality," *Journal of Political Economy*, 92, 852–874.