TECHNICAL WORKING PAPER SERIES

# CONVERGENCE PROPERTIES OF THE LIKELIHOOD OF COMPUTED DYNAMIC MODELS

Jesús Fernández-Villaverde
Juan F. Rubio-Ramírez
Manuel S. Santos

Technical Working Paper 315
http://www.nber.org/papers/T0315

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2005

We thank the 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 the project SES-0338997. Any views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta or of the Federal Reserve System. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Convergence Properties of the Likelihood of Computed Dynamic Models
Jesús Fernández-Villaverde, Juan F. Rubio-Ramírez, and Manuel S. Santos
NBER Technical Working Paper No. 315
October 2005
JEL No. C11, C15, D9, E10, E32


## ABSTRACT

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.



Jesús Fernández-Villaverde
University of Pennsylvania
Department of Economics
160 McNeil Building
3718 Locust Walk
Philadelphia, PA 19104-6297
and NBER
jesusfv@econ.upenn.edu

Juan F. Rubio-Ramírez
Federal Reserve Bank of Atlanta
juan.rubio@atl.frb.org

Manuel S. Santos
manuel.santos@asu.edu

## 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, IO, 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 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 IMF, 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. Section 1 presents one particularly simple example of non-convergence. 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 argmax and lim operators. But, we can impose mildly 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).

A related issue that is different from the focus of this paper is how to evaluate the likelihood when that likelihood function is intractable given some policy rules. This evaluation is usually performed by simulation methods (see Gouriéroux and Monfort, (1996)). Pakes and Pollard (1989) provide results regarding the convergence and asymptotics of simulation estimators. Of course, both problems can exist at the same time: We may need to approximate the decision rule of the agents and, even with that approximation, resort to simulation methods to evaluate the likelihood. This would be the case, for example, if we want to evaluate the likelihood function of the neoclassical growth model when the solution method is nonlinear.

The rest of the paper is organized as follows. Section 2 presents an example where the sequence of approximated likelihoods does not converge to the exact likelihood. Section 3 fixes an environment to discuss the convergence of the likelihood. Our main result concerning convergence is contained in section 4. Section 5 narrows down the speed of convergence, its relation to the sample size, and proposes a likelihood ratio rest to diagnose inference problems. Section 6 presents our findings regarding the convergence of maximum likelihood point estimates. Section 7 studies an application that confirms that the results of the paper hold in practice. Section 8 concludes. An appendix gathers all the proofs of the results in the paper.

## 2. An Example of Nonconvergence

We begin by presenting an example to show how, in general, we cannot assume the convergence of the approximated likelihood to the exact likelihood. This example is built around a discrete policy function. This policy function will be approximated in such a way that the sequence of approximated policy functions converges to the exact policy function but the sequence of approximated likelihoods does not. Our example may not surprise the reader since the convergence failure relies on the presence of multiple ergodic sets. However, it illustrates in a transparent environment our point that we cannot just assume the convergence of the likelihood.

Let us think about the following dynamic discrete choice problem. An agent has to choose the current state $S_t$ among three possible states $S = \{1, 2, 3\}$. After choosing the state, the agent gets a random endowment $y_t = \varepsilon_{i,t}$ if $S_t = i$, where $\varepsilon_{i,t}$ is normally distributed with standard deviation $\sigma_i$. The period utility function is $u(y_t, S_t, S_{t-1})$. This utility depends on the current endowment, the current state $S_t$, and on the state $S_{t-1}$ chosen last period. The presence of $S_{t-1}$ links the current choice with future payoffs, which are discounted at rate $\beta$. Also, the agent has access to a randomization device.

The utility function, the discount factor, and the randomization device are such that the exact policy function of the agent is given by:

$$\varphi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

This policy function is interpreted as follows. If the agent chose state 1 in the last period, she will choose state 1 in the current period with probability 1 (first row of the matrix). If the agent chose state 2 in the last period, she will choose state 2 with probability 1/2 and state 3 with probability 1/2 (second row of the matrix). The agent will behave in the same way if she chose state 3 in the last period (last row of the matrix).[1]

[1] It is possible to find utility functions and discount factors that imply this policy function. We omit details in the interest of space.

This policy function generates two ergodic distributions for the states of the economy: $(1,0,0)$ and $\left(0,\frac{1}{2},\frac{1}{2}\right)$. The presence of these two nonoverlapping ergodic distributions implies that, in order to write the likelihood function, we need to specify where the initial state of the economy $S_0$ is coming from. Following Lubik and Schorfheide (2004), we assume that there is a sunspot that picks one of the two distributions. The sunspot has probability $\pi_A$ to signal the first ergodic distribution and $\pi_B$ to signal the second (where $\pi_A + \pi_B = 1$).

If the economist observes a sequence of endowments $y^T$, the likelihood conditional on the first state $S_0$ of those data is:

$$L\left(y^T;\gamma,S_0=i\right)=\begin{cases}\prod_{t=1}^{T}\phi\left(\frac{y_t}{\sigma_1}\right) & \text{if } S_0=1\\ \frac{\prod_{t=2}^{T}\left(\phi\left(\frac{y_t}{\sigma_2}\right)+\phi\left(\frac{y_t}{\sigma_3}\right)\right)}{2}\phi\left(\frac{y_t}{\sigma_2}\right) & \text{if } S_0=2\\ \frac{\prod_{t=2}^{T}\left(\phi\left(\frac{y_t}{\sigma_2}\right)+\phi\left(\frac{y_t}{\sigma_3}\right)\right)}{2}\phi\left(\frac{y_t}{\sigma_3}\right) & \text{if } S_0=3\end{cases} \tag{1}$$

where $\gamma$ is the vector of parameters of the model and $\phi(\cdot)$ is the standardized normal density. Equation (1) and the sunspot distribution imply that the likelihood function equals:

$$L\left(y^T;\gamma\right)=L\left(y^T;\gamma,S_0=1\right)\pi_A+\left(\frac{L\left(y^T;\gamma,S_0=2\right)}{2}+\frac{L\left(y^T;\gamma,S_0=3\right)}{2}\right)\pi_B \tag{2}$$

Now let us assume that the economist cannot compute the exact policy function $\varphi$, but only an approximated policy function $\varphi_j$ of the form:

$$\varphi_j=\begin{bmatrix}1-\delta_j & \frac{\delta_j}{2} & \frac{\delta_j}{2}\\ 0 & \frac{1}{2} & \frac{1}{2}\\ 0 & \frac{1}{2} & \frac{1}{2}\end{bmatrix}$$

where $0<\delta_j<1$ is the maximum absolute error in the approximation of the policy function and $j$ is an index of the accuracy of the approximation. Suppose that the solution method is such that the economist computes exactly $y_t=\varepsilon_{i,t}$ if $S_t=i$ for all $j$ and that $\delta_j\to 0$ as $j\to\infty$. Given these two properties of the solution method, no matter how good our approximated policy function is (i.e., no matter how small $\delta_j$ is), the ergodic distribution for

the states of the economy is $\left(0, \frac{1}{2}, \frac{1}{2}\right)$ for all $j$.

Therefore, the economist evaluates the approximated likelihood:

$$L_j\left(y^T;\gamma\right) = \frac{L\left(y^T;\gamma, S_0 = 2\right)}{2} + \frac{L\left(y^T;\gamma, S_0 = 3\right)}{2} \tag{3}$$

for all $j$. Comparing (2) and (3), we see that, as $\delta_j \to 0$, $\varphi_j \to \varphi$, but $L_j\left(y^T;\gamma\right) \nrightarrow L\left(y^T;\gamma\right)$.

This example shows that the sequence of approximated likelihoods may fail to converge. Hence, it is important to find conditions that ensure the convergence of the approximated likelihoods and to assess the rate of convergence. Also, this example presents several elements that will be important in our results: the continuity of the exact policy function, the convergence of the sequence of approximated policy functions, the maximum error of the approximated policy function, and the stationary distribution of states of the economy.

## 3. 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)):

$$S_t = \varphi\left(S_{t-1}, W_t; \gamma\right) \tag{4}$$

$$Y_t = g\left(S_t, V_t; \gamma\right). \tag{5}$$

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 $S \subset R^l$. Often, we will use the measurable space $(S, \mathcal{S})$ where $\mathcal{S}$ is the Borel $\sigma - field$. The variables $W_t$ and $V_t$ are $i.i.d.$ 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, $\gamma$, which belongs to the compact set $\Upsilon \subset R^n$, 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 (4) is known as the transition equation, since it governs the evolution of states over time. Equation (5) 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 $W_{2,t} = W_t$ for $\forall t$, 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 $S^T = \{S_t\}_{t=0}^T$ for $\forall T$. Let $y^T$ be a realization of the random variable $Y^T$ for $\forall T$. 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 $\gamma$, 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 a unique 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 is 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 $\gamma$, there exists a unique invariant distribution for $S$, $\mu^*(S;\gamma)$, that has a Radon-Nykodim derivative with respect to the Lebesgue measure.*

Let us also make the following assumption:

**Assumption 3.** *For all $\gamma$ and $t$, the following system of equations:*

$$\begin{aligned} S_1 &= \varphi(S_0, (W_{1,1}, W_{2,1}); \gamma) \\ y_m &= g(S_m, V_m; \gamma) \qquad \text{for } m = 1, 2, ..., t \\ S_m &= \varphi(S_{m-1}, (W_{1,m}, W_{2,m}); \gamma) \qquad \text{for } m = 2, 3, ..., t \end{aligned}$$

*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 $\gamma$ 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 $\gamma$ 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 $\gamma$ is:

$$L(y^T; \gamma) = \prod_{t=1}^{T} p(y_t|y^{t-1}; \gamma) = \prod_{t=1}^{T} \int \int p(y_t|W_1^t, S_0, y^{t-1}; \gamma) p(W_1^t, S_0|y^{t-1}; \gamma) dW_1^t dS_0. \quad (6)$$

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 $\gamma$ 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 (6) in the following recursive way:

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

for all $\gamma$. This structure will be useful for proving our results in the next sections.

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

## 4. Convergence of the Likelihood

If the researcher knows the transition and measurement equations, $\varphi\left(\cdot,\cdot;\gamma\right)$ and $g\left(\cdot,\cdot;\gamma\right)$, the evaluation of the likelihood function (6) 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, $\varphi_j\left(\cdot,\cdot;\gamma\right)$ and $g_j\left(\cdot,\cdot;\gamma\right)$. 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 $\varphi_j\left(\cdot,\cdot;\gamma\right)$ and $g_j\left(\cdot,\cdot;\gamma\right)$ converge to their exact values as $j$ goes to infinity. For example, the dynamic programming algorithm allows for more vertex points, perturbation approaches for a higher order of the expansion, and projection methods for more basis functions.

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

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

given $\gamma$, $L_j\left(y^T;\gamma\right) \to L\left(y^T;\gamma\right)$ 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 $\gamma$ and $t$, the conditional probability $p\left(y_t|W_1^t, S_0, y^{t-1};\gamma\right)$ 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 $t$, $p\left(y_t|W_1^t, S_0, y^{t-1};\gamma\right) \in C\left(W_1^t, S_0\right)$.*

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

We now prove that for all $j$, $\gamma$, and $t$, the conditional probability $p_j\left(y_t|W_1^t, S_0, y^{t-1};\gamma\right)$ 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 parallel structure to that on our previous section.

First, we assume:

**Assumption 5.** *For all $j$ and $\gamma$, the functions $\varphi_j(\cdot,\cdot;\gamma)$ and $g_j(\cdot,\cdot;\gamma)$ are continuous. For all $j$ and $\gamma$, $\varphi_j(\cdot,\cdot;\gamma)$ 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 $\gamma$, the invariant distribution for $S$, $\mu_j^*(S;\gamma)$, has a Radon-Nykodim 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 $j$, $\gamma$, and $t$, the following system of equations:*

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

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

As before, assumptions 5 and 7 imply that for all $j$, $\gamma$, 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 $\left(v_j^t,s_j^t,w_{j,2}^t\right)$ rather than the cumbersome expression $\left(v_j^t(W_1^t,S_0,y^t;\gamma), s_j^t(W_1^t,S_0,y^t;\gamma), w_{j,2}^t(W_1^t,S_0,y^t;\gamma)\right)$. 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 $j$, $\gamma$, 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 (PMLE) of the approximated model as $\widehat{\gamma}_j\left(y^T\right) \equiv \arg\max_{\gamma\in\Upsilon} p_j\left(y^T;\gamma\right)$ and require the approximated model to explain the data even if it does so with arbitrarily low probability:

**Assumption 8.** *For all $j$, $\gamma$, 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 $L_j\left(y^T;\gamma\right)$ is bounded.

**Lemma 2.** *Let $\gamma \in \Upsilon$. Under assumptions 5 and 7, for all $j$ and $t$, $p_j\left(y_t|W_1^t, S_0, y^{t-1};\gamma\right) \in C\left(W_1^t, S_0\right)$ 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 $\gamma \in \Upsilon$, $\varphi_j\left(\cdot,\cdot;\gamma\right) \to \varphi\left(\cdot,\cdot;\gamma\right)$. Then, under assumptions 1 and 5, every sequence of invariant distributions $\left\{\mu_j^*\left(S;\gamma\right)\right\}$ converges weakly to the invariant distribution $\mu^*\left(S;\gamma\right)$ associated with $\varphi\left(\cdot,\cdot;\gamma\right)$.*

### 4.1. Main Result: Convergence of the Likelihood Function

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

**Assumption 9.** *For all $j$ and $\gamma$, $\varphi_j\left(\cdot,\cdot;\gamma\right)$ and $g_j\left(\cdot,\cdot;\gamma\right)$ 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 to 9, if $\varphi_j(\cdot,\cdot;\gamma) \to \varphi(\cdot,\cdot;\gamma)$ and $g_j(\cdot,\cdot;\gamma) \to g(\cdot,\cdot;\gamma)$, then*

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

The result is key for applied work. It states that for any given $\gamma$, 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.

### 4.2. 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, that 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, while 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-Nykodim 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 univocally 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-Nykodim 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 to 8. Finally, assumption 9 is fundamental for lemma 4.

### 4.3. 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 $\gamma$ and $\gamma'$, the ratio of likelihood functions converges.

[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 falls 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, Chapters 11 and 12).

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 $\gamma$ and $\gamma'$, it follows that:*

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

The second application of the result directly affects Bayesian inference. There are two main objects of interest in the Bayesian paradigm: the marginal likelihood of the model and the posterior of the parameters. The marginal likelihood of the model is defined as $p\left(y^T\right) = \int_\Upsilon L\left(y^T;\gamma\right)\pi\left(\gamma\right)d\gamma$, while the marginal likelihood of the approximated model is $p_j\left(y^T\right) = \int_\Upsilon L_j\left(y^T;\gamma\right)\pi\left(\gamma\right)d\gamma$. 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)).

Given that $L\left(y^T;\gamma\right)$ and $L_j\left(y^T;\gamma\right)$ are bounded, 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 $\gamma$, it follows that $p_j\left(y^T\right) \to p\left(y^T\right)$.*

Given some prior distribution of the parameters, $\pi\left(\gamma\right)$, the posterior is given by $p\left(\gamma|y^T\right) \propto L\left(y^T;\gamma\right)\pi\left(\gamma\right)$ for the exact likelihood, and $p_j\left(\gamma|y^T\right) \propto L_j\left(y^T;\gamma\right)\pi\left(\gamma\right)$ 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 $\gamma$, it follows that $p_j\left(\gamma|y^T\right) \to p\left(\gamma|y^T\right)$.*

The posterior distribution of the parameters of the model –beyond expressing our conditional belief– is also useful for evaluating expectations of the form $E\left(h\left(\gamma\right)|y^T\right)$, in which $h\left(\gamma\right)$ 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 $E\left(h\left(\gamma\right)|y^T\right) = \frac{1}{p(y^T)}\int_\Upsilon h\left(\gamma\right) L\left(y^T;\gamma\right)\pi\left(\gamma\right)d\gamma$ and the approximated model $E_j\left(h\left(\gamma\right)|y^T\right) = \frac{1}{p_j(y^T)}\int_\Upsilon h\left(\gamma\right) L_j\left(y^T;\gamma\right)\pi\left(\gamma\right)d\gamma$. Then:

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

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 argmax and limit operator. Below, we provide additional assumptions to prove uniform convergence of the likelihood function and, consequently, to prove the convergence of the PMLE $\widehat{\gamma}_j\left(y^T\right) \to \widehat{\gamma}\left(y^T\right)$.

## 5. Speed of Convergence of the Likelihood

The goal of this section is to analyze the speed of convergence of the approximated likelihood function, $L_j\left(y^T;\gamma\right)$, to the exact likelihood function, $L\left(y^T;\gamma\right)$ for a fixed $\gamma$. Given a bound for the difference between the approximated and exact transition and measurement equations, $\left\|\varphi_j\left(\cdot,\cdot;\gamma\right) - \varphi\left(\cdot,\cdot;\gamma\right)\right\| \leq \delta$ and $\left\|g_j\left(\cdot,\cdot;\gamma\right) - g\left(\cdot,\cdot;\gamma\right)\right\| \leq \delta$, we will obtain a bound for the difference between the approximated and exact likelihood functions $\left|\prod_{t=1}^T p_j\left(y_t|y^{t-1};\gamma\right) - \prod_{t=1}^T p\left(y_t|y^{t-1};\gamma\right)\right|$.

Let us introduce some additional assumptions needed in the section:

**Assumption 10.** *For all $\gamma$, the densities of $W_t$ and $V_t$ are differentiable, with bounded partial derivatives.*

**Assumption 11.** *For all $\gamma$, $\varphi\left(\cdot,\cdot;\gamma\right)$ and $g\left(\cdot,\cdot;\gamma\right)$ 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, $p\left(y_t|W_1^t, S_0, y^{t-1};\gamma\right)$ is continuously differentiable with bounded partial derivatives with respect to $S_0$ for all $t$.*

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

Once we have the continuity and differentiability of $p\left(y_t|W_1^t, S_0, y^{t-1}; \gamma\right)$, the next step is to bound the difference $\left|p_j\left(y_t|W_1^t, S_0, y^{t-1}; \gamma\right) - p\left(y_t|W_1^t, S_0, y^{t-1}; \gamma\right)\right|$. This difference will be a key component when we evaluate the differences between likelihoods. We then parameterize both $\varphi_j\left(\cdot, \cdot; \gamma\right) = \varphi\left(\cdot, \cdot; \gamma, \theta_j\right)$, and $g_j\left(\cdot, \cdot; \gamma\right) = g\left(\cdot, \cdot; \gamma, \theta_j\right)$, where $\theta_j \in \Phi$, $\forall j$, for a compact subset $\Phi \in R^M$, in such a way that they have bounded partial derivatives with respect to $\theta$, 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 below. 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 to 11, if $\left\|\varphi_j\left(\cdot, \cdot; \gamma\right) - \varphi\left(\cdot, \cdot; \gamma\right)\right\| \leq \delta$ and $\left\|g_j\left(\cdot, \cdot; \gamma\right) - g\left(\cdot, \cdot; \gamma\right)\right\| \leq \delta$, then there exists a positive constant $\chi$ such that for all $t$:*

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

*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 $\varphi$, which is equivalent to their Condition C.

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

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

*for all $S$, $S'$, and where $Q\left(\cdot; \gamma\right)$ 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 close-form solution.

Now we are ready to prove the main result of the 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 to 11, if $\left\|\varphi_j(\cdot,\cdot;\gamma) - \varphi(\cdot,\cdot;\gamma)\right\| \leq \delta$ and $\|g_j(\cdot,\cdot;\gamma) - g(\cdot,\cdot;\gamma)\| \leq \delta$, there are some positive constants $B$ and $L$ such for all $T$ :*

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

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 $\delta$.[4]

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 $\delta$ 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 $\chi$, can be estimated from repeated solutions of the model under different numerical approximations. Consequently, the researcher can bound the difference between the exact and approximated likelihood and use proposition 2 as a guide to determine the accuracy $\delta$ that she needs to ask from her solution method.

Second, in order to guarantee asymptotic convergence in the estimation of dynamic mod-

[4]Santos (2000) shows that for a class of dynamic optimization problems, the approximation error of the policy function $\delta$ is of the same order of magnitude as the Euler equation residual. Since Euler errors are easy to estimate, we can replace $\delta$ by an Euler error estimate and obtain a bound of the same order of magnitude.

els, 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 likelihood 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 $\delta$ and $\delta'$ such that $\delta < \delta'$. Then, for observations $y^T$, we can compute:

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

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

Vuong (1989) develops the asymptotic behavior of this statistic to compare competing models that are non-nested, overlapping, or nested and 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 our 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 $\frac{p_j\left(y^T\right)}{p_{j'}\left(y^T\right)}$ 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.

## 6. Convergence of the Maximum Likelihood Estimates

In section 4 we established the convergence of the approximated likelihood function and the convergence of Bayesian estimates. However, we mentioned that we could not guarantee the convergence of the PMLE. The reason is that, under pointwise convergence, we cannot, in general, swap the lim and the argmax operators. To fill this gap in our analysis, this section provides some conditions under which the PMLE of the approximated likelihood function, $\widehat{\gamma}_j\left(y^T\right)$, will converge to the PMLE of the exact likelihood function, $\widehat{\gamma}\left(y^T\right)$. In particular, we show that if the policy functions converge uniformly in the parameter space, i.e., for any $\delta$, there is an $N$ such that $\forall j \geq N$, $\left\|\varphi_j\left(\cdot,\cdot;\cdot\right) - \varphi\left(\cdot,\cdot;\cdot\right)\right\| \leq \delta$ and $\left\|g_j\left(\cdot,\cdot;\cdot\right) - g\left(\cdot,\cdot;\cdot\right)\right\| \leq \delta$ for all $S, W, V$, and $\gamma$, then the likelihood function also converges uniformly, implying the convergence of the PMLE.

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

Our first step is to show that if the policy functions converge uniformly in the parameter space, then $p_j\left(y_t|W_1^t, S_0, y^{t-1}; \gamma\right)$ converges uniformly to $p\left(y_t|W_1^t, S_0, y^{t-1}; \gamma\right)$. To accomplish this goal, we restrict the way in which $\gamma$ can enter the densities of $W_t$ and $V_t$:

**Assumption 12.** *The densities of $W_t$ and $V_t$ are continuous with respect to $\gamma$.*

Analogously, we modify assumptions 1, 5, 10, and 11:

**Assumption 13.** *The bounds in assumptions 1, 5, 10, and 11 are independent of $\gamma$.*

And, finally, we substitute the previous parametrization of the approximated transition and measurement equations by the following new assumption:

**Assumption 14.** *For all $j$, $\varphi_j\left(\cdot,\cdot;\cdot\right)\left(=\varphi\left(\cdot,\cdot;\cdot,\theta_j\right)\right)$ and $g_j\left(\cdot,\cdot;\cdot\right)\left(=g\left(\cdot,\cdot;\cdot,\theta_j\right)\right)$ have bounded partial derivatives with respect to $\theta$, as a function of $S$, $W$, $V$, and $\gamma$. The bounds are independent of $j$.*

These three new assumptions guarantee that all the bounds are uniform on $\gamma$. The main practical consequence of these new assumption is that the researcher needs to check the appropriate behavior of the policy function of the model over the whole space of parameter values of interest and not just at one particular value of the parameters. This may turn difficult if the theoretical behavior of the model changes over the parameter space: for example, if the policy functions become discontinuous when one particular parameter gets larger.

Armed with our stronger assumptions, we can modify lemma 6 to get:

**Lemma 7.** *Under assumptions 1 to 11, and assumptions 12 to 14, if the policy functions converge uniformly in the parameter space, then there is an $N$ such that $\forall j \geq N$:*

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

*for all $\gamma$, $S_0$, and $W_1^t$ but a finite number of points and for some finite constant $\chi$.*

We can also modify proposition 2 to get:

**Proposition 3.** *Let condition 1 hold. Under assumptions 1 to 11, and assumptions 12 to 14, if the policy functions converge uniformly in the parameter space, then there is an $N$ such that $\forall j \geq N$:*

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

*for all $\gamma$ and for some finite constant $B$ and $L$.*

Proposition 3 implies that if the policy functions converge uniformly in the parameter space, then the approximated likelihood function also converges uniformly to the exact likelihood function. Uniform convergence of the likelihood function implies convergence of the maximizer and, therefore, of the PMLE. Formally:

**Corollary 5.** *Let condition 1 hold. Under assumptions 1 to 11, and assumption 14, if the policy functions converge uniformly in the parameter space, then $\widehat{\gamma}_j \left( y^T \right) \to \widehat{\gamma} \left( y^T \right)$.*

Finally, note that even with uniform convergence, we cannot deliver the convergence of the partial derivatives of the approximated likelihood function. This problem limits our ability to interpret the standard errors and confidence intervals built under classical methods.

## 7. An Application

We now present 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 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 the 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 policymaking. 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 log utility and full depreciation. Why do we pick this particular version of the model? Because for this case we know the exact likelihood since the model has a closed-form solution in logs suitable to evaluation with the Kalman filter. Thus, we have the exact likelihood to compare against the approximated likelihood implied by different numerical approximations.

We analyze the case when the researcher does not know that the model has an exact closed-form solution. Instead, the researcher solves for the optimal policy functions using value function iteration and evaluate the likelihood of the model with a Particle filter (see Fernández-Villaverde and Rubio-Ramírez, (2004), for a description of this filter).

Note that the value function iteration and the Particle filter are only used as numerical procedures to compute two unknown (from the researcher's perspective) functions: the policy function and the likelihood function of the model. If we allowed the number of grid points in the value function iteration and the number of particles in the Particle filter to go to infinity, the combination of value function iteration and the Particle filter would deliver *exactly* the same likelihood function than the result of applying the exact loglinear solution of the model and the Kalman filter.

In practice, since we will have a finite number of grid points and a finite number of particles, the value of the likelihood will be affected by two approximation errors: one in the computation of the optimal policy function and a second one in the Particle filter. However,

we show below that this second error is orders of magnitude smaller than the first error. Thus, the differences in likelihoods and inference documented in the application are fundamentally the result of the approximation error in the computation of the policy function, our object of interest.

We could have studied a more sophisticated case where, instead of having full depreciation, the neoclassical growth model is calibrated to match basic characteristics of the U.S. data. This exercise would have the advantage of being more realistic. However, it would imply that we cannot evaluate the exact likelihood of the model since the model does not have a closed-form solution. Consequently, we could not compare the exact likelihood function with the approximated one, limiting the usefulness of the example.

### 7.1. The Neoclassical Growth Model

Let $y^T = (y_0, ..., y_T)$ be some given data, where $y_t \in R^2$ for all $0 \leq t \leq T$. The components of $y_t$ are output and gross investment. We want to calculate the likelihood of data $y^T$ implied by the neoclassical growth model where, in addition, we observe $y^T$ with measurement error $V_t$. Let $V_t \sim N(0, \Lambda)$, where $\Lambda$ is a diagonal matrix with $\sigma_1^2$ and $\sigma_2^2$, as diagonal entries. We assume two observables to keep the dimensionality of the model low while we capture most of its dynamics.

In the neoclassical growth model there is a representative household whose preferences over consumption $c_t$ and leisure $1 - l_t$ are represented by the utility function:

$$U = E_0 \sum_{t=0}^{\infty} \beta^t \{\xi \log c_t + (1 - \xi) \log (1 - l_t)\}$$

where $\beta \in (0, 1)$ is the discount factor, $\xi$ pins down labor supply, and $E_0$ is the conditional expectation operator.

The only good in this economy is produced according to the Cobb-Douglas production function $e^{z_t} \lambda k_t^{\alpha} l_t^{1-\alpha}$, where $k_t$ is the aggregate capital stock, $l_t$ is the aggregate labor input, $\lambda$ is a scale parameter, and $z_t$ is the technology level. $z_t$ follows an AR(1) $z_t = \rho z_{t-1} + \epsilon_t$ with $\epsilon_t \sim \mathcal{N}(0, \sigma)$. We consider the stationary case (i.e., $|\rho| < 1$). The law of motion for capital is $k_{t+1} = i_t + (1 - \eta) k_t$, where $i_t$ is investment and $\eta$ is the depreciation factor. The economy satisfies

the resource constraint $c_t + i_t = e^{z_t} \lambda k_t^{\alpha} l_t^{1-\alpha}$. Finally, let $\gamma = \{\tau, \alpha, \beta, \rho, \xi, \eta, \lambda, \sigma, \sigma_1, \sigma_2\}$ be the vector of structural parameters of the model.

A competitive equilibrium can be defined in a standard way. Since both welfare theorems hold, we solve the equivalent and simpler social planner's problem. We can think about this problem as finding the optimal policies for consumption $c(\cdot,\cdot)$, labor $l(\cdot,\cdot)$, and next period's capital $k'(\cdot,\cdot)$ that characterize the optimal choices as functions of the two state variables, capital and the technology level.[6]

### 7.2. The Likelihood Function for a Particular Calibration

As described in the introduction, we set, unrealistically but rather conveniently for our point, $\eta = 1$. In this case the model has two important and useful features. First, the income and the substitution effect from a productivity shock to labor supply exactly cancel each other. Consequently, $l_t$ is constant and equal to:

$$l_t = l = \frac{(1-\alpha)\,\xi}{(1-\alpha)\,\xi + (1-\xi)\,(1-\alpha\beta)}$$

Second, the policy function for capital is $k_{t+1} = \alpha\beta e^{z_t} \lambda k_t^{\alpha} l^{1-\alpha}$, or in logs:

$$\log k_{t+1} = \log \alpha\beta\lambda l^{1-\alpha} + \alpha \log k_t + \rho z_{t-1} + \epsilon_t$$

These two properties allow for a closed-form solution of the model and thus, for the easy evaluation of the likelihood of the model.

We derive now the exact and the approximated likelihood function of the model.

[6] For simplicity, we omit the issue of the support of the innovations to the model. Our theorems require bounded support of their densities, while our assumption of normality of $\epsilon_t$ implies that its support is the whole real line. We can fix this problem assuming that the normal distribution is truncated above and below by a number bigger than any number that the floating point arithmetic of the computer can evaluate. Analogously, we ignore that in the computer, we can use only the computable reals instead of the real line.

### 7.2.1. The Exact Likelihood

Since the policy function for capital is linear in logs, we have the transition equation for the model:

$$\begin{pmatrix} 1 \\ \log k_{t+1} \\ z_t \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \log \alpha\beta\lambda l^{1-\alpha} & \alpha & \rho \\ 0 & 0 & \rho \end{pmatrix} \begin{pmatrix} 1 \\ \log k_t \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \epsilon_t. \tag{7}$$

As described before our observables are log output ($\log output_t$) and log investment ($\log i_t$) subject to a linearly additive measurement error $V_t = \begin{pmatrix} v_{1,t} & v_{2,t} \end{pmatrix}'$. Let $V_t \sim N(0, \Lambda)$, where $\Lambda$ is a diagonal matrix with $\sigma_1^2$ and $\sigma_2^2$, as diagonal elements. Then:

$$\begin{pmatrix} \log output_t \\ \log i_t \end{pmatrix} = \begin{pmatrix} -\log \alpha\beta\lambda l^{1-\alpha} & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ \log k_{t+1} \\ z_t \end{pmatrix} + \begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix}. \tag{8}$$

Given that (7) and (8) are linear, we can evaluate the exact likelihood of the model given some data with the Kalman filter.

### 7.2.2. The Approximated Likelihood

Now, let us suppose that the researcher does not know that the model is loglinear. Then, she solves the social planner's problem using value function iteration over a grid of points of capital and productivity. To simplify the exercise, we assume that the researcher knows $l$ and that labor is constant over time. This solution method implies a policy function for capital $k_{t+1} = g_j(z_t, k_t; \gamma)$, where $j$ denotes that this policy function is an approximation. We select value function iteration because it is one of the most commonly used nonlinear solution methods, because it satisfies our assumptions regarding the approximated transition and measurement equations, and because it is a method for which we have plenty of convergence theorems (see Santos and Vigo, (1998)). In particular, we know that as the grid gets finer, $g_j(\rho z_{t-1} + \epsilon_t, k_t; \gamma) \to \alpha\beta e^{z_t} \lambda k_t^{\alpha} l^{1-\alpha}$.

The approximated likelihood function to evaluate is derived from the system:

$$\begin{aligned} k_{t+1} &= g_j\left(\rho z_{t-1} + \epsilon_t, k_t; \gamma\right) \\ z_t &= \rho z_{t-1} + \epsilon_t \end{aligned} \tag{9}$$

and

$$\begin{pmatrix} \log output_t \\ \log i_t \end{pmatrix} = \begin{pmatrix} -\log \alpha\beta\lambda l^{1-\alpha} \\ 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \log k_{t+1} \\ z_t \end{pmatrix} + \begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix}. \tag{10}$$

Given that (9) and (10) constitute a nonlinear system, we evaluate the approximated likelihood of the model given some data with a Particle filter.

The use of the Particle filter is only due to the fact that the approximated policy function $k_{t+1} = g_j\left(z_t, k_t; \gamma\right)$ generated by value function iteration is nonlinear, even if the exact policy function is in fact (log-) linear.

### 7.3. Differences in Likelihoods

How different are the likelihoods of the approximated and of the exact model? To answer this question, we generate a sample size of 200 observations from the exact loglinear model with the calibration: $\alpha = 0.4$, $\beta = 0.99$, $\lambda = 1$, $\rho = 0.95$, $\xi = 0.356$, $\sigma = 0.007$, $\sigma_1 = 0.001$, $\sigma_2 = 0.002$, and the scale factor $\lambda$ to get $\lambda l_t^{1-\alpha} = 1$.

We emphasize two points. First, the sample size of 200 observations aims to replicate the average size of U.S. macro data. In that sense our goal is to show how the problems hinted by our theoretical results may appear in a real life context.[7] Second, the exact model from which we generate our sample is *loglinear*. Consequently, *none* of the results in our application will depend on the presence of nonlinearities. A researcher that test for nonlinearities in the simulated data will not find any since, by construction, those nonlinearities do not exist in our data.

[7] Of course, some applications, for example in finance, have much longer sample sizes since data is collected every day or even by every transaction. The problems of approximating the likelihood would be much more severe on those circumstances.

We solve the model using three different grids: a coarse one with 2,000 points (50 in the capital axis and 40 points along the technology axis), an intermediate grid with 4,000 points (100 in the capital axis and 40 points along the technology axis), and fine grid with 40,000 points (1,000 in the capital axis and 40 points along the technology axis). Along the technology axis we evaluate the corresponding integral using quadrature. We keep fixed the number of points along the technology axis to illustrate more sharply how a refinement of the policy function along one particular dimension improves the likelihood. Given this parameterization, the maximum value of the absolute difference between the exact and approximated policy functions, $\delta$, takes the values:

Table 7.1: $\delta$ as a Function of the Grid Size

| Grid Points | $\delta$ |
|---|---|
| 2,000 | 0.001357 |
| 4,000 | 0.000674 |
| 40,000 | 0.000076 |

To interpret this number it is useful to think about its welfare implications. Even with only 2,000 points in the grid, the optimization problem is sufficiently well behaved that the welfare loss from using the approximated policy rule instead of the exact loglinear policy function is less than one-twentieth of a percent in terms of consumption.

The use of 2,000 points (50 in the capital axis and 40 points along the technology axis) may look too few for a standard macroeconomic application. However, in the literature of estimation of structural models, it is often the case that, because of computational reasons, a small number of grid points is used. To estimate the model, the value function needs to be solved repeatedly for different parameter values. Consequently, a relative small number of grid points is a common choice. See, for example, Keane and Wolpin (1994) and their discussion concerning the number of grid points.

Figure 1 plots the absolute value difference between the log likelihoods of the exact and the approximated model $\left|\log L\left(y^T;\gamma\right) - \log L_j\left(y^T;\gamma\right)\right|$ as a function of the sample size for the three grids. To minimize the impact of the error coming from the Particle filter, we created a swarm of 100,000 particles, well beyond the 20,000 required to achieve stability

of the estimation of the likelihood (see Fernández-Villaverde and Rubio-Ramírez, 2004, for details). In this way, the difference in the log likelihoods attributable to the approximation is several orders of magnitude bigger than difference attributable to the Particle filter.

Figure 1 illustrates the results from sections 3 and 4. Proposition 1 states that, for a fixed sample size, as $\delta$ goes to zero $\left|\log L\left(y^T;\gamma\right) - \log L_j\left(y^T;\gamma\right)\right|$ also goes to zero. This result is also confirmed by Figure 1. Proposition 2 proves that the absolute value difference between the log likelihoods of the exact and the approximated model is proportional to $\delta$. Therefore, if we reduce $\delta$ by half, the absolute value difference between the likelihoods should also be approximately reduced by half. This result is also confirmed by Figure 1.

A second implication of proposition 2 is that for a fixed $\delta$, as the sample size increases, the absolute value difference between the likelihoods increases linearly with the sample size. In addition, the slope of the increase is proportional to $\delta$. In Figure 1 we see how the larger the sample size, the larger the difference between the log likelihoods for any value of $\delta$. We plot the log differences because the size of the likelihood in levels will make the plot difficult to read. We need to remember that, in this case, a linear growth in time will be plotted as a parabola. Indeed in Figure 1, we the difference in logs grows at a decreasing rate, implying a linear rate in levels.

The surprising lesson of this figure is how bad the approximation of the likelihood is with the grid of 2,000 points even if a naive welfare comparison criterion would have suggested that the approximation was acceptable. In contrast, when we use 40,000 points, the approximated likelihood stays very close to the exact one, even at the end of the sample.

### 7.4. Impact on Inference

Can the likelihood differences documented in Figure 1 affect inference in an important way? The answer is yes.

Imagine that we have three different researchers. Each of them is trying to estimate the parameters of the neoclassical growth model and decide if the neoclassical growth model with full depreciation is a good description of the data. To do so, we give each of them the same sample of 200 observations which we generated from the exact loglinear 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 a particle filter and maximum likelihood. The only difference is that the first researcher uses the approximation of the model with 2,000 points in the grid, the second researcher the approximation with 4,000 grid points, and the third researcher the approximation 40,000 points.

All three researchers estimate the structural parameters $\gamma$ and compare the fit of their model against a simple alternative: a VAR(1) estimated with the same observables and same sample. Note that a VAR(1) is misspecified, since the dynamics of the neoclassical growth model, from which we have simulated the data, imply a VAR($\infty$).

Regarding model comparison, the researchers follow two alternatives. First, from a classical perspective, they undertake the comparison using Vuong's (1989) likelihood ratio test for model selection. as we discussed before, Vuong;s test is flexible enough to fit our needs to compare the different numerical solutions of the neoclassical growth model against each other or against a VAR(1). Second, the researchers compute Bayes factors using a degenerate prior that puts all the mass in the true parameter value. The Bayes factor also allows to compare competing models that are non-nested, overlapping, or nested and whether both, one, or neither is misspecified.

### 7.4.1. Impact on Model Comparison I: A Classical Perspective

Let $L_j\left(y^{2-200};\gamma\right)$ be the likelihood of the neoclassical growth model solved using $j$ grid points and the sample from the second to the two hundredth observation evaluated at $\gamma$. Let $G\left(y^{2-200};\theta\right)$ be the likelihood of the VAR(1) evaluated at $\theta$, where $\theta$ is the vector of parameters of the VAR. We drop the first observation since the VAR uses it to initialize the autoregression. We could evaluate the likelihood of the unconditional VAR and exploit all the 200 observations. The answers are nearly identical given that we initialize our simulated data at the deterministic steady state. Evaluating the conditional VAR makes the next explanation easier to follow. Also $L_j\left(y_t;\gamma\right)$ and $G\left(y_t;\theta\right)$ are the likelihoods of the observation at period $t$.

Define:

$$LR_{j,2-200}\left(\widehat{\gamma}_j\left(y^T\right),\widehat{\theta}\left(y^T\right)\right)=\sum_{t=2}^{200}\log\frac{L_j\left(y_t;\widehat{\gamma}_j\left(y^T\right)\right)}{G\left(y_t;\widehat{\theta}\left(y^T\right)\right)}$$

as the likelihood ratio between the neoclassical growth model and the VAR(1), where $\widehat{\gamma}_j\left(y^T\right)$ and $\widehat{\theta}\left(y^T\right)$ are the maximum likelihood estimates of $\gamma$ and $\theta$, and

$$\widehat{\omega}^2_{j,2-200} = \frac{1}{199}\sum_{t=2}^{200}\left[\log\frac{L_j\left(y_t;\widehat{\gamma}_j\left(y^T\right)\right)}{G\left(y_t;\widehat{\theta}\left(y^T\right)\right)}\right]^2 - \left[\frac{1}{199}\sum_{t=2}^{200}\log\frac{L_j\left(y_t;\widehat{\gamma}_j\left(y^T\right)\right)}{G\left(y_t;\widehat{\theta}\left(y^T\right)\right)}\right]^2$$

as its estimated variance.

Vuong shows that, under the null that:

$$H_0 : E^0\left[\sum_{t=2}^{200}\log\frac{L_j\left(y_t;\widehat{\gamma}_j\left(y^T\right)\right)}{G\left(y_t;\widehat{\theta}\left(y^T\right)\right)}\right] = 0$$

where $E^0$ is taken with respect the true data generation process (in our case the neoclassical growth model with depreciation and *exact* solution), the statistic

$$199^{-0.5}LR_{j,2-200}\left(\widehat{\gamma}_j\left(y^T\right),\widehat{\theta}\left(y^T\right)\right)/\widehat{\omega}_{j,2-200} \xrightarrow{D} \mathcal{N}\left(0,1\right).$$

We can implement this test using the data that generated Figure 1. For $\widehat{\theta}\left(y^T\right)$, we pick the maximum likelihood estimates given the sample. For $\widehat{\gamma}_j\left(y^T\right)$ we take, however, the true values we used to generate the data. In that way we eliminate the small sample problem for the neoclassical growth model.

We look first at the case of the researcher that uses 2,000 points. This researcher finds a statistic of $-4.27$ and overwhelmingly rejects the neoclassical growth model in favor of the VAR(1). Note how misleading the inference is: the data are generated by the same model that the researcher is using except that is using a slightly different policy function because of numerical reasons. Despite using the correct model, the accumulation of likelihood errors in just 199 observations is such that the researcher will reject the right model. What happens with the researcher that uses 4,000 points? In her case, she computes a statistic of -0.09 and concludes that she cannot tell the two models apart.

Finally, what will happen with the researcher that uses 40,000 points? She finds a statistic of 1.07, and she (marginally) rejects the VAR(1) in favor of the neoclassical growth model.

The third researcher is then the only one making the right inference, despite the fact that all three researchers are using the same model except for the choice of the number of grid points. Why is she making the right decision? Because the use of 40,000 grid points reduces the error in the policy function so much more than the number required by a simple welfare comparison that she is immune to the bias induced by the approximation error.

We can run a version of the likelihood ratio test between two different approximations of the neoclassical growth model. Those comparisons provide us with an example of how the likelihood ratio test helps to diagnose the problems created by the numerical approximation to the policy function. The value of the test comparing the solution with 2,000 points and with 4,000 points is -4.93, strongly indicating that 2,000 grid points are too few. The test comparing the solution with 4,000 points and the solution with 40,000 points is -1.40, also supporting (although less overwhelmingly) that the 4,000 are still not enough. Only later, when we increase the number of grid points to 40,000 the test delivers a solid answer of non-significativity: the likelihood ratio test between 40,000 points and the exact solution is only -0.11.

These results show how to use Vuong's method to select the accuracy of the numerical solution of the model. We should increase that accuracy until the value of the likelihood ratio is such that the 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 will protect against some of the worst forms of incorrect inference we documented.

### 7.4.2. Impact on Model Comparison II: A Bayesian Perspective

Now we implement the Bayesian perspective to the model comparison. In order to avoid the complication of specifying a prior for the parameter values and finding the posterior, we can assume that the researcher has a prior that puts all the mass in the true parameter value. In that way, as in the classical approach, we eliminate the small sample problem for the neoclassical growth model.

With that prior, the (log) Bayes factor is just the difference between the loglikelihoods of two competing models. Table 7.2 we give the values of the loglikelihoods for 2,000, 4,000,

and 40,000 grid points, for the exact likelihood, and for the VAR(1).

Table 7.2: Value of the Loglikelihoods

| | |
|---|---|
| 2,000 Grid Points | 1,558.53 |
| 4,000 Grid Points | 1,624.50 |
| 40,000 Grid Points | 1,632.31 |
| Exact Loglikelihood | 1,632.34 |
| VAR(1) | 1,625.24 |

A good way to read these number is to use Jeffreys' (1961) rule: if one hypothesis is more than 100 times more likely than the other, the evidence is decisive in its favor. This translates into differences in logmarginal likelihoods of 4.6 or higher between two models or two versions of a model.

This rule shows first the model with 2,000 grid points is easily defeated by the VAR(1) since the logdifference in favor of the VAR is nearly 67. Second, the model with 4,000 grid points and the VAR(1) are difficult to distinguish since the logdifference is 0.74. Finally, the model with 40,000 grid points performs well ahead of the VAR(1), with a difference of the loglikelihoods of 7.1. These results are identical to the findings from the classical test, with the partial exception that the Bayes factor provides more decisive evidence in favor the model with 40,000 grid points over the VAR(1), which was not favored beyond reasonable doubt in the likelihood ratio test because of a relatively high estimate of the variance $\widehat{\omega}_{j,2-200}$.

Also, we can see how the Bayes factor can be used as a diagnostic devise for the problems induced in the likelihood by the numerical approximation. For example, the difference in the loglikelihoods with 4,000 grid points and with 40,000 grid points, nearly 66, clearly indicates that 4,000 are not enough points in the grid. However, 40,000 grid points, which induce a difference with the exact loglikelihood of only 0.03 are enough to provide a good approximation to the likelihood.

In the same spirit as in the classical approach, these results show how we can use the Bayes factor to select the level of accuracy needed to avoid inference problems. We should increase the accuracy of the solution of the model until the Bayes factor stabilizes and cannot distinguish between two different versions of the model.

### 7.4.3. Impact on MLE

Table 7.3 reports the MLE of the parameters of the neoclassical growth model as a function of the grid points. As predicted by our results on the convergence of MLE estimates, the more refined the grid (the lower the $\delta$), the better the estimates.

Table 7.3: MLE as a Function of the Grid Size

| | Exact | 2,000 | 4,000 | 40,000 |
|---|---|---|---|---|
| $\alpha$ | 0.4000 | 0.4014 | 0.4000 | 0.4000 |
| $\beta$ | 0.9896 | 0.9862 | 0.9895 | 0.9896 |
| $\rho$ | 0.9500 | 0.9507 | 0.9506 | 0.9500 |
| $\sigma$ | 0.0070 | 0.0068 | 0.0069 | 0.0070 |

Is the inference mistake relevant? Given the point estimates for $\beta$, the researcher using 2,000 points in the grid would estimate a steady state interest rate *140* basis points higher than the exact one, a researcher using 4,000 points would only be *10* basis points off, and a researcher using 40,000 would get an almost perfect point estimate. Given the importance of the estimate of $\beta$ (and its inverse, the steady state interest rate) for policy making institutions like the Federal Reserve System, these point estimates reveal that the inference mistake induced by the approximation error could be relevant for practitioners.

## 7.5. Concluding Remarks on the Application

We have studied the consequences of using approximated likelihood functions instead of the exact likelihoods when we estimate the neoclassical growth model. We have documented the quantitative importance of our theoretical findings and how the likelihood ratio test and the Bayes factor can be used to check for accuracy of the solution.

Is this example realistic? We pick a representative model, the neoclassical growth model with 200 observations. However, we calibrated the model in an unrealistic way and we confronted it against simulated data. What will happen with real data and a realistic calibration? That case is more difficult to gauge because we cannot evaluate the exact likelihood. However, Fernández-Villaverde and Rubio-Ramírez (2005) attempt to answer this question estimating

the model linear and nonlinearly. They document differences in the values of the logmarginal likelihood of around 93 in favor of the nonlinear approach and they report important changes in point estimates of the parameters of the model. This evidence strongly suggests that the problems on inference induced by approximating the policy function appears in *real* data and affects *real* inferences.

## 8. 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. But 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 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)), IO (Pakes, Ostrovsky, and Berry, (2005)), and macroeconomics (Lubik and Schorfheide, (2004)).

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

*and*

*Federal Reserve Bank of Atlanta, Research Department, 1000 Peachtree St NE, Atlanta, GA 30309-4470; Juan.Rubio@atl.frb.org; http://www.econ.umn.edu/~rubio/*

*and*

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

## 9. 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 $p(y_t|W_1^t, S_0, y^{t-1}; \gamma) \in C(W_1^t, S_0)$. ■

**Proof of Lemma 2.** The proof follows the same steps as the proof of the previous lemma. ■

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, page 31, DePree and Swartz, (1988)).*

**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, page 157, Dieudonné, (1960)).*

**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.[8] 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. ■

**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

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

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:

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

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

and

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

Let

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

Thus the likelihood functions are

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

and

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

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:

$$\prod_{t=1}^{T} \widetilde{p}_j\left(y_t|y^{t-1};\gamma\right) \to \prod_{t=1}^{T} p\left(y_t|y^{t-1};\gamma\right). \tag{11}$$

If we define

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

it follows that

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

Note that $W_{1,t}$ has bounded support and bounded density. Also, lemma 2 shows that $p_j\left(y_t|W_1^t,S_0,y^{t-1};\gamma\right)$ is continuous except in a finite number of points, with bounded support, and hence it is bounded. Consequently, $f_{j,T}\left(S_0;\gamma\right)$ is bounded. In addition, step 1 shows that $p_j\left(y_t|W_1^t,S_0,y^{t-1};\gamma\right)\to p\left(y_t|W_1^t,S_0,y^{t-1};\gamma\right)$, except in a finite number of points. Hence, $f_{j,T}\left(S_0;\gamma\right)\to f_T\left(S_0;\gamma\right)$, but in a finite number of points.

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

$$\left|\prod_{t=1}^{T}p_j\left(y_t|y^{t-1};\gamma\right)-\prod_{t=1}^{T}\widetilde{p}_j\left(y_t|y^{t-1};\gamma\right)\right|\leq\int\left|f_{j,T}\left(S_0;\gamma\right)-f_T\left(S_0;\gamma\right)\right|\mu_j^*\left(dS_0;\gamma\right).\tag{12}$$

Since $\mu_j^*\left(dS_0;\gamma\right)$ has a density with respect to the Lebesgue measure (assumption 6), the right-hand side of (12) is less than $\varepsilon$. We then conclude that:

$$\prod_{t=1}^{T}p_j\left(y_t|y^{t-1};\gamma\right)\to\prod_{t=1}^{T}\widetilde{p}_j\left(y_t|y^{t-1};\gamma\right).\tag{13}$$

To complete the proof, we put together the convergence results (11) and (13). ■

**Proof of Corollary 1.** Let $\gamma,\gamma'\in\Upsilon$. Proposition 1 shows that $L_j\left(y^T;\gamma\right)\to L\left(y^T;\gamma\right)$ and $L_j\left(y^T;\gamma'\right)\to L\left(y^T;\gamma'\right)$. Moreover, assumptions 4 and 8 imply that $L\left(y^T;\gamma\right)\geq\xi>0$ and $L_j\left(y^T;\gamma\right)\geq\xi>0$ for all $j$. Therefore

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

■

**Proof of Corollary 2.** Let $\gamma\in\Upsilon$. Since the approximated likelihoods $L_j\left(y^T;\gamma\right)$ and

$L\left(y^T;\gamma\right)$ are bounded and Riemann-integrable (because they are densities), we can apply Arzelà's theorem (see (Apostol), 1974, Theorem 9.12) and proposition 1 to get

$$\int_{\Upsilon} L_j\left(y^T;\gamma\right)\pi\left(\gamma\right)d\gamma \to \int_{\Upsilon} L\left(y^T;\gamma\right)\pi\left(\gamma\right)d\gamma.$$

■

**Proof of Corollary 3.** Let $\gamma \in \Upsilon$. Proposition 1 shows that $L_j\left(y^T;\gamma\right) \to L\left(y^T;\gamma\right)$. Then $L_j\left(y^T;\gamma\right)\pi\left(\gamma\right) \to L\left(y^T;\gamma\right)\pi\left(\gamma\right)$ and the result follows. ■

**Proof of Corollary 4.** Let $\gamma \in \Upsilon$. Proposition 1 shows that $L_j\left(y^T;\gamma\right) \to L\left(y^T;\gamma\right)$ and corollary 2 that $p_j\left(y^T;\gamma\right) \to p\left(y^T\right)$. The result follows from an application of Arzelà's theorem. ■

**Proof of Lemma 5.** Let $\gamma \in \Upsilon$. To prove that $p\left(y_t|W_1^t, S_0, y^{t-1};\gamma\right)$ is continuously differentiable with respect to $S_0$, we need to show that

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

exists and is continuous for all $i$.

Assumption 3 allows us to write

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

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

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

exists and it is bounded for all $t$ and all $i$. ■

**Proof of Lemma 6.** Let $\gamma \in \Upsilon$. Let $(v^t, s^t, w_2^t)$ be the unique solution to the following

system of equations:

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

and let $\left(v_j^t, s_j^t, w_{j,2}^t\right)$ be the unique solution to the approximated system of equations:[9]

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

By assumption 1, functions $\varphi$ and $g$ are differentiable. In addition, assumption 4 implies that $\left|dy\left(v_t, w_{2,t}; \gamma\right)\right| \neq 0$ for all $t$. Since $\left\|\varphi_j\left(\cdot, \cdot; \gamma\right) - \varphi\left(\cdot, \cdot; \gamma\right)\right\| \leq \delta$ and $\left\|g_j\left(\cdot, \cdot; \gamma\right) - g\left(\cdot, \cdot; \gamma\right)\right\| \leq \delta$, by the implicit function theorem of Schwartz (see theorem G.2.3, page 32, Mas-Colell, (1985)), there exists a $\lambda\left(S_0, W_1^t\right)$ such that:

$$\left\|\left(v_j^t, s_j^t, w_{j,2}^t\right) - \left(v^t, s^t, w_2^t\right)\right\| \leq \lambda\left(S_0, W_1^t\right)\delta. \tag{14}$$

Since the model is stationary, equation (14) holds for all $t$.

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

$$\left|p\left(v_{j,t}; \gamma\right) p\left(w_{j,2,t}; \gamma\right) - p\left(v_t; \gamma\right) p\left(w_{2,t}; \gamma\right)\right| \leq \varepsilon\delta, \tag{15}$$

for all $S_0$ and $W_1^t$. As before, note that equation (15) also holds for all $t$.

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

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:

$$||dy(v_{j,t}, w_{j,2,t};\gamma)| - |dy(v_t, w_{2,t};\gamma)|| \leq L_y \lambda \delta, \tag{16}$$

for all $S_0$ and $W_1^t$.

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

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

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 $\varphi_j$, $\varphi$, $g_j$, and $g$ are Lipschitz. Therefore, $\exists \Psi_2$ such that:

$$|\det(dy_j(v_{j,t}, w_{j,2,t};\gamma)) - \det(dy(v_{j,t}, w_{j,2,t};\gamma))| \leq n! n \Psi_2^{n-1} \Psi_1 \delta, \tag{17}$$

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

Using equations (16) and (17) we get:

$$|\det(dy_j(v_{j,t}, w_{j,2,t};\gamma)) - \det(dy(v_t, w_{2,t};\gamma))| \leq \left(n! n \Psi_2^{n-1} \Psi_1 + L_y \lambda\right) \delta, \tag{18}$$

for all $S_0$ and $W_1^t$.

Now, let $\Psi_3 = \left(n! n \Psi_2^{n-1} \Psi_1 + L_y \lambda\right)$. We can put together equations (15) and (18) to find:

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

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

$$|p(v_{j,t};\gamma)\,p(w_{j,2,t};\gamma)\,|dy_j(v_{j,t},w_{j,2,t};\gamma)|-p(v_t;\gamma)\,p(w_{2,t};\gamma)\,|dy(v,w_{2,t};\gamma)||\leq B\delta(\varepsilon+\Psi_3)$$

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.

■

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

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

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:

$$\left|\prod_{s=1}^{T}p\left(y_t|y^{t-1};\gamma\right)-\prod_{s=1}^{T}\widetilde{p}_j\left(y_t|y^{t-1};\gamma\right)\right|\leq\frac{L\delta}{1-\alpha}. \tag{19}$$

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

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

$$=\int\left(\int\prod_{t=1}^{T}p(W_{1,t};\gamma)\left(p_j\left(y_t|W_1^t,S_0,y^{t-1};\gamma\right)-p\left(y_t|W_1^t,S_0,y^{t-1};\gamma\right)\right)dW_1^t\right)\mu_j^*(dS_0;\gamma). \tag{20}$$

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

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

is an upper bound to (20).

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

$$\left|\prod_{s=1}^{T}p_j\left(y_t|y^{t-1};\gamma\right)-\prod_{s=1}^{T}\widetilde{p}_j\left(y_t|y^{t-1};\gamma\right)\right|\leq TB\chi\delta. \tag{21}$$

Putting together (19) and (21) delivers the result. ■

**Proof of Lemma 7.** The proof is a modification of the proof of lemma 6. The argument is the same except in the following points:

1. We use assumption 14 to state that $\left\|\left(v_j^t, s_j^t, w_{j,2}^t\right)-\left(v^t, s^t, w_2^t\right)\right\|\leq\lambda\delta$ for all $\gamma$, all $S_0$ and $W_1^t$.

2. We need assumptions 10, 12, and 13 to show that the densities of $V_t$ and $W_t$ are absolutely continuous.

3. We need assumptions 11 and 13 to show that the determinant of the Jacobian matrix of $y_t$ with respect to $V_t, W_{2,t}$, $\left|dy\left(\cdot,\cdot;\cdot\right)\right|$, is Lipschitz. Also, by assumption 13, the Lipschitz constant $L_y$ is independent of $\gamma$.

4. We need assumptions 1, 5, and 13 to show existence of a constant $\Psi_1$, independent of $\gamma$, such that:

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

   for all $r$ and $s$, and for all $\gamma$, all $S_0$ and $W_1^t$, except in a finite number of points.

5. We need assumptions 1, 5, and 13 to prove existence of a constant $\Psi_2$, independent of $\gamma$, such that:

$$||dy_j\left(v_{j,t}, w_{j,2,t}; \gamma\right)| - |dy\left(v_{j,t}, w_{j,2,t}; \gamma\right)|| \leq n! n \Psi_2^{n-1} \Psi_1 \delta, \tag{22}$$

for all $\gamma$, all $S_0$ and $W_1^t$, but in a finite number of points.

6. Since $V$, $W$, and $\gamma$ have compact support, assumption 12 is important to guarantee that $p\left(v;\gamma\right)$, $p\left(w_2;\gamma\right)$ are bounded functions of $\gamma$, all $S_0$ and $W_1^t$. Assumptions 1 and 13 imply that $|dy\left(v, w_2;\gamma\right)|$ is also a bounded function of $\gamma$, all $S_0$ and $W_1^t$.

■

**Proof of Proposition 3.** The proof is a modification of the proof of proposition 2. The argument is the same except:

1. We use assumptions 12 and 13 to make the bounds $\frac{\partial f_T(S_0;\gamma)}{\partial S_{0,i}}$ independent of $\gamma$. Then $f_T\left(S_0;\gamma\right)$ is Lipschitz for all $t$ with a Lipschitz constant $L$ independent of $\gamma$, and the difference

$$\left| \prod_{s=1}^{T} p\left(y_t | y^{t-1};\gamma\right) - \prod_{s=1}^{T} \widetilde{p}_j\left(y_t | y^{t-1};\gamma\right) \right| \leq \frac{L\delta}{1-\alpha}, \tag{23}$$

holds for all $\gamma$.

2. Assumption 13 makes the bounds on $p\left(y_t | W_1^t, S_0, y^{t-1};\gamma\right)$ and $p_j\left(y_t | W_1^t, S_0, y^{t-1};\gamma\right)$ independent of $\gamma$. Therefore, the bound $B$ and all the expressions where it appears are independent of $\gamma$.

■

## References


[1] Apostol, T.M. (1974). *Mathematical Analysis*, $2^{nd}$ edition. Addison Wesley.

[2] Bajari, P., H. Hong and S. Ryan (2004). "Identification and Estimation of Discrete Games of Complete Information". *Mimeo*, Duke University.

[3] Benhabib, J. and R.E.A. Farmer (1999). "Indeterminacy and Sunspots in Macroeconomics", in The Handbook of Macroeconomics, John Taylor and Michael Woodford, eds., North Holland.

[4] Boldrin, M, and M. Woodford (1990). "Equilibrium Models Displaying Endogenous Fluctuations and Chaos: A Survey". *Journal of Monetary Economics* 25, 189-222.

[5] Brin, M and Y. Kifer (1987). "Dynamics of Markov Chains and Stable Manifolds for Random Diffeomorphisms". *Ergodic Theory and Dynamical Systems* 7, 351-374.

[6] Crawford, G.S. and M. Shum (2005). "Uncertainty and Learning in Pharmaceutical Demand". *Econometrica* 73, 1137-1173.

[7] Daula, T. and R. Moffitt (1995). "Estimating Dynamic Models of Quit Behavior: The Case of Military Reenlistment". *Journal of Labor Economics* 13, 499-523.

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

[9] DePree, J.D. and C.W. Swartz (1988). *Introduction to Real Analysis*. Wiley.

[10] Dieudonné, J. (1960). *Treatise on Analysis: Volume 1*. Academic Press.

[11] Duffie, D. and K.J. Singleton (1993). "Simulated Moments Estimation of Markov Models of Asset Prices". *Econometrica* 61, 929-952.

[12] Ellison, G. and D. Fudenberg (1993). "Rules of Thumb for Social Learning". *Journal of Political Economy* 101, 612-643.

[13] Fernández-Villaverde, J. and J. Rubio-Ramírez (2004b). "Estimating Nonlinear Dynamic Equilibrium Economies: a Likelihood Approach". *Penn Institute for Economic Research Working Paper 04-001.* University of Pennsylvania.

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

[15] Flinn, C.C. and J. Heckman (1982). "New Methods for Analyzing Structural Models of Labor Force Dynamics". *Journal of Econometrics* 18, 115-168.

[16] Foley, D. and M. Hellwig (1975). "Asset Management with Trading Uncertainty". *Review of Economic Studies* 42, 327-346.

[17] Futia, C. (1982). "Invariant Distributions and the limiting Behavior of Markovian Economic Models". *Econometrica* 50, 377-408.

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

[19] Gilleskie, D.B. (1998). "A Dynamic Stochastic Model of Medical Care Use and Work Absence". *Econometrica* 66, 1-45.

[20] Gouriéroux, C. and A. Monfort (1996). *Simulation-Based Econometric Methods.* Oxford University Press.

[21] Jofre-Bonet, M. and M. Pesendorfer (2003). "Estimation of a Dynamic Auction Game". *Econometrica* 71, 1443-1489.

[22] 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.

[23] Keane, M.P. and K.I. Wolpin (1997). "The Career Decisions of Young Men". *Journal of Political Economy* 105, 473-522.

[24] Keane, M.P. and R. Moffitt (1998). "A Structural Model of Multiple Welfare Program Participation and Labor Supply". *International Economic Review* 39, 553-589.

[25] Kehoe, T. and D.K. Levine (1985). "Comparative Statics and Perfect Foresight in Infinite Horizon Economies," *Econometrica* 53, 433-453.

[26] Lubik, T. and F. Schorfheide (2004). "Testing for Indeterminacy: An Application to U.S. Monetary Policy". *American Economic Review* 94, 190-217.

[27] Mas-Colell, A. (1985). *The Theory of General Economic Equilibrium: A Differentiable Approach.* Econometric Society Monographs, Cambridge University Press.

[28] Pakes, A. (1986). "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks". *Econometrica* 54, 755-785.

[29] Pakes, A., M. Ostrovsky, and S. Berry (2004). "Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Examples)". *Mimeo*, Harvard University.

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

[31] Peres, Y. and B. Solomyak (1996). "Absolute Continuity of Bernoulli Convolutions: A Simple Proof". *Mathematical Research Letters* 3, 231-239.

[32] Rabanal, P. and J. Rubio-Ramírez (2005). "Comparing New Keynesian Models of the Business Cycle: a Bayesian Approach". *Journal of Monetary Economics,* forthcoming.

[33] 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.

[34] Rust, J. (1987). "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher". *Econometrica* 55, 999-1033.

[35] Rust, J. and C. Phelan (1997). "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets". *Econometrica* 65, 781-831.

[36] Sanghvi, A.P. and M.J. Sobel (1976). "Bayesian Games as Stochastic Processes". *International Journal of Game Theory* 5, 1-22.

[37] Santos, M.S. (2000). "Accuracy of Numerical Solutions Using the Euler Equation Residuals". *Econometrica* 68, 1377-1402.

[38] Santos, M.S. (2004). "Simulation-Based Estimation of Dynamic Models with Continuous Equilibrium Solutions". *Journal of Mathematical Economics*, 40, 465-491.

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

[40] Santos, M.S. and J. Vigo (1998). "Analysis of Error for a Dynamic Programming Algorithm". *Econometrica* 66, 409-426.

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

[42] Schenk-Hoppé, K.R. and B. Schmalfuss (2001). "Random Fixed-Points in a Stochastic Solow Growth Model". *Journal of Mathematical Economics* 36, 19-30.

[43] Schmalensee, R. (1975). "Alternative Models of Bandit Selection". *Journal of Economic Theory* 10, 333-342.

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

[45] 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.

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

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

[48] Stachurski, J. (2002). "Stochastic Optimal Growth with Unbounded Shock". *Journal of Economic Theory* 106, 45–60.

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

[50] Wolpin, K.I. (1984). "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality". *Journal of Political Economy* 92, 852-874.

Figure 1: Absolute Value Difference between the Likelihoods as a Function of the Sample Size