FEDERAL RESERVE BANK of ATLANTA

Convergence Properties of the Likelihood of Computed Dynamic Models

Jesús Fernández-Villaverde, Juan Francisco Rubio-Ramírez, and Manuel Santos

Working Paper 2004-27
November 2004

WORKING PAPER SERIES

FEDERAL RESERVE BANK *of* ATLANTA WORKING PAPER SERIES

Convergence Properties of the Likelihood of Computed Dynamic Models

Jesús Fernández-Villaverde, Juan Francisco Rubio-Ramírez, and Manuel Santos

Working Paper 2004-27
November 2004

Abstract: This paper studies the econometrics of computed dynamic models. Since these models generally lack a closed-form solution, economists approximate the policy functions of the agents in the model with numerical methods. But this implies that, instead of the exact likelihood function, the researcher can evaluate only an approximated likelihood associated with the approximated policy function. What are the consequences for inference of the use of approximated likelihoods? First, we show that as the approximated policy function converges to the exact policy, the approximated likelihood also converges to the exact likelihood. Second, we prove that the approximated likelihood converges at the same rate as the approximated policy function. Third, we find that the error in the approximated likelihood gets compounded with the size of the sample. Fourth, we discuss convergence of Bayesian and classical estimates. We complete the paper with three applications to document the quantitative importance of our results.

JEL classification: C63, C68, E37

Key words: likelihood function, dynamic models, numerical approximation

---

The authors thank Jim Nason, Tom Sargent, Frank Schorfheide, Tao Zha, and participants at several seminars for useful comments. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the authors' responsibility.

Please address questions regarding content to Jesús Fernández-Villaverde, Department of Economics, University of Pennsylvania, 160 McNeil Building, 3718 Locust Walk, Philadelphia, Pennsylvania 19104, 215-898-1504, jesusfv@upenn.edu; Juan Francisco Rubio-Ramirez, Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street, NE, Atlanta, Georgia 30309-4470, 404-498-8057; juan.rubio@atl.frb.org; or Manuel Santos, BAC 551, Department of Economics, Arizona State University, Tempe, Arizona 85287-3806, 480-965-6335.

Convergence Properties of the Likelihood of Computed Dynamic Models

1. Introduction

This paper studies the following problem. Most dynamic models do not have a closed-form solution. Instead, the economist needs to approximate the solution using a numerical method. This approximation implies that when the 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 numerically approximated solution to the model. What are the effects on statistical inference of using an approximated likelihood instead of the exact likelihood function?

Over the last 20 years, there has been considerable progress in the field of dynamic models in economics, both at the micro and at the macro level. The popularity of this class of models has raised an interest in their estimation using a likelihood-based approach. There are several reasons for that interest. First, likelihood inference offers the flexibility to handle a large class of assumptions regarding preferences, technology, and information sets. Second, likelihood inference allows for the estimation of the whole range of parameters required to perform policy experiments. Third, the likelihood delivers good efficiency properties and small sample behavior even under potential model misspecifications.

Without being exhaustive, we enumerate a few examples of the successful estimation of dynamic models with a likelihood approach. In the area of discrete choice dynamic programming models, likelihood inference has been applied to a wide range of questions in industrial organization, labor economics, development, health economics, demography, and public finance. Among many others, we can cite Flinn and Heckman (1982), Miller (1984), Wolpin (1984), Pakes (1986), Rust (1987), Rosenzweig and Wolpin (1993), Daula and Moffitt (1995), Ferrall (1997), Keane and Wolpin (1997), Rust and Phelan (1997), Gilleskie (1998), and Keane and Moffitt (1998). In macroeconomics, examples of how to estimate dynamic general equilibrium economies using the likelihood function include Sargent (1989), McGrattan, Rogerson, and

Wright (1997), Landon-Lane (1999), DeJong, Ingram, and Whiteman (2000), Schorfheide (2000), Dib (2001), Otrok (2001), Ireland (2002), Fernández-Villaverde and Rubio-Ramírez (2003), Lubik and Schorfheide (2003), Rabanal and Rubio-Ramírez (2003), and Smets and Wouters (2003).

All these applications face a similar problem: how to evaluate the likelihood function of the model. A key difficulty in that evaluation is that dynamic models imply policy rules for the agents for which we do not have closed-form solutions except in a few cases. In practice, researchers circumvent that problem by approximating the policy rules using numerical methods and building the likelihood associated with those approximated policy rules.

But this approach implies that when we perform inference, the economist does not use the exact likelihood of the model under consideration but an approximated likelihood. Consequently, it is important to assess how the likelihood generated by numerically approximated policy functions relates to the exact likelihood. We need to ask ourselves questions such as: 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? What are the effect of the approximation on the parameter estimates? And on hypothesis testing?

These questions are important not only theoretically but also from an applied perspective. Numerical methods allow the user to control the error in the approximation. For example, we can add more points to the grid in the dynamic programming algorithm. However, the reduction in the error that we get with the additional points is achieved at the cost of speed. Given this trade-off between speed and accuracy, how many points are enough? Can we relate the error in the policy function created by the use of a grid to the error in the approximated likelihood? Do we need to make our choice of grid dependent on the size of the sample?

Unfortunately, not much is known about the convergence properties of the likelihood of computed dynamic models. To fill this gap, we build on the recent work by Santos and Peralta-Alva (2003) and Santos (2003), 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 using 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 the approximated likelihood functions. This extension raises a whole new range of issues not previously explored, as far as we know, either in economics or statistics.

First, we present an example where the sequence of approximated likelihoods does not converge to the exact likelihood even if the sequence of approximated policy functions converges to the exact policy function. This example motivates why, in general, we cannot assume the convergence of the approximated likelihood and why we need to find conditions under which this convergence is guaranteed.

Then, we develop our theoretical setting and derive our findings. Our most important result is that for given parameter values, as the approximated policy function converges to the exact policy function in the sup norm, the approximated likelihood function also converges to the exact likelihood if certain conditions are satisfied. This is a basic consistency result because it ensures convergence of likelihood ratios and of the marginal likelihoods.

We also 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. The intuition is as follows. Period by period, small errors in the policy function accumulate at the same rate at which the sample size grows. This means that as the sample size goes to infinity, a linear approximation will deliver an approximation of the likelihood that will fail to converge. This finding suggests that solution methods where reduction of the error is not possible, like linear approximations, may face difficulties with large samples.

Our third result regards 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. However, we can impose mildly 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.

We complete the paper with three economic applications, where we progressively docu-

ment how our results work in action. The applications illustrate how our findings are useful for practitioners and how the issues created by the use of approximated likelihood functions are quantitatively important.

An issue that is related to, but different from, the focus of this paper is how to evaluate the likelihood when that 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 wanted 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 sets up an environment to discuss the convergence of the likelihood. Section 4 shows our main result concerning convergence. Section 5 discusses the speed of convergence and its relation to the sample size. Section 6 presents our findings regarding the convergence of maximum likelihood point estimates. Section 7 studies three examples to see how the results of the paper hold in practice. Section 8 concludes. An appendix includes all the proofs of the results in the paper.

2. An Example of Nonconvergence

We now present an example to illustrate why, 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 likelihoods associated with it does not converge to the exact one even if the sequence of approximated policy functions converges to the exact policy function.

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 σ_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 this last argument links the current choice with future payoffs, which are discounted at rate β. 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:

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

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 1 with probability 1/2 and state 2 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]

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 this two nonoverlapping ergodic distributions implies that, in order to write the likelihood function, we need to specify where does the initial state of the economy S_0 come from. Following Lubik and Schorfheide (2004) we assume that there is a sunspot that picks one of the two distributions. The sunspot has probability π_A to signal the first ergodic distribution and π_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

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

first state S_0 of those data is:

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

where γ is a vector of structural parameters of the model and $\phi(\cdot)$ is the standardized normal density.

Equation (2) and the sunspot distribution imply that the unconditional likelihood of y^T is given by:

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

Now let us assume that, because of the use of a numerical method to solve the dynamic discrete choice problem, the economist cannot compute the exact policy function φ, but only an approximated policy function φ_j of the form:

$$\varphi_j=\begin{bmatrix}1-\delta_j & \frac{\delta_j}{2} & \frac{\delta_j}{2}\\ 0 & \frac{1}{2} & \frac{1}{2}\\ 0 & \frac{1}{2} & \frac{1}{2}\end{bmatrix} \tag{4}$$

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 (for example, in value function iteration, an index of the number of grid points). The solution method is flexible enough such that the economist can refine the approximation as much as she wants to guarantee that $\delta_j\to 0$ as $j\to\infty$. We also let the economist to be able to compute exactly $y_t=\varepsilon_{i,t}$ if $S_t=i$.

Given this problem, no matter how good our approximated policy function is (i.e., not matter how small δ_j is), the ergodic distribution for the states of the economy is $\left(0,\frac{1}{2},\frac{1}{2}\right)$ for

all j.

As a consequence, the unconditional approximated likelihood of y^T than the economist can evaluate in practice is:

$$L_j\left(y^T;\gamma\right) = \frac{L\left(y^T;\gamma,S_0=2\right)}{2} + \frac{L\left(y^T;\gamma,S_0=3\right)}{2} \tag{5}$$

for all j.

Comparing the two likelihoods (3) and (5), we can see that $\varphi_j \to \varphi$, but $L_j\left(y^T;\gamma\right) \nrightarrow L\left(y^T;\gamma\right)$.

This example has shown how the sequence of approximated likelihoods may fail to converge. It motivates why it is important to find conditions that ensure the convergence of the approximated likelihoods and to study the rate of convergence. Also it presents several elements that will be important in our results: the continuity (or discontinuity) of the exact policy function, the convergence of the sequence of approximated policy functions, the maximum error of the approximated policy function, and the role of an stationary distribution of states of the economy. We will discuss each of these elements below.

3. The Setting

In this section we present the environment in which we will work to investigate the convergence properties of the likelihood of computed dynamic models.

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

$$S_t = \varphi\left(S_{t-1},W_t;\gamma\right) \tag{6}$$

$$Y_t = g\left(S_t,V_t;\gamma\right). \tag{7}$$

Here S_t is a vector of state variables that characterize the evolution of the system. The state variables can be partitioned between a vector of endogenous state variables, K_t, and a vector of exogenous state variables Z_t. The vector of state variables $S_t = (K_t, Z_t)$ belongs to the

compact set $S = K \times Z \subset R^l \times R^m$. 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 the Euclidean space, with bounded and continuous densities. W_t and V_t are independent of each other. More involved stochastic structures can be accommodated by appropriately increasing the dimensionality of the state space. The observables in each period are stacked in a vector Y_t. If we have T periods of observations, we define $Y^T \equiv (Y_1', ..., Y_T')'$ with $Y^0 = \{\emptyset\}$. We assume that Y^T is distributed according to the probability density function $p_0^T(\cdot)$. Finally, γ, which belongs to the compact set $\Upsilon \subset R^s$, is the vector of structural parameters, i.e., those describing the preferences, technology, and information sets of the economy.

It is also the case that $\dim(W_t) + \dim(V_t) \geq \dim(Y_t)$. This assumption ensures that the model is not stochastically singular. We do not impose any restrictions on how those degrees of stochasticity are achieved. Fernández-Villaverde and Rubio-Ramírez (2003) discuss the issue in detail. Finally, let us partition $\{W_t\}$ into two sequences $\{W_{1,t}\}$ and $\{W_{2,t}\}$, such that $W_t = (W_{1,t}, W_{2,t})$ and $\dim(W_{2,t}) + \dim(V_t) = \dim(Y_t)$.

This partition is not restrictive. It accommodates the case where $\{W_{2,t}\}$ is a zero-dimensional sequence immediately. We could also allow $\{W_{1,t}\}$ to be a zero-dimensional sequence at the cost of heavier notation throughout the paper. At the same time, the partition is useful for increasing the class of models that can be studied, since it lets us deal with cases where $\dim(V_t) < \dim(Y_t)$ but $\dim(W_{2,t}) > 0$.

Equation (6) is known as the transition equation, since it governs the evolution of states over time. Equation (7) is called the measurement equation because it relates states and observables. Note that, abusing notation, we allow the possibility that the dimensionality of the shocks could be zero and that the states might be part of the observables without noise (for example, if g is the identity function along some dimension).

Before continuing with our analysis, we make the following assumptions:

Assumption 1. *$\varphi(\cdot, \cdot; \gamma)$ and $g(\cdot, \cdot; \gamma)$ are continuously differentiable, with bounded partial derivatives, for all γ.*

Assumption 1 arises naturally in a number of economic models. For example, the continuity of $\varphi(\cdot, \cdot; \gamma)$ often follows from results like Theorem 4.8 in Stokey, Lucas, and Prescott

(1989) that ensure the continuity and single-valuedness of the policy functions of agents.

Now we set out some definitions that will be useful in the rest of the paper. First, let $W_i^t = \{W_{i,m}\}_{m=1}^t$ and let w_i^t be a realization of the random variable W_i^t for $i = 1, 2$ and $\forall t$. Let $V^t = \{V_m\}_{m=1}^t$ and let v^t be a realization of the random variable V^t for $\forall t$. Let $S^t = \{S_m\}_{m=0}^t$ and s^t be a realization of the random variable S^t for $\forall t$. Let $Y^t = \{Y_m\}_{m=1}^t$ and y^t be a realization of the random variable Y^t for $\forall t$. We also define $W_i^0 = \{\emptyset\}$ and $y^0 = \{\emptyset\}$.

We introduce some additional constructs. Let $C(S)$ be the space of all continuous, $\mathcal{S}-$measurable, real-valued functions on S. Similarly, let $V(S)$ be the space of all bounded, $\mathcal{S}-$measurable, real-valued functions on S. We endow $C(S)$ with the norm $\|f\| = \sup_{s \in S} |f(s)|$ and induce a Banach space. For a vector-valued function $f = (\ldots, f^i, \ldots)$ we define $\|f\| = \max_{1 \leq i \leq l+m} \|f^i\|$. Convergence of a sequence of functions $\{f_j\}$ should be understood in the metric induced by this norm.

We define the operator Ψ from the space of probability measures on $\mathcal{S}$ into itself as:

$$(\Psi\mu)(A;\gamma) = \int P(s, A;\gamma)\, \mu(ds;\gamma) \text{ for all } A \in \mathcal{S} \tag{8}$$

where $P(\cdot, \cdot;\gamma)$ is a transition kernel on $(S, \mathcal{S})$ generated by the transition equation (6) evaluated at parameter values γ. Standard arguments show that there exists a fixed point of the operator Ψ for all γ. We will call this fixed point $\mu^*(S;\gamma)$, the invariant distribution for S on $\mathcal{S}$ of the dynamic model. Note that an invariant distribution for S also implies an invariant distribution for y_t through the measurement equation (7).

Now we assume that this invariant distribution has a density that we can use in our future derivations. With some extra work, this assumption can also be written directly in terms of the policy functions $\varphi(\cdot, \cdot;\gamma)$, and $g(\cdot, \cdot;\gamma)$.

Assumption 2. *The invariant distribution for S, $\mu^*(S;\gamma)$, has a Radom-Nykodim derivative with respect to the Lebesgue measure for all γ.*

With the invariant measure, we can define the likelihood of the data as follows. If y^T is a realization of the random variable $Y^T = \{Y_t\}_{t=1}^T$, its likelihood conditional on parameter

values γ is given by:

$$L\left(y^T;\gamma\right)=\prod_{t=1}^{T}p\left(y_t|y^{t-1};\gamma\right)=\prod_{t=1}^{T}\int p\left(y_t|y^{t-1},W_1^t,S_0;\gamma\right)p\left(W_1^t,S_0|y^{t-1};\gamma\right)dW_1^tdS_0, \quad (9)$$

where

$$p\left(W_1^t,S_0|y^{t-1};\gamma\right)=p\left(W_{1,t};\gamma\right)p\left(W_1^{t-1},S_0|y^{t-1};\gamma\right), \quad (10)$$

and

$$p\left(W_1^{t-1},S_0|y^{t-1};\gamma\right)=\frac{p\left(y_{t-1}|y^{t-2},W_1^{t-1},S_0;\gamma\right)p\left(W_1^{t-1},S_0|y^{t-2};\gamma\right)}{p\left(y_{t-1}|y^{t-2};\gamma\right)}, \quad (11)$$

for all t and γ. Finally, we set $p\left(W_1^0,S_0|y^0;\gamma\right)dS_0=\mu^*\left(dS_0;\gamma\right)$.

Pasting together (10) and (11), we can see that $p\left(W_1^t,S_0|y^{t-1};\gamma\right)$ has the following recursive structure:

$$p\left(W_1^t,S_0|y^{t-1};\gamma\right)=p\left(W_{1,t};\gamma\right)\frac{p\left(y_{t-1}|y^{t-2},W_1^{t-1},S_0;\gamma\right)p\left(W_1^{t-1},S_0|y^{t-2};\gamma\right)}{p\left(y_{t-1}|y^{t-2};\gamma\right)},$$

for all t, S_0 and γ.

Define the pseudo-maximum likelihood point estimate (PMLE) as

$$\widehat{\gamma}\left(y^T\right)\equiv\arg\max_{\gamma\in\Upsilon}p\left(y^T;\gamma\right).$$

Note that we do not assume that there exists a value γ^* such that $p\left(y^T;\gamma^*\right)=p_0^T\left(y^T\right)$ (hence, the term pseudo). Statistically, this means that the model may be misspecified. Far more important, from an economic perspective, this is a direct consequence of the fact that dynamic models are false by construction.

Now we make a basic and rather weak assumption about our ability to use the model to think about the data.

Assumption 3. *We can evaluate the conditional densities $p\left(y_t|W_1^t,y^{t-1},S_0;\gamma\right)$ for all t, S_0, W_1^t, and γ.*

Assumption 3 implies that for any realizations s_0, w_1^t, and y^t of the random variables S_0, W_1^t and Y^t, we can evaluate the probability of the model described by (6) and (7) of generating the observables. In other words, assumption 3 implies that for any s_0, w_1^t, and y^t, the following system of equations

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

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

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

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

To simplify the notation, we are going to write $\left(v^t, s^t, w_2^t\right)$, instead of the more cumbersome $\left(v^t\left(s_0, w_1^t, y^t; \gamma\right), s^t\left(s_0, w_1^t, y^t; \gamma\right), w_2^t\left(s_0, w_1^t, y^t; \gamma\right)\right)$. Then, we have

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

for all W_1^t, S_0, and γ, and all t, where $\left|dy\left(v_t, w_{2,t}; \gamma\right)\right|$ stands for the determinant of the Jacobian of y_t with respect to V_t and $W_{2,t}$ evaluated at v_t and $w_{2,t}$. Note that assumption 3 requires only the ability to evaluate the density; it does not require having a closed form for it. As a consequence, we allow numerical or simulation methods for this evaluation.

To avoid trivial problems, we assume that the model assigns positive probability to the data, y^T, for any initial S_0. This is formally reflected in the following assumption:

Assumption 4. *For all S_0, W_1^t, and γ the model gives some positive probability to the data y^T, i.e.*

$$p\left(y_t | y^{t-1}, W_1^t, S_0; \gamma\right) > \xi \geq 0,$$

for all t.

Assumption 4 and repeated applications of equation (11) lead us to write the likelihood

function, (9), in the following recursive way:

$$L\left(y^T;\gamma\right)=\prod_{t=1}^{T}p\left(y_t|y^{t-1};\gamma\right)=\int\left(\int\prod_{t=1}^{T}p\left(W_{1,t};\gamma\right)p\left(y_t|y^{t-1},W_1^t,S_0;\gamma\right)dW_1^t\right)\mu^*\left(dS_0;\gamma\right). \tag{12}$$

This structure will be useful proving the theorems in the next section.

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 (12) is conceptually a simple task, although potentially cumbersome to implement. 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 solution method we use to approximate 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 an increase in the number of points on the grid, 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 last two functions. The researcher can evaluate only 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 $L\left(y^T;\gamma\right)$? If so, at what speed? And what about the point estimates?

The objective of this section is to show that for any given value of the structural parameters, γ, 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 functions. Formally, we prove that for any given γ, the

following limit holds:

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

as $\varphi_j(\cdot,\cdot;\gamma)\rightarrow\varphi(\cdot,\cdot;\gamma)$ and $g_j(\cdot,\cdot;\gamma)\rightarrow g(\cdot,\cdot;\gamma)$. We argue that this convergence has several important consequences for estimation.

We organize the section as follows. First, lemma 1 replicates Theorem 3.2 in Santos and Peralta-Alva (2003). This theorem asserts the bilinear convergence of $\Psi_j\mu_j^*$ to $\Psi\mu^*$ for any given γ, where Ψ_j is the equivalent operator to (8) for the approximated transition equation, and μ_j^* is the fixed point of Ψ_j. Then, lemmas 2 and 3 show that for any given γ, $p\left(y_t|y^{t-1},W_1^t,S_0;\gamma\right)$ and $p\left(y_t|y^{t-1},W_1^t,S_0;\gamma\right)$ are continuous as a function of S_0. These lemmas are then used to show the main result of the section, the convergence of the likelihood, in proposition 5. Finally, we discuss the effects of this result on several aspects of inference.

4.1. Convergence of the Invariant Distribution

Let $\mu_j^*(S;\gamma)$ be the invariant distribution of S on $\mathcal{S}$ associated with the approximated functions $\varphi_j(\cdot,\cdot;\gamma)$ and $g_j(\cdot,\cdot;\gamma)$:

$$\mu_j^*(S;\gamma)=\left(\Psi_j\mu_j^*\right)(A;\gamma)=\int P_j(s,A;\gamma)\,\mu_j^*(ds;\gamma)\ \text{ for all } A\in\mathcal{S}$$

where $P_j(\cdot,\cdot;\gamma)$ is a transition kernel on $(S,\mathcal{S})$ induced by the approximated transition equation.

Assumption 5. *The invariant distribution for S, $\mu_j^*(S;\gamma)$, has a Radom-Nykodim derivative with respect to the Lebesgue measure for all γ.*

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

Under assumption 1, as the approximated functions converge to the exact functions, the invariant distributions generated by the approximations to the measurement and transition

equations will converge to the invariant distributions created by exact measurement and transition functions. This result is formally stated in Theorem 3.2 in Santos and Peralta-Alva (2003), that we reproduce here.

Lemma 1. *Let* $\gamma \in \Upsilon$, $\varphi_j(\cdot,\cdot;\gamma) \to \varphi(\cdot,\cdot;\gamma)$, *and* $g_j(\cdot,\cdot;\gamma) \to g(\cdot,\cdot;\gamma)$. *Then, under assumption 1, every weak limit point* $\mu^*(S;\gamma)$ *of* $\left\{\mu_j^*(S;\gamma)\right\}$ *is an invariant distribution associated with* $\varphi(\cdot,\cdot;\gamma)$ *and* $g(\cdot,\cdot;\gamma)$.

In other words, this lemma tells us that the invariant distribution correspondence is upper semicontinuous. As discussed by Santos and Peralta-Alva (2003), the theorem asserts the bilinear convergence of $\Psi_j\mu_j^*$ to $\Psi\mu^*$.

4.2. Continuity of Conditional Probabilities

We now proceed to show how the conditional probability $p(y_t|y^{t-1}, W_1^t, S_0;\gamma)$ is a continuous, real-valued function of S_0.

Lemma 2. *Let* $\gamma \in \Upsilon$. *Under assumptions 1 and 3,* $p(y_t|y^{t-1}, W_1^t, S_0;\gamma) \in C(S_0)$ *for all* t.

The proof of the lemma, as the proof of the other results in the paper, is technical, and it can be found in the appendix. From this lemma we can also derive that $L\left(y^T;\gamma\right)$ is bounded, since $p(y_t|y^{t-1}, W_1^t, S_0;\gamma)$ is continuous with bounded support. This point will be useful below.

We now need to prove that the conditional probability $p_j(y_t|y^{t-1}, W_1^t, S_0;\gamma)$ associated with the approximated transition and measurement equations is also a continuous, real-valued function of S_0. To do so, we assume that:

Assumption 6. $\varphi_j(\cdot,\cdot;\gamma)$ *and* $g_j(\cdot,\cdot;\gamma)$ *are continuous for all* γ, *and all* j. $\varphi_j(\cdot,\cdot;\gamma)$ *and* $g_j(\cdot,\cdot;\gamma)$ *are continuously differentiable at all points except in a finite number of points for all* γ *and all* j. *If they exist, partial derivatives are bounded, and the bounds are independent of* j.

Assumption 6 ensures continuity of $\varphi_j(\cdot,\cdot;\gamma)$ and $g_j(\cdot,\cdot;\gamma)$ at all points, while both functions could be no 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, the commonly used value function iteration with linear interpolation or the finite elements method as described in McGrattan (1999).

We also have the equivalent of assumption 3 for approximated functions:

Assumption 7. *We can evaluate the conditional densities $p_j(y_t|y^{t-1},\cdot,\cdot;\gamma)$ at all points except in a finite number of points for all t, all γ, and all j.*

As in the previous section, assumption 7 implies that for any s_0, w_1^t, and y^t, the following system of equations

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

$$y_m = g_j(S_m,V_m;\gamma) \qquad \text{for } m=1,2,...t$$

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

has a unique solution $\left(v_j^t(s_0,w_1^t,y^t;\gamma), s_j^t(s_0,w_1^t,y^t;\gamma), w_{j,2}^t(s_0,w_1^t,y^t;\gamma)\right)$, and that we can evaluate the probabilities $p\left(v_j^t(s_0,w_1^t,y^t;\gamma);\gamma\right)$ and $p\left(w_{j,2}^t(s_0,w_1^t,y^t;\gamma);\gamma\right)$.

As we did before and to simplify the notation, we write $\left(v_j^t, s_j^t, w_{j,2}^t\right)$, instead of the much more complicated $\left(v_j^t(s_0,w_1^t,y^t;\gamma), s_j^t(s_0,w_1^t,y^t;\gamma), w_{j,2}^t(s_0,w_1^t,y^t;\gamma)\right)$. Since assumption 6 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:

$$p_j\left(y_t|y^{t-1},\cdot,\cdot;\gamma\right) = p(v_{j,t};\gamma)\,p(w_{j,2,t};\gamma)\,|dy_j(v_{j,t},w_{j,2,t};\gamma)|$$

for all but a finite number of points for all t, all γ, and all j. Notice that the Jacobian 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 can explain the data even if it does so with arbitrarily low probability:

Assumption 8. *The model gives some positive probability to the data y^T, i.e.,*

$$p_j\left(y_t|y^{t-1},\cdot,\cdot;\gamma\right)\geq\xi>0.$$

at all points except in a finite number of points for all t, all γ, and all j.

Now we can prove the equivalent to lemma 2 for the approximated functions.

Lemma 3. *Let $\gamma\in\Upsilon$. Under assumptions 6 and 7, then $p_j\left(y_t|y^{t-1},W_1^t,S_0;\gamma\right)\in C\left(S_0\right)$ at all except in a finite number of points for all t and all j.*

As before in the case of the exact probability, this lemma ensures that $L_j\left(y^T;\gamma\right)$ is bounded.

4.3. Main Result: Convergence of the Likelihood Function

To prove convergence of the likelihood function, and since 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 $g\left(\cdot,\cdot;\gamma\right)$, we need to consider the convergence of such Jacobians as an intermediate step. To show that $d\varphi_j\left(\cdot,\cdot;\gamma\right)\rightarrow d\varphi\left(\cdot,\cdot;\gamma\right)$ and $dg_j\left(\cdot,\cdot;\gamma\right)\rightarrow dg\left(\cdot,\cdot;\gamma\right)$, as $\varphi_j\left(\cdot,\cdot;\gamma\right)\rightarrow\varphi\left(\cdot,\cdot;\gamma\right)$ and $g_j\left(\cdot,\cdot;\gamma\right)\rightarrow g\left(\cdot,\cdot;\gamma\right)$, we first need to assume that:

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

This assumption is satisfied naturally by most solution methods for dynamic models, since a common strategy is to find an approximation to the unknown functions using some well behaved basis, like polynomials. Our previous examples of the value function iteration and the finite elements method fit into this category. Other popular procedures such as linearization and perturbation methods, do as well (see Judd, 1998).

Our next lemma shows how assumption 9 ensures that wherever the transition and measurement equations are differentiable, $d\varphi_j\left(\cdot,\cdot;\gamma\right)\rightarrow d\varphi\left(\cdot,\cdot;\gamma\right)$ and $dg_j\left(\cdot,\cdot;\gamma\right)\rightarrow dg\left(\cdot,\cdot;\gamma\right)$, as $\varphi_j\left(\cdot,\cdot;\gamma\right)\rightarrow\varphi\left(\cdot,\cdot;\gamma\right)$ and $g_j\left(\cdot,\cdot;\gamma\right)\rightarrow g\left(\cdot,\cdot;\gamma\right)$.

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, 2, 3, and 4 to prove the main result of this section, the convergence of the likelihood function. Formally:

Proposition 5. *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:*

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

for all t.

The result is key for applied work: It states that for any given γ, as we get better and better approximations of the policy function in our dynamic model, the likelihood computed also converges to the exact likelihood. This finding provides a foundation for the 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 of the model. It is important to notice that proposition 5 shows only pointwise convergence of the likelihood function. Section 6 analyzes the additional assumptions needed to prove uniform convergence.

4.4. Applications of the Result

The result in proposition 5 has a number of implications. Here, we will highlight two of them. First, pointwise convergence implies that for any given γ and γ', the ratio of likelihood functions converges.

Corollary 6. *Let $\gamma, \gamma' \in \Upsilon$. Under assumptions 1 to 9, if $\varphi_j(\cdot,\cdot;\gamma) \to \varphi(\cdot,\cdot;\gamma)$, $g_j(\cdot,\cdot;\gamma) \to g(\cdot,\cdot;\gamma)$, $\varphi_j(\cdot,\cdot;\gamma') \to \varphi(\cdot,\cdot;\gamma')$, and $g_j(\cdot,\cdot;\gamma') \to g(\cdot,\cdot;\gamma')$, then:*

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

This result is useful in all contexts in which likelihood ratios are built, such as in classical hypothesis testing or comparison of models (Vuong, 1989), or when implementing the Metropolis-Hastings algorithm.

The second implication of the result comes from its direct effects for Bayesian inference. There are two main objects of interest in the Bayesian paradigm: the marginal likelihood of the model, $p\left(y^t\right)$, and the posterior distribution of the parameters, $p\left(\gamma|y^t\right)$.

The marginal likelihood of the exact 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 (see Geweke, 1998, for details).

Given that $L\left(y^T;\gamma\right)$, and $L_j\left(y^T;\gamma\right)$ are bounded, an application of Arzelà's Theorem shows the convergence of the marginal likelihood when the approximated likelihood converges pointwise.

Corollary 7. *Under assumptions 1 to 9, if* $\varphi_j\left(\cdot,\cdot;\gamma\right) \to \varphi\left(\cdot,\cdot;\gamma\right)$ *and* $g_j\left(\cdot,\cdot;\gamma\right) \to g\left(\cdot,\cdot;\gamma\right)$ *for all* γ*, then* $p_j\left(y^t\right) \to p\left(y^t\right)$.

The second object of interest for Bayesians is the posterior distribution of the parameters. Given some prior distribution of the parameters, $\pi\left(\gamma\right)$, the posterior is given by:

$$p\left(\gamma|y^t\right) \propto L\left(y^T;\gamma\right)\pi\left(\gamma\right)$$

if we have the exact likelihood, and

$$p_j\left(\gamma|y^t\right) \propto L_j\left(y^T;\gamma\right)\pi\left(\gamma\right)$$

if we have the approximated likelihood. Thus, proposition 5 implies that we also have convergence of the posterior as stated by the next corollary.

Corollary 8. *Let* $\gamma \in \Upsilon$. *Under assumptions 1 to 9, if* $\varphi_j\left(\cdot,\cdot;\gamma\right) \to \varphi\left(\cdot,\cdot;\gamma\right)$ *and* $g_j\left(\cdot,\cdot;\gamma\right) \to g\left(\cdot,\cdot;\gamma\right)$*, then* $p_j\left(\gamma|y^t\right) \to p\left(\gamma|y^t\right)$.

The posterior distribution of the parameters of the model, beyond its intrinsic interest as 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 functions of interest include loss functions for point estimation and point prediction, indicator functions for percentile statements, moment conditions, predictive intervals, or turning point probabilities.

If we develop the expectation:

$$E\left(h\left(\gamma\right)|y^{t}\right)=\frac{1}{p\left(y^{T}\right)}\int_{\Upsilon}h\left(\gamma\right)L\left(y^{T};\gamma\right)\pi\left(\gamma\right)d\gamma$$

Analogously, for the approximated likelihood, we have:

$$E_{j}\left(h\left(\gamma\right)|y^{t}\right)=\frac{1}{p_{j}\left(y^{T}\right)}\int_{\Upsilon}h\left(\gamma\right)L_{j}\left(y^{T};\gamma\right)\pi\left(\gamma\right)d\gamma.$$

Then, we can prove the following corollary:

Corollary 9. *Under assumptions 1 to 9, if* $\varphi_{j}\left(\cdot,\cdot;\gamma\right)\rightarrow\varphi\left(\cdot,\cdot;\gamma\right)$, $g_{j}\left(\cdot,\cdot;\gamma\right)\rightarrow g\left(\cdot,\cdot;\gamma\right)$ *for all* γ, *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* $E_{j}\left(h\left(\gamma\right)|y^{t}\right)\rightarrow E\left(h\left(\gamma\right)|y^{t}\right)$.

These three corollaries illustrate how most work within the Bayesian framework is covered by simple extensions of the convergence of the likelihood result.

4.5. Limitations of the Result

We have briefly discussed several applications of this section's results, the convergence of the likelihood function. However, important as it is, the result is also limited. Proposition 5 shows pointwise convergence of the likelihood function for any given γ. Unfortunately, pointwise convergence does not imply convergence on the PMLE estimate of γ, since for that result we need uniform convergence of the likelihood. In section 6 we show the additional assumptions needed for uniform convergence. More problematic will be the attempts to show convergence of the estimates of standard errors, since they require statements about the convergence of the derivative of the likelihood.

5. Speed of Convergence of the Likelihood

The objective of this section is to analyze, for any given value of γ, the speed of convergence of the approximated likelihood function, $L_j\left(y^T;\gamma\right)$, to the exact likelihood function, $L\left(y^T;\gamma\right)$. 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:

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

We organize the section as follows. First, we prove in lemma 10 that $p\left(y_t|y^{t-1},W_1^t,S_0;\gamma\right)$ is Lipschitz with respect to S_0 . Second, we use this result in lemma 12 to bound the difference between

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

Finally, we employ lemmas 10 and 12 to bound the difference between the approximated and exact likelihood functions in proposition 13.

Let us introduce some additional assumptions we need in the section:

Assumption 10. *The densities of W_t and V_t are differentiable, with bounded partial derivative, for all γ.*

Assumption 11. *$\varphi\left(\cdot,\cdot;\gamma\right)$ and $g\left(\cdot,\cdot;\gamma\right)$ are twice continuously differentiable, with bounded second partial derivative, for all γ.*

Now we can prove that:

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

Furthermore, we also have that:

Corollary 11. *Let $\gamma \in \Upsilon$. Under assumptions 1, 3, 10, and 11, $p\left(y_t|y^{t-1}, W_1^t, S_0; \gamma\right)$ is Lipschitz with respect to S_0 for all t, with Lipschitz constant L_p.*

Once we have the continuity of $p\left(y_t|y^{t-1}, W_1^t, S_0; \gamma\right)$, the next step is to bound the difference between:

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

because this difference will be a key component when we evaluate the differences between likelihoods.

To do so, we parametrize both $\varphi_j\left(\cdot, \cdot; \gamma\right)$, and $g_j\left(\cdot, \cdot; \gamma\right)$ in the following way; $\varphi_j\left(\cdot, \cdot; \gamma\right) = \varphi\left(\cdot, \cdot; \gamma, \theta_j\right)$, and $g_j\left(\cdot, \cdot; \gamma\right) = g\left(\cdot, \cdot; \gamma, \theta_j\right)$, where $\theta_j \in \Phi$, $\forall j$, where Φ is a compact subset of R^M. The restrictions that this parametrization implies for the family of policy functions that we can study are stated formally in the following assumption.

Assumption 12. *$\varphi_j\left(\cdot, \cdot; \gamma\right) (= \varphi\left(\cdot, \cdot; \gamma, \theta_j\right))$, and $g_j\left(\cdot, \cdot; \gamma\right) (= g\left(\cdot, \cdot; \gamma, \theta_j\right))$ have bounded partial derivatives with respect to θ, as a function of S, W, and V. The bounds are independent of j.*

Lemma 12. *Let $\gamma \in \Upsilon$. Under assumptions 1 to 12, 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:*

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

for all but a finite number of points, for some finite χ, and all t.

In the next proposition we apply Theorem 3.7 of Santos and Peralta-Alva (2003). Before doing so we impose a contractivity condition on φ. This restriction is equivalent to Condition C in Santos and Peralta-Alva (2003).

Condition 1. *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 γ.

Condition C 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 Vigo, 1998), in learning models (Schmalensee, 1975, and Ellison and Fudenberg, 1993) and in some stochastic games (Sanghvi and Sobel, 1976). Also, it is a common condition in the literature on Markov chains (Stenflo, 2001).

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 13. *Let $\gamma \in \Upsilon$, and let condition 1 hold. Under assumptions 1 to 12, if $\left\|\varphi_j(\cdot,\cdot;\gamma) - \varphi(\cdot,\cdot;\gamma)\right\| \leq \delta$ and $\|g_j(\cdot,\cdot;\gamma) - g(\cdot,\cdot;\gamma)\| \leq \delta$, then*

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

for some finite B and L.

Proposition 13 states that the difference between the likelihoods is bounded by a linear function of the length of the sample of observations, T, and the bound on the error in the transition and measurement equation δ. This result means that in order to guarantee convergence in the estimation of dynamic models, the error in the policy function must depend on the length of the sample: the longer the sample, the smaller the policy function error. Otherwise, the bound in the difference between the approximated and the exact likelihood goes to infinity.

The intuition is as follows. Small errors in the policy function accumulate at the same rate at which the sample size grows. This problem is not very relevant if, for example, we are calibrating the model à la Kydland-Prescott and computing simulated moments, since the policy errors can cancel each other out when finding a mean or a variance (we are just stating that a generalized law of large numbers holds, as shown by Santos and Peralta-Alva,

2003). However, the errors in the policy function do not cancel out in the likelihood, since the likelihood records their magnitude regardless of their sign.

What are the practical implications of proposition 13? We highlight three. First, that there is an inherent limitation in the use of linearization methods to estimate dynamic equilibrium models. After Sargent (1989), a large literature has followed the strategy of linearizing a dynamic model and estimating it with the Kalman filter. Examples include Ireland (2002), Landon-Lane (1999), McGrattan, Rogerson, and Wright (1997), Rabanal and Rubio-Ramírez (2003), Schorfheide (2000), and Smets and Wouters (2003), among many others.

However, proposition 13 shows that this linear approach to estimation of nonlinear models is due to fail as the sample size grows. The reason is that linearization (either in levels or in logs) fixes the policy function error, and this error cannot be improved upon without losing the linearity of the state-space representation required by the Kalman filter.[2] As a consequence, as the sample size grows, the bound on the divergence between the exact and approximated likelihood also grows to infinity.

Even if the bias introduced in small samples by the linearization is difficult to gauge in general, our examples in the next section suggest that the error may be quite important for the sample sizes commonly used in macroeconomics (quarterly, postwar U.S. data, around 200 observations). Proposition 13 and the numerical evidence should be interpreted, at least, as word of caution regarding the indiscriminate use of linearization. Also, it suggests that justifying linearization methods based on small errors in the policy function may be misleading for estimation purposes.

A second implication of proposition 13 is that when we use nonlinear methods to solve and estimate a dynamic model, we may want to make the accuracy of the solution a function of the sample size. Larger samples are, of course, more informative than smaller ones, but to squeeze the extra information, we need to avoid the accumulation of policy function errors over the sample.

[2] Tricks such as bias correction in the linearization (i.e., linearizing around a point that is not the deterministic steady state to get a more accurate solution as in Collard and Juillard, 2001) or changes of variables (Fernández-Villaverde and Rubio-Ramírez, 2004c) are not a solution to this problem because they only make δ smaller but not dependent on the size of the sample.

The third implication is that when we build likelihood ratios between a model for which we can compute the exact likelihood (for example a VAR) and a model for which we need to approximate the likelihood (like a dynamic general equilibrium model), the accumulation of errors in the likelihood of the latter model as the sample size grows may lead to a likelihood ratio test (or analogously to a Bayes factor) that delivers an incorrect conclusion.

Proposition 13 also suffers from limitations. First, as all bounds, it is not clear whether it is tight and, consequently, informative for practitioners. Second, it does not offer a constructive way to evaluate the different constants in the bound. Finally, the bound depends on δ, an unknown constant, because to find it we will need to use the exact policy function that, by the nature of our exercise, is unknown.

To partially address the first two limitations, we will offer some numerical evidence in section 6 that indicates that the bound is informative and that we can estimate the constants for certain examples. With respect to the third limitation, in some cases we can link δ with the Euler Equation errors, which are easily computed. Santos (2000) shows that for a class of dynamic optimization problems, the approximation error of the policy function δ is of the same order of magnitude as the size of the Euler equation residual. As a consequence, we could substitute δ for an Euler error estimate and obtain a bound of the same order of magnitude.

6. Convergence of the Maximum Likelihood Estimates

In section 4 we showed 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 was that since our convergence was pointwise, we could not 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 δ, there is an N such that $\forall j \geq N$,

$$\left\|\varphi_j(\cdot,\cdot;\cdot) - \varphi(\cdot,\cdot;\cdot)\right\| \leq \delta \text{ and } \left\|g_j(\cdot,\cdot;\cdot) - g(\cdot,\cdot;\cdot)\right\| \leq \delta$$

for all S, W, V, and γ, then the likelihood function also converges uniformly, implying the convergence of the PMLE.

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

Assumption 13. *The densities of W_t and V_t are continuous with respect to γ.*

Analogously, we modify assumptions 1, 6, 10, and 11:

Assumption 14. *The bounds in assumptions 1, 6, 10, and 11 are independent of γ.*

And, finally, we substitute 12 by the following new assumption:

Assumption 15. *$\varphi_j(\cdot,\cdot;\cdot)\,(=\varphi(\cdot,\cdot;\cdot,\theta_j))$ and $g_j(\cdot,\cdot;\cdot)\,(=g(\cdot,\cdot;\cdot,\theta_j))$ have bounded partial derivatives with respect to θ, as a function of S, W, V, and γ. The bounds are independent of j.*

These new three assumptions assume that all the bounds are uniform on γ. Armed with our stronger assumptions, we can modify lemma 12 to get:

Lemma 14. *Under assumptions 1 to 11, and assumptions 13 to 15, if the policy functions converge uniformly in the parameter space, then there is an N such that $\forall j \geq N$:*

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

for all but a finite number of points, for some finite χ, all t, and all γ.

We can also modify proposition 13 to get:

Proposition 15. *Let condition 1 hold. Under assumptions 1 to 11, and assumptions 13 to 15, 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 some finite B and L, and all γ.

Proposition 15 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 maximum and, therefore, of the PMLE. Formally:

Corollary 16. *Let condition 1 hold. Under assumptions 1 to 11, and assumption 15, 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 using classical methods.

7. Three Applications

In this section we present three examples, ordered in terms of complexity, to illustrate how our results hold in real-life applications. First, we study the case where the exact optimal policy function follows a simple autoregressive process. The approximated policy function is also an autoregressive process but with slightly different parameters. This example gives us a feeling for how the results work in a stylized environment without the need to be explicit

about the underlying economic theory. Then, we study a linearized neoclassical growth model. We look at a case where, instead of the exact linear policy function, we also employ a linear policy rule, but with slightly different coefficients. Finally, we analyze a nonlinear neoclassical growth model. Thanks to a carefully chosen calibration, this model has a closed-form solution that allows us to evaluate the likelihood. We compare the exact likelihood with the one we would evaluate if we computed an approximated optimal policy function using value function iteration on a grid.

7.1. An AR(1) Optimal Policy Function Example

As mentioned before, we first study a stylized environment. This simple example, however, is already rich enough to highlight most of the theoretical results in the paper.

Let $y^T = (y_0, ..., y_T)$ be some given data, where $y_t \in R$ for all $0 \leq t \leq T$. Let us assume that there is an unspecified dynamic economic model that implies the following optimal policy function for the agent:

$$y_t = \rho y_{t-1} + \sigma \varepsilon_t, \tag{13}$$

where $\varepsilon_t \sim iid\ \mathcal{N}(0,1)$. Let $\gamma = \{\rho, \sigma\}$. We will call this the exact policy function of the model.

Note that we can write model (13) in the state space form of equations (6) and (7), making (6) equal to equation (13) and (7) equal to the identity function.

The likelihood for this model given data y^T is equal to:

$$L\left(y^T; \gamma\right) = \frac{1}{\sqrt{\frac{2\pi\sigma^2}{1-\rho^2}}} \exp\left(-\frac{1}{2}\frac{1-\rho^2}{\sigma^2} y_0\right) \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y_t - \rho y_{t-1}\right)^2}{2\sigma^2}\right).$$

Now, let us imagine that, instead of the exact model (13), for some reason, the economist can compute only the approximated policy function:

$$y_t = \varphi_j\left(y_{t-1}, \varepsilon_t; \gamma\right), \tag{14}$$

where $\varphi_j\left(y_{t-1}, \varepsilon_t; \gamma\right) = \left(\rho + \nu_j\right) y_{t-1} + \left(\sigma + \omega_j\right) \varepsilon_t$ and ν_j and ω_j are different from zero. Let

us define $\rho_j = \rho + \nu_j$ and $\sigma_j = \sigma + \omega_j$. We will call (14) the approximated policy function of the model.

Given our data y^T, the likelihood of the model (14) is:

$$L_j\left(y^T;\gamma\right) = \frac{1}{\sqrt{\frac{2\pi\sigma_j^2}{1-\rho_j^2}}}\exp\left(-\frac{1}{2}\frac{1-\rho_j^2}{\sigma_j^2}y_0\right)\prod_{t=1}^{T}\frac{1}{\sqrt{2\pi\sigma_j^2}}\exp\left(-\frac{\left(y_t-\rho_j y_{t-1}\right)^2}{2\sigma_j^2}\right).$$

In the terminology of section 5, we write $\delta = \max\left(\left|\nu_j\right|,\left|\omega_j\right|\right)$. The data y^t will be a sample of 1000 observations generated randomly from process (13).[3]

Figure 6.1.1 plots the absolute value difference between the likelihoods of the exact and the approximated model, $\left|L\left(y^T;\gamma\right) - L_j\left(y^T;\gamma\right)\right|$, as a function of δ for a range between 0 and 0.3. We can see how, as δ goes to zero, the absolute value difference between the likelihoods also goes to zero. This result matches the theoretical predictions reported in proposition 5 of section 4.

We illustrate next the results from section 5. Proposition 13 states that for a fixed sample size, the absolute value difference between the likelihoods of the exact and the approximated model is proportional to δ. Therefore, if we reduce δ by half, the absolute value difference between the likelihoods should also be approximately reduced by half. This result is confirmed by the simulation. Table 6.1.1 reports the absolute value difference between the likelihoods for different values of δ.

[3]For simplicity of exposition, in the three applications, we are omitting 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 ε_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 forget about the restriction that in the computer, we can use only the computable reals instead of the real line.

Table 6.1.1: Absolute Value Difference between the Likelihoods as a Function of δ

δ	$\left\|L\left(y^T;\gamma\right) - L_j\left(y^T;\gamma\right)\right\|$
0.3000	48.884
0.1500	22.162
0.0750	10.163
0.0375	4.8075
0.0188	2.3298

A second implication of proposition 13 is that for a fixed δ, 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 δ. Figure 6.1.2 shows the absolute value difference between the likelihoods for different δ as a function of the sample size. As expected, the larger the sample size, the larger the difference for any value of δ. Figure 6.1.2 also shows, to better emphasize the slope, a fitted line to the likelihood. As expected, the slopes are proportional to δ.

These two results, the linearity of the difference of the likelihoods on δ and the sample size, emphasize the usefulness for practitioners of the bound in proposition 13.

We checked that all of the results above hold for different sequences of δ, for changes of only one of the two parameters, for different values of ρ and σ, for different data y^t, and for different sample sizes. In that sense, our reported numbers are to be interpreted as a representative sample of our findings.

7.2. A Linear Neoclassical Growth Model Example

Now we study an example more explicitly motivated by economic theory. We pick the stochastic neoclassical growth model with leisure, linearize it around the steady state, and ask what will happen if we incur an error in the coefficients of the optimal linear policy function.

Let $y^T = (y_0, ..., y_T)$ be some given data, where $y_t \in R^3$ for all $0 \leq t \leq T$. The components of y_t are output, hours worked, and gross investment. Let us assume 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 Λ is a diagonal matrix with σ_1^2, σ_2^2, and σ_3^2, as diagonal elements.

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

$$U = E_0 \sum_{t=1}^{\infty} \beta^{t-1} \frac{\left(c_t^{\xi} (1 - l_t)^{1-\xi}\right)^{1-\tau}}{1 - \tau}$$

where $\beta \in (0, 1)$ is the discount factor, τ controls the elasticity of intertemporal substitution, ξ pins down labor supply, and E_0 is the conditional expectation operator.

The only good in this economy is produced according to the 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, λ 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 η 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 $\gamma = \{\tau, \alpha, \beta, \rho, \xi, \eta, \lambda, \sigma, \sigma_1, \sigma_2, \sigma_3\}$.

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 policy functions for consumption $c(\cdot, \cdot)$, labor $l(\cdot, \cdot)$, and next period's capital $k'(\cdot, \cdot)$ that deliver the optimal choices as functions of the two state variables, capital and the technology level.

A way to solve the model is to linearize its first order conditions and resource constraint around its deterministic steady state. Such procedure delivers an optimal linear policy function. Then, the state-space representation has the following form:

$$S_t = G(\gamma) + A(\gamma) S_{t-1} + B(\gamma) W_t, \tag{15}$$

and

$$Y_t = F(\gamma) + C(\gamma) S_t + D(\gamma) V_t, \tag{16}$$

where $A(\gamma)$, $B(\gamma)$, $C(\gamma)$, $D(\gamma)$, $G(\gamma)$, and $F(\gamma)$ are matrices with the required dimensions,

which depend on the parameters of the model collected in vector γ. Note how this representation is nothing more than a particular case of (6) and (7), where $W_t = \epsilon_t$, and V_t are three measurement errors. Let $L(y^t; \gamma)$ be the likelihood function associated with (15) and (16). With a bit of abuse of the language, we will call this state-space representation the exact model. This is also the sense in which we name this example a linear neoclassical growth model.

Let us now assume that we cannot evaluate (15) and (16), but only approximated versions of them of the form:

$$S_t = G_j(\gamma) + A_j(\gamma) S_{t-1} + B_j(\gamma) W_t, \tag{17}$$

and

$$Y_t = F_j(\gamma) + C_j(\gamma) S_t + D_j(\gamma) V_t, \tag{18}$$

where again, $W_t = \epsilon_t$, and V_t are three independent measurement errors with mean zero and variances σ^1, σ^2, and σ^3. Then we have

$$\begin{aligned} \delta_1 &= \max\left(\|G_j(\gamma) - G(\gamma)\|, \|A_j(\gamma) - A(\gamma)\|, \|B_j(\gamma) - B(\gamma)\|\right), \\ \delta_2 &= \max\left(\|F_j(\gamma) - F(\gamma)\|, \|C_j(\gamma) - C(\gamma)\|, \|D_j(\gamma) - D(\gamma)\|\right), \end{aligned}$$

and $\delta = \max(\delta_1, \delta_2)$. Finally, let $L_j\left(y^T; \gamma\right)$ be the likelihood function of y^Tassociated with (17) and (18).

We generate a sample size of 200 observations, roughly the size of postwar U.S. macro data, to give a feeling for the behavior of the likelihood in realistic applications. For the same reason, we set the parameters at standard values: $\tau = 2$, $\alpha = 0.4$, $\beta = 0.989$, $\rho = 0.96$, $\xi = 0.356$, $\eta = 0.02$, $\lambda = 1$, $\sigma = 0.007$, $\sigma_1 = 0.016$, $\sigma_2 = 0.011$ and $\sigma_3 = 0.087$. To perturb our matrices, $A(\gamma)$, $B(\gamma)$, $C(\gamma)$, $D(\gamma)$, $G(\gamma)$, and $F(\gamma)$, we add to each of their elements a normal random number. This perturbation, plus the use of the sup norm, implies that δ is equal to the biggest of these random numbers. By controlling the standard deviation of the normal random number, we can play with the size of δ.[4]

[4]In real-life applications of the linear models, economists, in fact, incur a trivially small δ because we use floating point arithmetic.

Figure 6.2.1 plots the absolute value difference between the likelihoods of the exact and the approximated model $\left|L\left(y^T;\gamma\right) - L_j\left(y^T;\gamma\right)\right|$ as a function of δ for our sample. As before, as δ goes to zero, the absolute value difference between the likelihoods also goes to zero.

Table 6.2.1 reports the absolute value difference between the likelihoods for different values of δ and for a fixed sample size. These numbers replicate the results of proposition 13, i.e., the absolute value difference between the likelihoods is roughly proportional to δ, showing once more the informativeness of the bound.

Table 6.2.1: Absolute Value Difference between the Likelihoods as a Function of δ

δ	$\left\|L\left(y^T;\gamma\right) - L_j\left(y^T;\gamma\right)\right\|$
$2.36e-006$	1225.8
$1.61e-006$	788.37
$1.21e-006$	620.19
$1.10e-006$	542.6
$1.00e-006$	401.72

Figure 6.2.2 shows the absolute value difference between the loglikelihoods for different δ's as a function of the sample size. 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. Figure 6.2.2. reveals again that, as the sample size grows, the difference between the likelihoods for any value of δ becomes larger and that the difference grows at a linear rate.

7.3. A Nonlinear Neoclassical Growth Model Example

In the previous exercise we assumed that both the exact model and the approximated model were linear dynamic systems. This assumption allowed us to evaluate the exact and approximated likelihood with the Kalman filter. However, we did not motivate why we were using an approximated model or how the perturbations came about. In this exercise, we address these issues.

We will propose a version of the neoclassical growth model for which we know the likelihood because the model has a closed-form solution in logs suitable to evaluation with the Kalman filter. We will study what happens when the researcher does not know this exact closed-form solution, and, instead, she solves for the optimal policy functions using value function iteration. Since the solution from the value function iteration is not linear, the state-space representation is also nonlinear. As a consequence, we cannot apply the Kalman filter. Instead, we use a Sequential Monte Carlo method. Fernández-Villaverde and Rubio-Ramírez (2004a) show how to implement that technique in a model similar to this one. This implies that we will have two approximation errors: one in the computation of the optimal policy function and a second one in the Sequential Monte Carlo. We discuss how we address this problem below.

We take the neoclassical growth model in the previous section, where a representative household maximizes:

$$U = E_0 \sum_{t=1}^{\infty} \beta^{t-1} \frac{\left(c_t^{\xi} \left(1 - l_t\right)^{1-\xi}\right)^{1-\tau}}{1-\tau},$$

subject to the resource constraint, $c_t + i_t = e^{z_t} \lambda k_t^{\alpha} l_t^{1-\alpha}$, the law of motion for capital $k_{t+1} = i_t + (1-\eta)k_t$, and the evolution of technology $z_t = \rho z_{t-1} + \epsilon_t$.

We set $\tau = 1$ and, unrealistically but rather useful for our point, $\eta = 1$. In this case, the income and the substitution effect to a productivity shock in labor supply exactly cancel each other. Consequently, l_t is constant over time.

Exploiting this feature of labor supply, we can use the method of undetermined coefficients to find the exact policy function for labor:

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

and for capital $k_{t+1} = \alpha\beta e^{z_t} \lambda k_t^{\alpha} l^{1-\alpha}$.

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

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

As observables, we assume that we have data on 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 Λ is a diagonal matrix with σ_1^2 and σ_2^2, as diagonal elements.:

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

We drop labor from the observables because it is constant over time, and any movement in it will be trivially attributed to measurement error. We can apply the Kalman filter to the transition and measurement equations above and evaluate the exact likelihood of the model given some data. Finally, let $\gamma = \{\alpha, \beta, \lambda, \rho, \xi, \sigma, \sigma_1, \sigma_2\}$.

Now, let us suppose that we have a researcher who does not know the exact solution for the optimal policy function for capital (although, to simplify, we assume that the researcher realizes that labor is constant). Instead, the economist solves the social planner's problem using value function iteration over a grid of points of capital and productivity, and linear interpolation. 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 assumption 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 more points are introduced in the grid, we have:

$$g_j(\rho z_{t-1} + \epsilon_t, k_t; \gamma) \to \alpha\beta e^{z_t} \lambda k_t^{\alpha} l^{1-\alpha}.$$

The approximated likelihood function to evaluate is implied by the state-space form:

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

and

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

The nonlinearity of this state-space form asks for a Sequential Monte Carlo algorithm to evaluate the likelihood function of the approximated model.

How different are the likelihoods of the approximated and the exact model? To answer this question, we generate a sample size of 200 observations with the calibration: $\alpha = 0.4$, $\beta = 0.989$, $\lambda = 1$, $\rho = 0.95$, $\xi = 0.356$, $\sigma = 0.007$, $\sigma_1 = 0.001$, $\sigma_2 = 0.002$, and the scale factor λ to get $\lambda l_t^{1-\alpha} = 1$. Then, we solve the model using three different grids: a coarse one with 10 points in the capital axis, an intermediate grid with 100 points, and fine grid with 1000 points. Along the technology axis we have 40 points in the grid, and we evaluate the corresponding integral using quadrature, so the support of the technology shock is continuous. 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 parametrization, the δ's are as follows:

Table 6.3.1: δ as a Function of the Capital Grid

Capital Grid Points	δ
10	0.008545
100	0.000674
1000	0.000076

To interpret this number it is useful to think about its welfare implications. Even with only 10 points in the capital grid, this problem is sufficiently well behaved that the welfare

loss around the deterministic steady state from using the approximated policy rule instead of the exact one is less than one-tenth of a percent in terms of consumption.

Figure 6.3.1 plots the absolute value difference between the approximated and the exact loglikelihoods as a function of the sample size for three different capital grids. To minimize the impact of the error coming from the Sequential Monte Carlo, we created a swarm of 100.000 particles, well beyond the 20.000 required to achieve stability of the estimation of the likelihoods (see Fernández-Villaverde and Rubio-Ramírez, 2004a, for details on this issue). In that way, we computed that the difference in the likelihood attributable to the simulation is, with more than a 99 percent probability, several orders of magnitude smaller than the reported total differences in likelihoods.

As in the previous two examples, we see how the larger the sample size, the larger the difference between the likelihoods for any value of δ and how that the difference 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 capital grid of 10 points even if a naive welfare comparison criterion would have suggested that the approximation was acceptable. In contrast, when we use 1000 points, the approximated likelihood stays very close to the exact one, even at the end of the sample. This exercise emphasizes that a control of the accuracy of the solution of the policy function as a function of the sample size is important to guarantee a good behavior of the likelihood.

Because of space considerations we do not offer a full study of the implications of the differences in the likelihood for point estimates. We refer the interested reader to Fernández-Villaverde and Rubio-Ramírez (2004b) in which the authors present a thorough analysis of the impact on estimation of using different approximations to the optimal policy functions. Suffice it to say that Fernández-Villaverde and Rubio-Ramírez (2004b) document important differences in point estimates and that they show that those differences have a relevant impact on the empirical predictions of the model.

8. Conclusions

In this paper we have studied the consequences of using approximated likelihood functions instead of the exact likelihoods when we estimate computed dynamic 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 obtain more data. Our three applications have documented the quantitative importance of our findings.

There are several additional issues that we have not considered and that we leave for future analysis. First, it would be important to eliminate the assumption of continuity of the transition and measurement equations. A large class of models in economics, especially in micro applications, implies choices with jumps and discontinuities. Second, we could relax some of the assumptions required to deliver uniform convergence of the likelihood function and the consequent convergence of the maximum likelihood estimates. Related to this, establishing results concerning the convergence of standard error estimates will complete the findings regarding classical estimation. Finally, it would be useful to extend the framework of this paper to cover game-theoretic settings that may create, among other characteristics, a multiplicity of equilibria. The econometric advances in Jofre-Bonet and Pesendorfer (2003), Aguirregabiria and Mira (2004), Bajari, Hong and Ryan (2004), and Pakes, Ostrovsky and Berry (2004), among others, open an important field of research in empirical applications where some aspects of the dynamic model are approximated. Results in this area will help to fine tune the performance of the developed estimators.

9. Appendix

We include in this appendix the proofs of the results in the paper.

Proof of Lemma 2.
Let $\gamma \in \Upsilon$. Assumption 3 implies that:

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

but we know from assumption 1 that v_t and $w_{2,t}$ are continuous functions of S_0 and that $\left|dy\left(v_t, w_{2,t}; \gamma\right)\right|$ 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 $p\left(y_t|y^{t-1}, W_1^t, S_0; \gamma\right) \in C\left(S_0\right)$. ■

Proof of Lemma 3. As in the proof of the previous lemma, but substituting the jacobian $\left|dy\left(v_t, w_{2,t}; \gamma\right)\right|$ for the approximated one $\left|dy_j\left(v_t, w_{2,t}; \gamma\right)\right|$. Since assumption 6 ensures that $\left|dy_j\left(v_t, w_{2,t}; \gamma\right)\right|$ is continuous at all but in a finite number of points, we have that $p\left(y_t|y^{t-1}, W_1^t, S_0; \gamma\right) \in C\left(S_0\right)$ except in a finite number of points. ■

In the proof of Lemma 4 we use the following well known theorems (see Dieudonné for their proofs, 1960):

Theorem 17. *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 convergence subsequence $\{a''_n\} \subset \{a'_n\}$ such that $a''_n \to a$.*

Theorem 18. *If $f_n \to f$ in the sup norm, and $f'_n \to g$ in the sup norm, then $g = f'$.*

Proof of Lemma 4. Assumption 9 implies that $\left\{\varphi_j\left(\cdot, \cdot; \gamma\right)\right\}$, and $\left\{g_j\left(\cdot, \cdot; \gamma\right)\right\}$ have uniformly bounded second derivatives, then $\left\{d\varphi_j\left(\cdot, \cdot; \gamma\right)\right\}$ and $\left\{dg_j\left(\cdot, \cdot; \gamma\right)\right\}$ is a family of equicontinuous functions. Therefore, the Arzelà-Ascoli theorem implies that every subsequence of $\left\{\varphi_j\left(\cdot, \cdot; \gamma\right)\right\}$ and $\left\{g_j\left(\cdot, \cdot; \gamma\right)\right\}$ have a convergence subsequence in the C^1 topology.[5] Since $\left\{\varphi_j\left(\cdot, \cdot; \gamma\right)\right\}$ and $\left\{g_j\left(\cdot, \cdot; \gamma\right)\right\}$ converge to $\varphi\left(\cdot, \cdot; \gamma\right)$ and $g\left(\cdot, \cdot; \gamma\right)$ respectively,

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

every subsequence of $\{\varphi_j(\cdot,\cdot;\gamma)\}$ and $\{g_j(\cdot,\cdot;\gamma)\}$ has a convergence subsequence in the sup norm to $\varphi(\cdot,\cdot;\gamma)$ and $g(\cdot,\cdot;\gamma)$. Therefore, theorem 18 implies that every subsequence of $\{d\varphi_j(\cdot,\cdot;\gamma)\}$ and $\{dg_j(\cdot,\cdot;\gamma)\}$ has a convergence subsequence in the sup norm to $d\varphi(\cdot,\cdot;\gamma)$ and $dg(\cdot,\cdot;\gamma)$. Hence, this last result and theorem 17 imply that $d\varphi_j(\cdot,\cdot;\gamma) \to d\varphi(\cdot,\cdot;\gamma)$ and $dg_j(\cdot,\cdot;\gamma) \to dg(\cdot,\cdot;\gamma)$. ■

Proof of Proposition 5. Let $\gamma \in \Upsilon$. The proof is divided into two steps. First, we show that:

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

for all t, except in a finite number of points. This is a technical result that we will use in step two. Second, we prove that:

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

Step 1. We show convergence of $p_j(y_t|y^{t-1},W_1^t,S_0;\gamma)$. Assumption 3 allows us to write:

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

while by assumption 7 we have:

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

except in a finite number of points.

First, remember that assumption 1 implies that $\varphi(\cdot,\cdot;\gamma)$, $g(\cdot,\cdot;\gamma)$, and their partial derivatives are continuous. Second, note that assumption 6 states that $\varphi_j(\cdot,\cdot;\gamma)$ and $g_j(\cdot,\cdot;\gamma)$ are continuous, while their partial derivatives are continuous at all but in 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 in a finite number of points, and we can assert

that:

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

except in a finite number of points.

Step 2. Assumptions 4 and 8 allow us to write:

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

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

and

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

Define

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

therefore:

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

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 2, $f_T\left(S_0;\gamma\right)$ is continuous. Therefore, we can apply corollary 3.3 of Santos and Peralta-Alva (2003) to show that:

$$\prod_{t=1}^{T}\widetilde{p}_j\left(y_t|y^{t-1};\gamma\right)\to\prod_{t=1}^{T}p\left(y_t|y^{t-1};\gamma\right). \tag{19}$$

If we define

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

we get:

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

Note that $W_{1,t}$ has bounded support and bounded density. Also, lemma 3 shows that $p_j\left(y_t|y^{t-1},W_1^t,S_0;\gamma\right)$ is continuous except in a finite number of points, with bounded support, and hence it is bounded. Therefore $f_{j,T}\left(S_0;\gamma\right)$ is bounded. In addition, step 1 shows:

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

except in a finite number of points. Hence, $f_{j,T}\left(S_0;\gamma\right)\rightarrow f_T\left(S_0;\gamma\right)$, 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,

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

$$\leq\int\left|f_{j,T}\left(S_0;\gamma\right)-f_T\left(S_0;\gamma\right)\right|\mu_j^*\left(dS_0;\gamma\right)<\varepsilon \tag{20}$$

and we can conclude that:

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

To close the proof, we put together the convergence results (19) and (21). ■

Proof of Corollary 6. Let $\gamma, \gamma' \in \Upsilon$. Proposition 5 shows that:

$$L_j\left(y^T;\gamma\right) \to L\left(y^T;\gamma\right)$$

and

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

Therefore, since assumption 4 implies that $L\left(y^T;\gamma\right) \geq \xi > 0$, and assumption 8 states that $L_j\left(y^T;\gamma\right) \geq \xi > 0$ for all j, we have:

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

■

Proof of Corollary 7. Let $\gamma \in \Upsilon$. Proposition 5 shows that:

$$L_j\left(y^T;\gamma\right) \to L\left(y^T;\gamma\right).$$

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) 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 8. Let $\gamma \in \Upsilon$. Proposition 5 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 9. Let $\gamma \in \Upsilon$. Proposition 5 shows that:

$$L_j\left(y^T;\gamma\right) \to L\left(y^T;\gamma\right).$$

and corollary 7 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 10. Let $\gamma \in \Upsilon$. To prove that $p\left(y_t|y^{t-1},W_1^t,S_0;\gamma\right)$ is continuously differentiable with respect to S_0, we need to show that:

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

exists and is continuous for all i.

Assumption 3 allows us to write:

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

for all t. Since in addition V_t and $W_{2,t}$ have bounded densities, assumptions 1, 10, and 11 imply that $\frac{\partial p\left(y_t|y^{t-1},W_1^t,S_0;\gamma\right)}{\partial S_{0,i}}$ exists and it is bounded for all t and all i. ■

Proof of Lemma 12. Let $\gamma \in \Upsilon$. Let s_0 and w_1^t be a realization of the random variables S_0 and W_1^t. Let (v^t,s^t,w_2^t) be the unique solution to the following system of equations:

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

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

and

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

and let $\left(v_j^t, s_j^t, w_{j,2}^t\right)$ be the unique solution to the approximated system of equations:[6]

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

$$y_m = g_j\left(S_m, V_m; \gamma\right) \qquad \text{for } m = 1, 2, ...t,$$

and

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

Assumption 1 implies that φ 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 $\left\|g_j\left(\cdot,\cdot;\gamma\right) - g\left(\cdot,\cdot;\gamma\right)\right\| \leq \delta$, the implicit function theorem of Schwartz (see theorem G.2.3, page 32, Mas-Colell, 1985) implies that there exists a $\lambda\left(s_0, w_1^t\right)$ such that:

$$\left\|\left(v_j^t, s_j^t, w_{j,2}^t\right) - \left(v^t, s^t, w_2^t\right)\right\| \leq \lambda\left(s_0, w_1^t\right)\delta. \tag{22}$$

Since we are using the sup norm, equation (22) holds for all t.

Notice that $\lambda\left(s_0, w_1^t\right)$ depends on the the derivatives of $\varphi_j\left(\cdot,\cdot;\gamma\right)$ and $g_j\left(\cdot,\cdot;\gamma\right)$ with respect to θ_j. These derivatives are bounded independently of j by assumption 12. 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.

Since V_t and W_t have continuous densities, assumption 10 implies that those densities are absolutely continuous. Then, $\exists\varepsilon$ such that:

$$\left|p\left(v_{j,t};\gamma\right)p\left(w_{j,2,t};\gamma\right) - p\left(v_t;\gamma\right)p\left(w_{2,t};\gamma\right)\right| \leq \varepsilon\delta, \tag{23}$$

for all s_0 and w_1^t. As before, since we are using the sup norm, equation (23) also holds for all t.

[6]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 relationship explicit.

Assumption 11 delivers that the determinant of the Jacobian 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:

$$||dy(v_{j,t}, w_{j,2,t};\gamma)| - |dy(v_t, w_{2,t};\gamma)|| \leq L_y \lambda \delta, \tag{24}$$

for all s_0 and w_1^t.

Assumption 9 and the fact that $\left\|\varphi_j(\cdot,\cdot;\gamma) - \varphi(\cdot,\cdot;\gamma)\right\| \leq \delta$ and $\|g_j(\cdot,\cdot;\gamma) - g(\cdot,\cdot;\gamma)\| \leq \delta$ imply that $\left\|d\varphi_j(\cdot,\cdot;\gamma) - d\varphi(\cdot,\cdot;\gamma)\right\| \leq \delta$ and $\|dg_j(\cdot,\cdot;\gamma) - dg(\cdot,\cdot;\gamma)\| \leq \delta$ except in a finite number of points. Then, by assumptions 1 and 6, we know that $\exists \Psi_1$ such that:

$$|dy_j(v_{j,t}, w_{j,2,t};\gamma)[r,s] - dy(v_{j,t}, w_{j,2,t};\gamma)[r,s]| < \Psi_1 \delta \tag{25}$$

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 to $n \times n$ matrices such that $|A[i,j] - B[i,j]| < \Psi_1 \delta$ and $|A[i,j]|, |B[i,j]| < \Psi_2$, then $||A| - |B|| < n! n \Psi_2^{n-1} \Psi_1 \delta$. In addition, assumptions 1 and 6 also imply that φ_j, φ, g_j, and g are Lipschitz. Therefore $\exists \Psi_2$ such that:

$$||dy_j(v_{j,t}, w_{j,2,t};\gamma)| - |dy(v_{j,t}, w_{j,2,t};\gamma)|| \leq n! n \Psi_2^{n-1} \Psi_1 \delta, \tag{26}$$

for all s_0 and w_1^t, except in a finite number of points.

Using equations (24) and (26) we get:

$$||dy_j(v_{j,t}, w_{j,2,t};\gamma)| - |dy(v_t, w_{2,t};\gamma)|| \leq \left(n! n \Psi_2^{n-1} \Psi_1 + L_y \lambda\right) \delta, \tag{27}$$

for all s_0 and w_1^t.

Now let $\Psi_3 = \left(n! n \Psi_2^{n-1} \Psi_1 + L_y \lambda\right)$. Since

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

except in a finite number of points and

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

we can put together equations (23) and (27) to find:

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

for all s_0 and w_1^t, except in a finite number of points.

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

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

for all s_0 and w_1^t, but in a finite number of points. If we let $\chi=B\left(\varepsilon+\Psi_3\right)$, the lemma is proved. ■

Proof of Proposition 13. Let $\gamma\in\Upsilon$. Define $f_T\left(S_0;\gamma\right)$ as in the proof of proposition 5 and note that:

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

is bounded because corollary 11 bounds $\frac{\partial p\left(y_t|,y^{t-1},W_1^t,S_0;\gamma\right)}{\partial S_{0,i}}$ for all t and i, and lemma 2 bounds $p\left(y_s|y^{s-1},W_1^s,S_0;\gamma\right)$ for all s. Therefore, $f_T\left(S_0;\gamma\right)$ is Lipschitz for all t with Lipschitz constant L (the Lipschitz constant will be different for each t, but since t is finite, we can set a global L).

Therefore, since condition 1 holds, we can apply Theorem 3.7 of Santos and Peralta-Alva

(2003) to $f_T(S_0;\gamma)$ to get:

$$\left|\prod_{s=1}^{T} p\left(y_t|y^{t-1};\gamma\right) - \prod_{s=1}^{T} \widetilde{p}_j\left(y_t|y^{t-1};\gamma\right)\right| = \frac{L\delta}{1-\alpha}. \tag{28}$$

Note now that using the values for the likelihoods in the proof of proposition 5, we have:

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

$$= \int \left(\int \prod_{t=1}^{T} p\left(W_{1,t};\gamma\right) \left(p_j\left(y_t|y^{t-1},W_1^t,S_0;\gamma\right) - p\left(y_t|y^{t-1},W_1^t,S_0;\gamma\right)\right) dW_1^t \right) \mu_j^*\left(dS_0;\gamma\right). \tag{29}$$

Lemmas 2 and 3 show that $p\left(y_t|y^{t-1},W_1^t,S_0;\gamma\right)$ and $p_j\left(y_t|y^{t-1},W_1^t,S_0;\gamma\right)$ are bounded for all t and j. Thus, we can define a constant B such that:

$$\int \left(\int \prod_{t=1}^{T} p\left(W_{1,t};\gamma\right) B \sum_{t=1}^{T} \left|p_j\left(y_t|y^{t-1},W_1^t,S_0;\gamma\right) - p\left(y_t|y^{t-1},W_1^t,S_0;\gamma\right)\right| dW_1^t \right) \mu_j^*\left(dS_0;\gamma\right) \tag{30}$$

is an upper bound to (29).

Lemma 12 shows that:

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

for all t, and for all W_1^t, and S_0 but for a finite number of points. Therefore,

$$\left|\prod_{s=1}^{T} p_j\left(y_t|y^{t-1};\gamma\right) - \prod_{s=1}^{T} \widetilde{p}_j\left(y_t|y^{t-1};\gamma\right)\right| \leq TB\chi\delta, \tag{31}$$

Putting together (28) and (31), we have:

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

■

Proof of Lemma 14. The proof is a modification of the proof of lemma 12. The argument is the same except in the following points:

1. We use assumption 15 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 γ, s_0 and w_1^t.

2. We need assumptions 10, 13, and 14 to show that the densities of V_t and W_t are absolutely continuous.

3. We need assumptions 11 and 14 to show that the determinant of the Jacobian of y_t with respect to $V_t, W_{2,t}$, $|dy\left(\cdot,\cdot;\cdot\right)|$, is Lipschitz. Also, by assumption 14, the Lipschitz constant L_y is independent of γ.

4. We need assumptions 1, 6, and 14 to show existence of a constant Ψ_1, independent of γ, 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 γ, s_0 and w_1^t, except in a finite number of points.

5. We need assumptions 1, 6, and 14 to prove existence of a constant Ψ_2, independent of γ, such that:

$$\left||dy_j\left(v_{j,t}, w_{j,2,t}; \gamma\right)| - |dy\left(v_{j,t}, w_{j,2,t}; \gamma\right)|\right| \leq n!n\Psi_2^{n-1}\Psi_1\delta, \tag{32}$$

   for all γ, s_0 and w_1^t, but in a finite number of points.

6. Since V, W, and γ have compact support, assumption 13 is important to guaranty that $p\left(v;\gamma\right)$, $p\left(w_2;\gamma\right)$ are bounded functions of γ, s_0 and w_1^t. Assumptions 1 and 14 imply that $|dy\left(v, w_2;\gamma\right)|$ is also a bounded function of γ, s_0 and w_1^t.

■

Proof of Proposition 15.

The proof is a modification of the proof of proposition 13. The argument is the same except:

1. We use assumptions 13 and 14 to make the bounds $\frac{\partial f_T(S_0;\gamma)}{\partial S_{0,i}}$ independent of γ. Then $f_T(S_0;\gamma)$ is Lipschitz for all t with a Lipschitz constant L independent of γ, and the difference
$$\left|\prod_{s=1}^{T} p\left(y_t|y^{t-1};\gamma\right) - \prod_{s=1}^{T} \widetilde{p}_j\left(y_t|y^{t-1};\gamma\right)\right| = \frac{L\delta}{1-\alpha}, \tag{33}$$
holds for all γ.

2. Assumption 14 makes the bounds on $p\left(y_t|y^{t-1}, W_1^t, S_0;\gamma\right)$ and $p_j\left(y_t|y^{t-1}, W_1^t, S_0;\gamma\right)$ independent of γ. Therefore, the bound B and all the expressions where it appears are independent of γ.

■

References

[1] Aguirregabiria, V. and P. Mira (2004). "Sequential Estimation of Dynamic Discrete Games". *Mimeo*, Boston University.

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

[3] Collard, F. and M. Juillard (2001). "Perturbation Methods for Rational Expectations Models". *Mimeo*, CEPREMAP.

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

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

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

[7] Dib, A. (2001). "An Estimated Canadian DSGE Model with Nominal and Real Rigidities". *Bank of Canada Working Paper* 2001-26.

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

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

[10] Fernández-Villaverde, J. and J. Rubio-Ramírez (2003). "Comparing Dynamic Equilibrium Models to Data: A Bayesian Approach". *Journal of Econometrics,* forthcoming.

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

[12] Fernández-Villaverde, J. and J. Rubio-Ramírez (2004b). "Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood". *Journal of Applied Econometrics,* forthcoming.

[13] Fernández-Villaverde, J. and J. Rubio-Ramírez (2004c). "Some Results on the Solution of the Neoclassical Growth Model". *Penn Institute for Economic Research Working Paper 04-002.* University of Pennsylvania.

[14] Ferrall, C. (1997). "Unemployment Insurance Eligibility and the Transition from School to Work in Canada and the United States". *Journal of Business and Economic Statistics* 15, 115-129.

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

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

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

[20] Ireland, P. (2002). "Technology Shocks in the New Keynesian Model". *Mimeo,* Boston College.

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

[22] Judd, K.L. (1998). *Numerical Methods in Economics*. MIT Press.

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

[26] Lubik, T. and F. Schorfheide (2003). "Do Central Banks Respond to Exchange Rates? A Structural Investigation". *Mimeo*, University of Pennsylvania.

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

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

[29] McGrattan, E.R. (1999). "Application of Weighted Residuals Methods to Dynamic Economic Models" in R. Marimón and A. Scott (eds.) *Computational Methods for the Study of Dynamic Economies*. Oxford University Press.

[30] McGrattan, E., R. Rogerson, and R. Wright (1997). "An Equilibrium Model of the Business Cycle with Household Production and Fiscal Policy". *International Economic Review* 33, 573-601.

[31] Miller, R.A. (1984). "Job Matching and Occupational Choice". *Journal of Political Economy* 92, 1086-1120.

[32] Otrok, C. (2001). "On Measuring the Welfare Cost of Business Cycles". *Journal of Monetary Economics* 47, 61-92.

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

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

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

[36] Rabanal, P. and J. Rubio-Ramírez (2003). "Comparing New Keynesian Models of the Business Cycle: a Bayesian Approach". *Federal Reserve Bank of Atlanta Working Paper 2003-30.*

[37] 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.

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

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

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

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

[42] Santos, M.S. (2003). "Simulation-Based Estimation of Dynamic Models with Continuous Equilibrium Solutions". *Mimeo*, Arizona State University.

[43] Santos, M.S. and A. Peralta-Alva (2003). "Accuracy of Simulations for Stochastic Dynamic Models". *Mimeo*, Arizona State University.

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

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

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

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

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

[49] Smets, F. and R. Wouters (2003). "Shocks and Frictions in US Business Cycle Fluctuations: A Bayesian DSGE Approach". *Mimeo,* European Central Bank.

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

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

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

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

Figure 6.1.1: Absolute Value Difference between the Likelihoods as a Function of δ

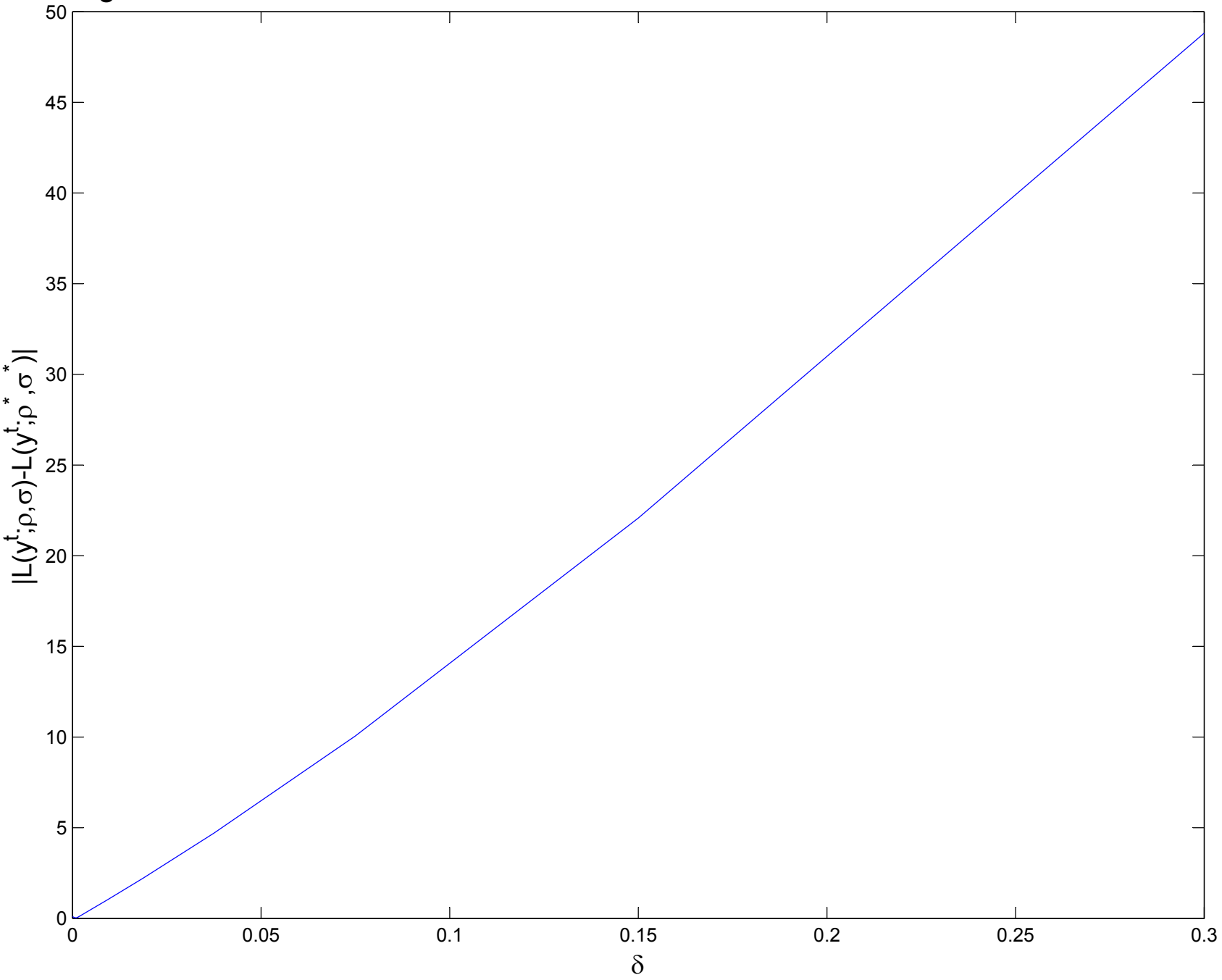

Figure 6.1.2: Absolute Value Difference between the Likelihoods as a Function of the Sample Size

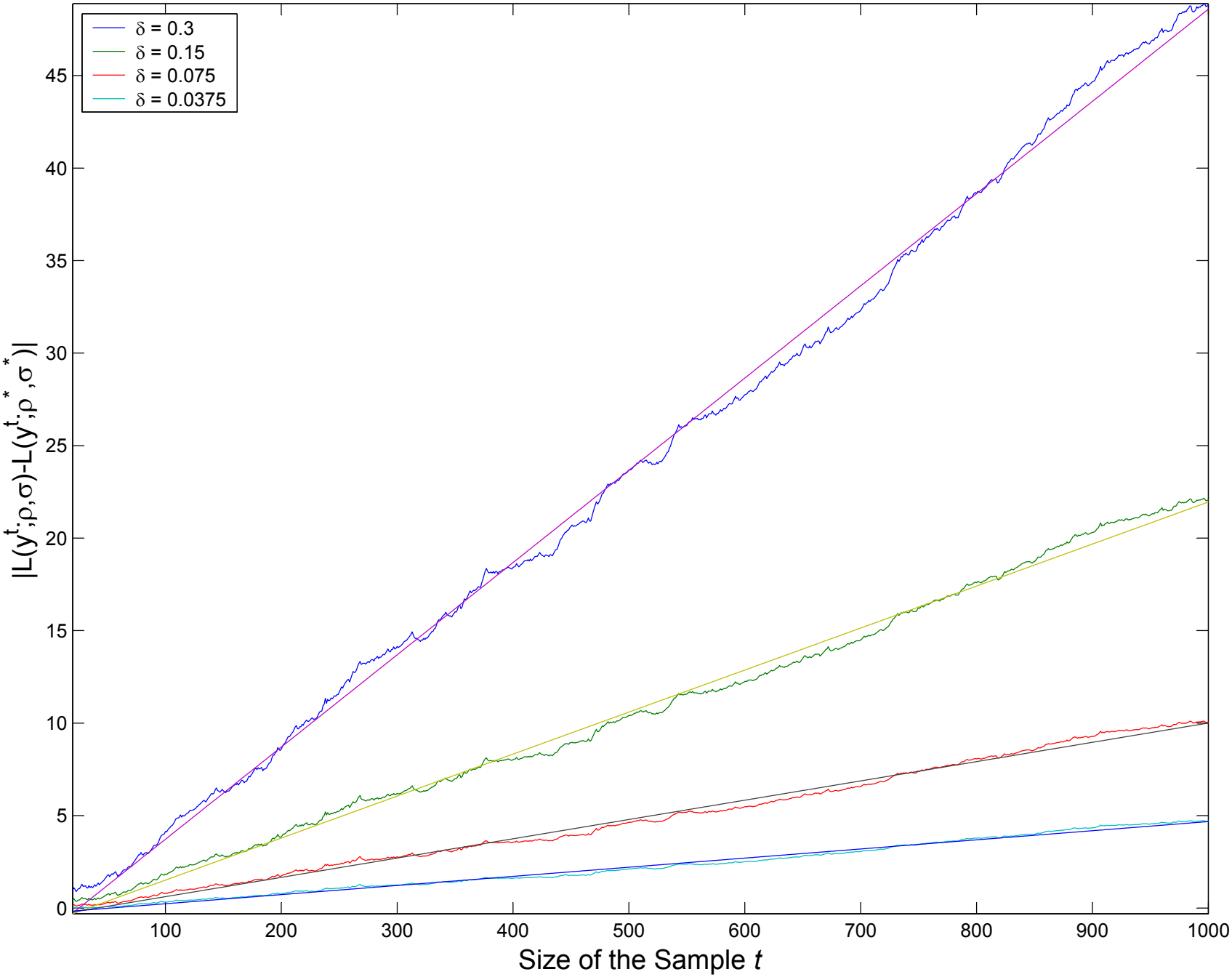

Figure 6.2.1: Absolute Value Difference between the Likelihoods as a function of δ

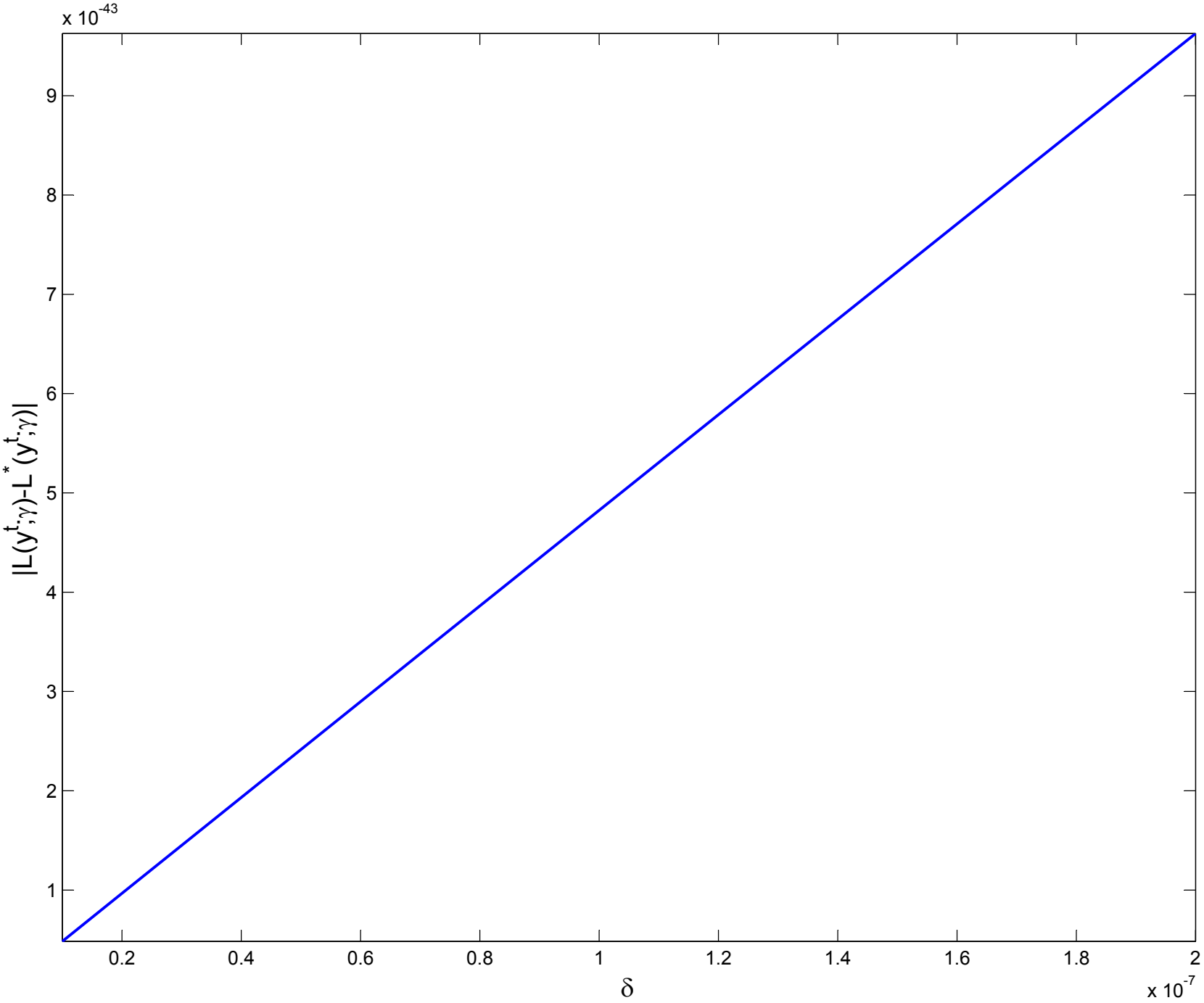

Figure 6.2.2: Absolute Value Difference between the Likelihoods as a Function of the Sample Size

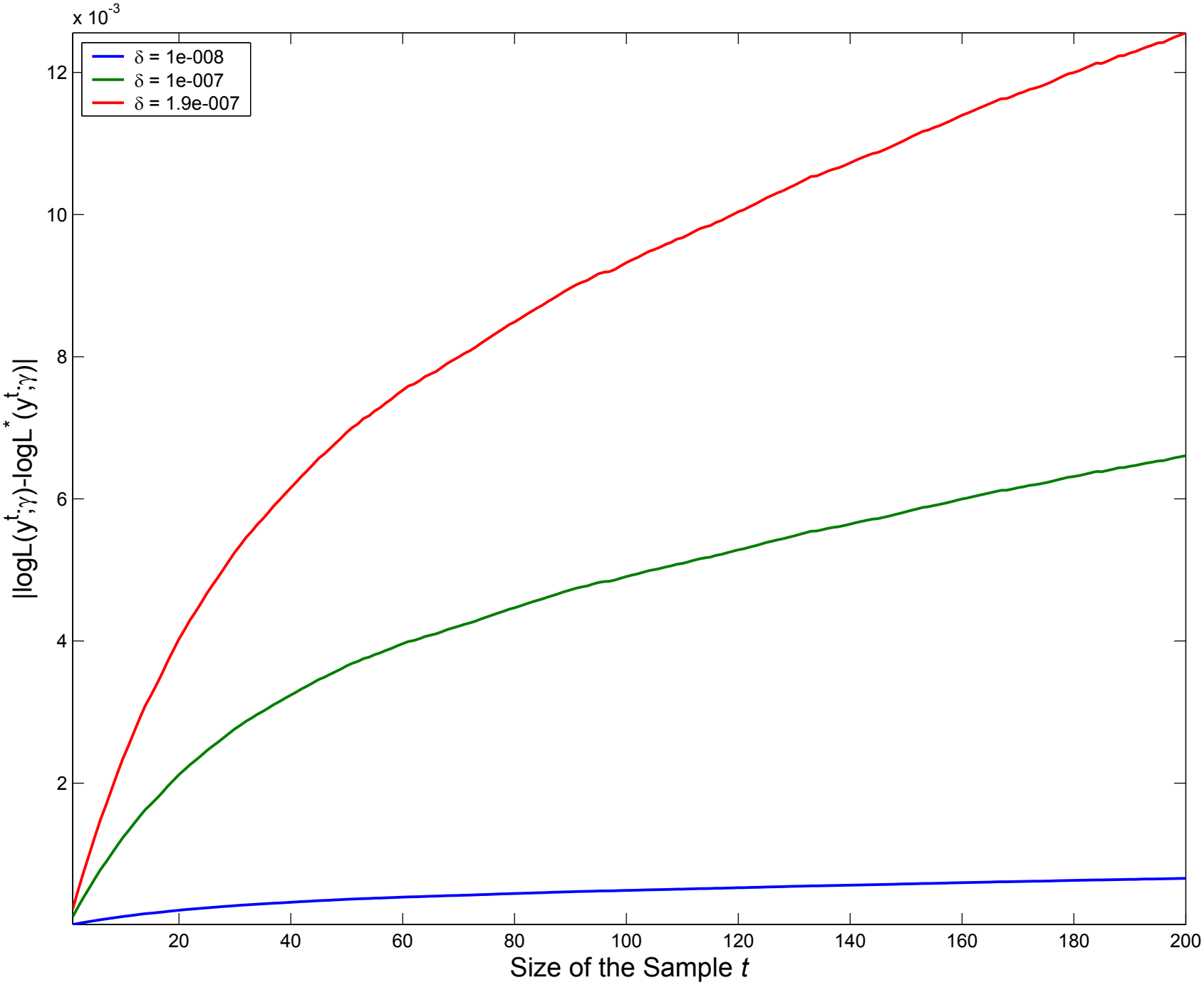

Figure 6.3.1: Absolute Value Difference between the Likelihoods as a Function of the Sample Size