Econometrica, Vol. 77, No. 6 (November, 2009), 2009–2017

COMMENTS ON "CONVERGENCE PROPERTIES OF THE LIKELIHOOD OF COMPUTED DYNAMIC MODELS"

BY DANIEL ACKERBERG, JOHN GEWEKE, AND JINYONG HAHN

We show by counterexample that Proposition 2 in Fernández-Villaverde, Rubio-Ramírez, and Santos (*Econometrica* (2006), 74, 93–119) is false. We also show that even if their Proposition 2 were corrected, it would be irrelevant for parameter estimates. As a more constructive contribution, we consider the effects of approximation error on parameter estimation, and conclude that second order approximation errors in the policy function have at most second order effects on parameter estimates.

KEYWORDS: Approximation error.

1. INTRODUCTION

FERNÁNDEZ-VILLAVERDE, RUBIO-RAMÍREZ, AND SANTOS (2006; FRS hereafter) consider likelihood based estimation of economic models which cannot be solved analytically and must be solved numerically or with some other form of approximation. This approximated model (e.g., an optimal policy function in the dynamic environment of FRS) is used in place of the exact model to form an approximated likelihood function. The approximated likelihood is then used for statistical inference.

FRS posed the question, "What are the effects on statistical inference of using an approximated likelihood instead of the exact likelihood?" Assuming that the approximated model is converging to the exact model, they obtained two results. First, given a fixed number of observations T, they found conditions under which the approximated likelihood function converges to the exact likelihood function. This is an important contribution, since if the approximate likelihood function converges to something other than the exact likelihood function, it is hard to see how one could ever perform meaningful inference.

In the second part of the paper (Section 4), the authors consider implications of the relative *speed* at which the approximated model (in their case, the policy function) converges to the exact model. Their main conclusion here is that "*second order approximation errors in the policy function*, which almost always are ignored by researchers, have *first order effects on the likelihood function*" (emphasis in the introduction of original text). More specifically, Proposition 2 in their paper states that the difference between the approximated likelihood function and the exact likelihood function is bounded by a function that includes a term of the form $TB\chi\delta$, where δ is a bound on the approximation error of the model, and where B and χ are constants that do not depend on T. This result implies that as T increases, one may need the approximation error δ to be shrinking at a rate faster than T for the approximated likelihood to converge to the exact likelihood. Given the authors' emphasis on the result, a reader might additionally conclude that, for meaningful inference on the para-

 DOI: 10.3982/ECTA7669

meter vector γ_0, one may also need the approximation error δ to disappear at a rate faster than T.

This comment has two parts. First, we show by counterexample that Proposition 2 in FRS is false. In a simple model that satisfies the assumptions of the FRS framework, we find a *sharp* bound on the likelihood function that can either be larger or smaller (depending on parameters) than the bound claimed by Proposition 2. We note that Geweke (2007) was the first to point out the error in Proposition 2 in FRS. He showed that the upper bound derived there is incorrect. He also pointed out the logical error of using an upper bound that is not sharp to make conclusions regarding the effects of approximation error. Our note goes one step further, using the counterexample to show that even if their Proposition 2 were corrected, it would be irrelevant for parameter inference.

Second, as a more constructive contribution, we extend the results of FRS to explicitly consider the effects of approximation error on parameter inference from a classical perspective.[1] What is relevant for classical maximum likelihood inference on γ_0 is not the behavior of the approximated likelihood function per se (as considered by FRS), but the behavior of the *maximizer* of the approximated likelihood function. Denote this maximizer as $\widehat{\gamma}_j$ and call it the pseudo-maximum likelihood estimator (PMLE). Our first result shows that as T increases, the difference between $\widehat{\gamma}_j$ and the true parameter vector γ_0 converges to something bounded by a term of the *same order* as the approximation error. Hence, we conclude that *second order approximation errors in the policy function* have at most *second order effects on parameter inference*.

We then investigate the consistency and asymptotic normality of the PMLE. The analysis is a straightforward extension of that used in the simulated maximum likelihood literature (e.g., Gouriéroux and Monfort (1991), Hajivassiliou and Ruud (1994)). We first show that as long as the approximation error converges to zero at *any* rate (as T increases), $\widehat{\gamma}_j$ is a consistent estimate of γ_0.[2] Regarding asymptotic normality, we show that as long as the approximation error disappears at a rate *faster* than $\sqrt{T}$, the approximation error does not affect the asymptotic distribution of the maximum likelihood estimator. In other words, the asymptotic distribution of $\sqrt{T}(\widehat{\gamma}_j - \gamma_0)$ is normal with mean 0 and variance given by the inverse of the information matrix of the exact model. Under our assumptions, this information matrix can be consistently estimated in the standard way.

[1]Note that our extension is done under some additional regularity conditions that are not assumed by FRS.

[2]A working paper version of Fernández-Villaverde, Rubio-Ramírez, and Santos (2005) shows that if the approximated likelihood function converges to the exact likelihood function, the maximum likelihood estimate using the approximated likelihood converges to the value of the parameter vector that maximizes the exact likelihood function. However, this is done for fixed T and thus does not imply anything about consistency or asymptotic normality.

We conclude that the relative impact of approximation error is not as large as one might conclude from a reading of FRS. The effects of approximation error on classical inference regarding γ_0 are of the same magnitude as the approximation error itself. There is no sense in which the effects of approximation error on point estimates worsen as the sample size T increases. To put the result in context, it is helpful to compare the effects of approximation error to the effects of simulation error in maximum likelihood estimation. In both cases, there are "bias" terms that must disappear at a rate faster than $\sqrt{T}$ to not affect the asymptotic distribution of the estimator. However, we emphasize that these are only asymptotic results. One's estimates are only going to be as accurate as one's approximations, and we believe one should make concerted efforts to make these approximations as precise as possible. We also note that it is beyond the scope of this comment to characterize the convergence behavior of approximation error using various approximation techniques.[3]

2. SIMPLIFIED VERSION OF FRS'S PROPOSITION 2

We consider a very simplified, static version of the model studied by FRS. This simple model is sufficient to illustrate the key points of this comment. Note that since our simplified model is just a restricted version of the model of FRS,[4] our counterexample to FRS's Proposition 2 is in fact a valid counterexample.

Our simple model is

$$y_t = g(v_t; \gamma),$$

where v_t is an independent and identically distributed unobservable, and γ is an unknown parameter which will be assumed to be a scalar for simplicity of notation. As in FRS, we will assume that we cannot compute the function $g(\cdot; \gamma)$ exactly. Alternatively, it can be approximated by a function $g_j(\cdot; \gamma)$, where j indexes the accuracy of the approximation. This approximated model $g_j(\cdot; \gamma)$ will in turn generate an approximation $p_j(\cdot; \gamma)$ to the true density $p(\cdot; \gamma)$ of y_t.

Lemma 6 of FRS then implies that, under some regularity conditions, if $\|g_j(\cdot; \gamma) - g(\cdot; \gamma)\| \leq \delta$, where $\|\cdot\|$ is the sup norm, then there exists some constant $\chi > 0$ such that

$$|p_j(y_t; \gamma) - p(y_t; \gamma)| \leq \chi\delta. \tag{1}$$

[3]For example, quadrature and interpolation.

[4]Formally, one can get to our simplified model from the FRS framework by simply adding the additional assumptions that $S_t = \{\cdot\}$ and $W_t = \{\cdot\}$.

Their Proposition 2 further implies that

$$\left|\prod_{t=1}^{T} p_j(y_t;\gamma) - \prod_{t=1}^{T} p(y_t;\gamma)\right| \le TB\chi\delta, \tag{2}$$

where, like χ, B is a constant that does not depend on T.[5]

3. COUNTEREXAMPLE TO PROPOSITION 2

We now show by counterexample that Proposition 2 of FRS is incorrect. Consider the model $y_t = g(v_t;\gamma) = \gamma + v_t$, where $v_t \sim N(0,\sigma^2)$. Suppose that σ^2 is known, so the only parameter to estimate is γ. Suppose that instead of using the true model $g(v_t;\gamma) = \gamma + v_t$, the econometrician uses an approximated model $g_j(v_t;\gamma) = \gamma + \delta + v_t$. For now, consider the approximation error $\delta \neq 0$ to be a constant. Again, note that this simple model satisfies the assumptions of the FRS framework.

The exact and approximated likelihoods for y_t are, respectively,

$$p(y_t;\gamma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_t-\gamma)^2}{2\sigma^2}\right)$$

and

$$p_j(y_t;\gamma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_t-\delta-\gamma)^2}{2\sigma^2}\right).$$

It can be shown[6] that the difference in these two individual likelihoods can be bounded by

$$|p(y_t;\gamma) - p_j(y_t;\gamma)| \le \chi(\delta)|\delta|, \tag{3}$$

where $\chi(\delta)$ is such that

$$\chi(\delta) \le \frac{\exp\left(-\frac{1}{2}\right)}{\sigma\sqrt{2\pi}} \tag{4}$$

and

$$\frac{1}{\sigma\sqrt{2\pi}} \le \liminf_{|c|\to\infty} |c|\chi(c) \le \limsup_{|c|\to\infty} |c|\chi(c) \le \frac{2}{\sigma\sqrt{2\pi}}. \tag{5}$$

[5]The second term in the published version of Proposition 2 of FRS disappears in our simple example since $S_t = \{\cdot\}$ and $W_t = \{\cdot\}$.

[6]Proofs of (3), (4), and (5) are available in the a Supplemental Material (Ackerberg, Geweke, and Hahn (2009)).

It is also shown that the difference in the joint likelihoods can be bounded by

$$\left|\prod_{t=1}^{T} p_j(y_t;\gamma) - \prod_{t=1}^{T} p(y_t;\gamma)\right| \le \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{T-1} |\sqrt{T}\delta|\chi(\sqrt{T}\delta) \tag{6}$$

and that this bound is sharp.

Consider the case where $\sigma < 1/\sqrt{2\pi}$. Because $|\sqrt{T}\delta|\chi(\sqrt{T}\delta)$ is bounded from above and away from zero for T large, our bound is of order K^{T-1}, where $K > 1$. Comparing our bound to the bound derived in Proposition 2 of FRS (i.e., (2)), it is obvious that for large enough T, our bound will be strictly larger. Given that our bound is sharp, this contradicts Proposition 2 of FRS. If $\sigma > 1/\sqrt{2\pi}$, then our bound is of order K^{T-1} with $K < 1$, which implies that in other cases the bound in Proposition 2 of FRS is too big.

Given this counterexample, one might be inclined to try to find a valid bound for the joint likelihood of more general models. However, further consideration of this simple counterexample suggests that this might not be a fruitful endeavor. Note that the maximum likelihood error (MLE) of γ in the simple counterexample is $\widehat{\gamma} = \overline{y}$. The MLE using the approximate model is $\widehat{\gamma}_j = \overline{y} - \delta$. Hence, the effects of approximation error on inference regarding γ is of the same order of magnitude as the approximation error in the model $g_j(v_t;\gamma)$. In addition, the impact of a fixed level of approximation error does not depend on T. These results are true regardless of the value of the standard deviation σ.

In contrast, the dependence of our sharp bound (6) on T depends dramatically on the value of σ. When $\sigma < 1/\sqrt{2\pi}$, the bound increases exponentially in T (for fixed δ). When $\sigma > 1/\sqrt{2\pi}$, the bound actually shrinks. Thus, while the effects of approximation error on $\widehat{\gamma}_j$ do not depend on σ, the effect of approximation error on the bound of the joint likelihood depends critically on σ. The counterexample is therefore suggestive that the way in which these joint likelihood function bounds depend on T may not be relevant for studying the effects of approximation error on inference, even in more general models. In this simple example, the Bayesian posterior means (with flat priors) are equivalent to the MLEs. Hence, this evidence is also suggestive that the dependence of a joint likelihood bound on T is not relevant for both classical inference and at least some aspects of Bayesian inference.

In the next two sections, we provide a constructive analysis of the effects of approximation error on classical maximum likelihood inference. This analysis is based on the differences in the true and the approximating average log likelihoods rather than the differences in their levels.

4. THE EFFECTS OF APPROXIMATION ERROR ON CLASSICAL INFERENCE

We now extend the results in FRS to examine the effects of approximation error on classical inference regarding γ. This is done in the context of the static model of the previous section, but we suspect that our results could also be shown in FRS's more general framework. We first show that we can bound the effect of approximation error on parameter inference by a term of the *same* order of magnitude as the approximation error. In other words, we show that second order approximation errors in the model have at most second order effects on inference regarding the model's parameters. In the following section, we explicitly analyze the effect of approximation error on the asymptotic distribution of maximum likelihood estimators.

Denote the true value of γ by γ_0. Define $\widehat{\gamma}_j$ as the pseudo-maximum likelihood estimator (PMLE) which maximizes the approximated joint log likelihood function, that is,

$$\widehat{\gamma}_j = \arg\max_{\gamma} \frac{1}{T}\sum_{t=1}^{T} \log p_j(y_t; \gamma).$$

To characterize the magnitude of the effect of approximation error on estimation, we will investigate how the probability limit of $\widehat{\gamma}_j$ depends on the approximation error.

We use the Sobolev norm to measure the degree of approximation:

DEFINITION 1: We define

$$\Delta_j \equiv \max\left\{\sup_{y_t,\gamma}\left|\frac{\partial^k}{\partial\gamma^k}\log p_j(y_t;\gamma) - \frac{\partial^k}{\partial\gamma^k}\log p(y_t;\gamma)\right|, k = 0, 1, 2\right\}.$$

The Δ_j measures how well the individual log likelihood p_j approximates both the level and the shape of the exact log likelihood. This approximation error in the individual log likelihood is generated by the difference between the exact model $(g(v_t;\gamma))$ and the approximated model $(g_j(v_t;\gamma))$. One could derive this bound Δ_j from lower level assumptions on bounds relating $g_j(v_t;\gamma)$, $g(v_t;\gamma)$, and their derivatives.[7] For this comment, it is sufficient to observe that this approximation error bound Δ_j will not depend on the sample size T. This is because in our simple model, none of $g_j(v_t;\gamma)$, $g(v_t;\gamma)$, or the distribution of v_t depends on the sample size T. We will assume that the index j is such that the approximation error gets small as j gets larger:

CONDITION 1: *We assume* $\Delta_j \to 0$ *as* $j \to \infty$.

We will assume the following standard regularity conditions on the exact likelihood, which can be found in, for example, Newey and McFadden (1994; NM hereafter):

[7]Given that we bound differences in shapes as well as levels between the approximated and true individual log likelihoods, this might require extra regularity conditions (on $g_j(v_t;\gamma)$, $g(v_t;\gamma)$, and the distribution of v_t) in addition to those assumed by FRS.

CONDITION 2: (i) *If* $\gamma \neq \gamma_0$, *then* $p(y_t;\gamma) \neq p(y_t;\gamma_0)$; (ii) $\gamma_0 \in \Gamma$, *which is compact*; (iii) $\log p(y_t;\gamma)$ *is continuous at each* $\gamma \in \Gamma$ *with probability* 1; (iv) $E[\sup_{\gamma\in\Gamma} |\log p(y_t;\gamma)|] < \infty$.

Given these conditions, NM (Lemma 2.2) implies that the function $Q_0(\gamma) \equiv E[p(y_t;\gamma)]$ is uniquely maximized at γ_0. We will strengthen this identification result by making the following additional assumption:

CONDITION 3: *We assume* $Q_0''(\gamma_0) < 0$.

Last, we assume additional regularity conditions on both the exact individual likelihood $p(y_t;\gamma)$ (Condition 4) and the approximated individual likelihood $p_j(y_t;\gamma)$ (Condition 5):

CONDITION 4: (i) $\log p(y_t;\gamma)$ *is twice continuously differentiable*; (ii) *there exists some* $d(y_t)$ *with* $E[d(y_t)] < \infty$ *such that* $\gamma \in \Gamma$ $|\log p(y_t;\gamma)| \leq d(y_t)$, $|\nabla_\gamma \log p(y_t;\gamma)| \leq d(y_t)$, *and* $|\nabla_{\gamma\gamma} \log p(y_t;\gamma)| \leq d(y_t)$ *for all* $\gamma \in \Gamma$.

CONDITION 5: (i) $\log p_j(y_t;\gamma)$ *is twice continuously differentiable*; (ii) *there exists some* $d(y_t)$ *with* $E[d(y_t)] < \infty$ *such that* $|\log p_j(y_t;\gamma)| \leq d(y_t)$, $|\nabla_\gamma \log p_j(y_t;\gamma)| \leq d(y_t)$, *and* $|\nabla_{\gamma\gamma} \log p_j(y_t;\gamma)| \leq d(y_t)$ *for all* $\gamma \in \Gamma$; (iii) γ_j *uniquely maximizes* $Q_j(\gamma) \equiv E[p_j(y_t;\gamma)]$.

Note that part (iii) of Condition 5 is not guaranteed by NM (Lemma 2.2) because $p_j(y_t;\gamma)$ is not the true likelihood of the data y_t. However, given we assume this additional identification condition, NM (Theorem 2.1) implies that the PMLE $\widehat{\gamma}_j$ converges to γ_j in probability, i.e.

THEOREM 1: *Suppose that Condition 5 is satisfied. Then*, $\widehat{\gamma}_j$ *converges to* γ_j *in probability as* $T \to \infty$ *while* j *is fixed*.

The proofs of all theorems are given in the Supplemental Material (Ackerberg, Geweke, and Hahn (2009)).

We next explicitly relate how this bound relates to the approximation error of p_j:

THEOREM 2: *Suppose that Conditions 1–5 are satisfied. Then there exists* $\zeta > 0$ *such that* $|\gamma_j - \gamma_0| \leq \zeta \cdot \Delta_j$.

The ζ in Theorem 2 does not depend on the sample size T. Hence, Theorem 2 states that the difference between the true parameter and the probability limit of the PMLE using the approximated likelihood is bounded by a term of the *same* magnitude as the approximation error Δ_j. As a result, we conclude that second order approximation errors in the model have at most second order effects on inference regarding parameters.

5. ASYMPTOTIC RESULTS WITH APPROXIMATION ERROR

Last we explicitly consider the effects of approximation errors on standard asymptotic approximations. For this purpose, we assume that the index j is a function $j(T)$ of the sample size T such that $j(T) \to \infty$ as $T \to \infty$. It is fairly obvious that one will need the approximation error to disappear asymptotically in order to obtain consistent estimates. What might be less obvious is the rate at which the approximation error needs to disappear, both for consistency and for standard asymptotic approximations to be valid. The results in this section are based on very standard arguments that have also been used in the simulated maximum likelihood literature (Gouriéroux and Monfort (1991), Hajivassiliou and Ruud (1994)); the proofs are in the Supplemental Material.

Our first result regards the consistency of the PMLE $\widehat{\gamma}_{j(T)}$.

THEOREM 3: *Suppose that Conditions 1–5 are satisfied. Then* $\widehat{\gamma}_{j(T)} = \gamma_0 + o_p(1)$ *as* $T \to \infty$ *and* $\Delta_{j(T)} \to 0$.

This result states that $\widehat{\gamma}_{j(T)}$ is a consistent estimate of γ_0 regardless of the rate at which $\Delta_{j(T)}$ converges to zero (as $T \to \infty$). Intuitively, the approximated likelihood converges to the exact likelihood (since $\Delta_{j(T)} \to 0$), and the maximum of the exact likelihood converges to γ_0 (since $T \to \infty$), providing the result.

Our next result considers the asymptotic distribution of $\sqrt{T}(\widehat{\gamma}_{j(T)} - \gamma_0)$ as $T \to \infty$ and $\Delta_{j(T)} \to 0$.

THEOREM 4: *Suppose that Conditions 1–5 are satisfied. Then*

$$\sqrt{T}\big(\widehat{\gamma}_{j(T)} - \gamma_0\big) \Rightarrow N(0, \mathcal{I}^{-1})$$

if $\sqrt{T}\Delta_{j(T)} \to 0$ *as* $T \to \infty$. *Here,* $\mathcal{I} = -Q_0''(\gamma_0)$ *denotes the Fisher information.*

To appreciate Theorem 4, note that standard arguments imply that the MLE that maximizes the exact joint likelihood, that is,

$$\widehat{\gamma}_0 = \arg\max_{\gamma} \frac{1}{T}\sum_{t=1}^{T} \log p(y_t; \gamma),$$

also has asymptotic distribution

$$\sqrt{T}(\widehat{\gamma}_0 - \gamma_0) \Rightarrow N(0, \mathcal{I}^{-1}). \tag{7}$$

A comparison of (7) to Theorem 4 implies that as long as $\Delta_{j(T)} \to 0$ at a rate *faster* than $\sqrt{T}$, approximation error *does not* affect the asymptotic distribution of the estimator. Intuitively, approximation error introduces a bias term of order $\Delta_{j(T)}$ into the PMLE. If this bias term disappears at a rate slower than $\sqrt{T}$, this bias term dominates the asymptotic distribution. If the term disappears faster than $\sqrt{T}$, it vanishes from the asymptotic distribution. Again, this is very reminiscent of the "bias term" that arises in simulated maximum likelihood estimation. That bias term (which is inversely proportional to the number of simulation draws) also needs to disappear at a rate faster than $\sqrt{T}$ for it not to affect the asymptotic distribution.

Last, we note that under our assumptions it is straightforward to show that $Q_j''(\widehat{\gamma}_{j(T)})$ is a consistent estimate of $Q_0''(\gamma_0)$. In other words, we can consistently estimate $\mathcal{I}$ using the approximated model. Thus, our results imply that as long as approximation error disappears at a rate faster than $\sqrt{T}$, it can be ignored for purposes of forming asymptotically valid confidence intervals and hypothesis tests.

REFERENCES

ACKERBERG, D., J. GEWEKE, AND J. HAHN (2009): "Supplement to 'Comments on 'Convergence Properties of the Likelihood of Computed Dynamic Models'," *Econometrica Supplemental Material*, 77, http://www.econometricsociety.org/ecta/Supmat/7669_proofs.pdf. [2012,2015]

FERNÁNDEZ-VILLAVERDE, J., J. F. RUBIO-RAMÍREZ, AND J. S. SANTOS (2005): "Convergence Properties of the Likelihood of Computed Dynamic Models," Technical Working Paper 315, NBER. [2010]

——— (2006): "Convergence Properties of the Likelihood of Computed Dynamic Models," *Econometrica*, 74, 93–119. [2009]

GEWEKE, J. (2007): "Convergence Properties of the Likelihood of Computed Dynamic Models: Comment," Unpublished Manuscript, University of Iowa, available at http://www.biz.uiowa.edu/faculty/jgeweke/papers/discussionA/comment.pdf. [2010]

GOURIÉROUX, C., AND A. MONFORT (1991): "Simulation Based Inference in Models With Heterogeneity," *Annales d'Économie et de Statistique*, 20/21, 69–107. [2010,2016]

HAJIVASSILIOU, V. A., AND P. A. RUUD (1994): "Classical Estimation Methods for LDV Models Using Simulation," in *Handbook of Econometrics*, Vol. IV, ed. by R. F. Engle and D. L. McFadden. Amsterdam: Elsevier. [2010,2016]

NEWEY, W. K., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Vol. IV, ed. by R. F. Engle and D. L. McFadden. Amsterdam: Elsevier. [2014]

Dept. of Economics, University of California Los Angeles, 8283 Bunche Hall, Los Angeles, CA 90095, U.S.A.; ackerberg@econ.ucla.edu,

University of Technology Sydney, City Campus, P.O. Box 123, Broadway, Sydney, NSW 2007, Australia and University of Colorado, Boulder, CO 80309, U.S.A.; john.geweke@uts.edu.au,

and

Dept. of Economics, University of California Los Angeles, 8283 Bunche Hall, Los Angeles, CA 90095, U.S.A.; hahn@econ.ucla.edu.

Manuscript received January, 2008; final revision received July, 2008.