Econometric Theory, **15**, 1999, 337–360. Printed in the United States of America.

STATISTICAL INFERENCE WITH SIMULATED LIKELIHOOD FUNCTIONS

LUNG-FEI LEE
Hong Kong University of Science and Technology

This paper considers classical test statistics, namely, the likelihood ratio, efficient score, and Wald statistics, for econometric models under simulation estimation. The simulated likelihood ratio, simulated efficient score, and simulated Wald test statistics are shown to be asymptotically equivalent. Because the simulated score vector can be asymptotically biased, limiting distributions of these simulated statistics can be asymptotically noncentral χ^2 distributed. This paper studies inference issues with various simulated test statistics. Monte Carlo results are also provided to compare and demonstrate finite sample properties of simulated test statistics.

1. INTRODUCTION

We consider classical inference statistics for models of which likelihood functions are simulated. Simulation estimation methods have been considered by, among others, Lerman and Manski (1981), McFadden (1989), Pakes and Pollard (1989), Hajivassiliou and McFadden (1990), Hendry and Richard (1992), and Laroque and Salanie (1993, 1994). Surveys on simulation estimation include those by Gourieroux and Monfort (1993), Hajivassiliou (1993), and Keane (1993). Most existing publications focus on simulators and estimation issues. In this paper, we focus on classical inference procedures such as the likelihood ratio (LR) test, the Wald (W) test, and the score (SC) test (or the Lagrange multiplier (LM) test) with simulated maximum likelihood estimators (SMLE's). These classical inference procedures have broad applications in econometrics (see, e.g., Engle, 1984; Godfrey, 1988). We investigate asymptotic distributions of various simulated classical statistics. It is shown that the limiting distributions of these simulated statistics are asymptotically χ^2 distributed; but under certain circumstances, they are asymptotically noncentral χ^2 distributed. We investigate adjusted statistics that may correct for possible asymptotic biases and incorporate various variance estimates. Monte Carlo experiments are performed to investigate finite sample properties of test statistics.

I thank the co-editor Professor Joel Horowitz and anonymous referees for helpful comments. This research has been supported by a DAG research grant from the Hong Kong University of Science and Technology. Address correspondence to: Lung-fei Lee, Department of Economics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong.

© 1999 Cambridge University Press 0266-4666/99 $9.50

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

A by-product of the Monte Carlo experiments is the investigation of the identification of specific dynamic discrete choice structures in a general model framework. Heckman (1981) discussed specific dynamic discrete choice models and their characterizations in terms of probability run patterns. He introduced a general dynamic model that contains many familiar dynamic structures, such as the Markov, Polya, renewal, and habit persistence state dependence. Under the normality assumption, the choice probabilities can be simulated by the GHK simulator (Geweke, 1991; Borsch-Supan and Hajivassiliou, 1993; Keane, 1994). With the GHK simulator, these dynamics models can be effectively estimated by the method of simulated likelihood provided the time dimension of a model is not too long (not more than 30 periods in a panel setting; see Lee, 1997). It remains of interest to investigate whether simulated test statistics can distinguish a specific dynamic structure from other dynamic structures in the general model.

This paper is organized as follows. Section 2 reviews the asymptotic properties of the simulated score vector and the SMLE. In Section 3, the asymptotic distributions for the simulated classical test statistics are derived under both the null hypothesis and a sequence of local alternatives. The simulated classical test statistics are shown to be asymptotically equivalent. Under the null hypothesis, the limiting distribution is a χ^2, but, under some circumstances, it can be a noncentral χ^2. For the latter case, possible bias-adjusted test statistics are considered in Section 4. In Section 5, simulated test statistics that incorporate various variance estimates of parameters are formulated. Monte Carlo results on the finite sample performance of these various simulated classical statistics are reported in Section 6. Conclusions are provided in Section 7. Proofs of all the asymptotic properties of simulated estimators and test statistics are presented in Appendixes A and B.

2. SIMULATED LIKELIHOOD AND ASYMPTOTIC DISTRIBUTIONS

Suppose that an econometric model is specified with a likelihood function $f(\theta, z)$ where θ is a k-dimensional parameter vector and z is a vector of random variables.[1] For the sample observation z_i, an unbiased simulator for $f(\theta, z_i)$ is $f_m(\theta, z_i) = 1/m \sum_{j=1}^{m} h(v_i^{(j)}, z_i, \theta)$, where $v_i^{(1)}, \ldots, v_i^{(m)}$ are m random draws for a random variable v and $E(h(v, z, \theta)|z) = f(\theta, z)$. The subscript i on v indicates that random draws are also independent across different sample observations.[2] For panel data with sample size n, the log-likelihood function is $\ln L_n(\theta) = \sum_{i=1}^{n} \ln f(\theta, z_i)$ and the simulated log likelihood function is

$$\ln L_{n,m}(\theta) = \sum_{i=1}^{n} \ln f_m(\theta, z_i). \tag{2.1}$$

The asymptotic distribution of the SMLE $\hat{\theta}_u$ from the maximization of (2.1) can be derived from a Taylor expansion. For notational simplicity, denote $S_{n,m}(\theta) =$

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

$(1/\sqrt{n})(\partial \ln L_{n,m}(\theta)/\partial\theta)$ and $H_{n,m}(\theta) = 1/n \sum_{i=1}^{n}(\partial^2 \ln f_m(\theta, z_i)/\partial\theta\partial\theta')$. Let θ_0 denote the true parameter vector. One has

$$\sqrt{n}(\hat{\theta}_u - \theta_0) = -[H_{n,m}(\bar{\theta})]^{-1} S_{n,m}(\theta_0), \tag{2.2}$$

where $\bar{\theta}$ lies between $\hat{\theta}_u$ and θ_0. By an expansion of a difference of ratios, the following proposition generalizes a result in Lee (1995).[3] The E_θ refers to the expectation operator with respect to $f(\theta, z)$ (and/or v) at a possible θ. The E without any parametric subscript refers to the expectation operator with respect to the true data generating process. Regularity conditions for our analysis and results are listed in Appendix A.

PROPOSITION 1. *Under the regularity assumptions* (*in Appendix* A),

$$S_{n,m}(\theta) = S_n(\theta) + P_{n,m}(\theta) + Q_{n,m}(\theta) + R_{n,m}(\theta), \tag{2.3}$$

where

$$S_n(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \ln f(\theta, z_i)}{\partial\theta} \left(= \frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta)}{\partial\theta} \right),$$

$$P_{n,m}(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{f(\theta, z_i)} \left\{ \left(\frac{\partial f_m(\theta, z_i)}{\partial\theta} - \frac{\partial f(\theta, z_i)}{\partial\theta} \right) - \frac{\partial \ln f(\theta, z_i)}{\partial\theta} (f_m(\theta, z_i) - f(\theta, z_i)) \right\},$$

$$Q_{n,m}(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{f^2(\theta, z_i)} \left\{ -\left(\frac{\partial f_m(\theta, z_i)}{\partial\theta} - \frac{\partial f(\theta, z_i)}{\partial\theta} \right)(f_m(\theta, z_i) - f(\theta, z_i)) + \frac{\partial \ln f(\theta, z_i)}{\partial\theta} (f_m(\theta, z_i) - f(\theta, z_i))^2 \right\},$$

and

$$R_{n,m}(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{f_m(\theta, z_i) f^2(\theta, z_i)} \times \left\{ \left(\frac{\partial f_m(\theta, z_i)}{\partial\theta} - \frac{\partial f(\theta, z_i)}{\partial\theta} \right)(f_m(\theta, z_i) - f(\theta, z_i))^2 + \frac{\partial \ln f(\theta, z_i)}{\partial\theta} (f_m(\theta, z_i) - f(\theta, z_i))^3 \right\}.$$

For any θ, $E_\theta(S_n(\theta)) = 0$ *and* $E(P_{n,m}(\theta)) = 0$ *but* $E(Q_{n,m}(\theta)) = n^{1/2} m^{-1} \mu(\theta)$, *where*

$$\mu(\theta) = E\left\{ \frac{1}{f^2(\theta, z)} \left[\frac{\partial \ln f(\theta, z)}{\partial\theta} E(h^2(v, z, \theta)|z) - E\left(h(v, z, \theta) \frac{\partial h(v, z, \theta)}{\partial\theta} \middle| z \right) \right] \right\}. \tag{2.4}$$

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

Furthermore, $S_n(\theta_0) = O_P(1)$; *and* $P_{n,m}(\theta) = O_P(m^{-1/2})$, $Q_{n,m}(\theta) - E(Q_{n,m}(\theta)) = O_P(m^{-1})$, *and* $R_{n,m}(\theta) = O_P(n^{1/2}m^{-3/2})$ *semiuniformly in* $\theta \in \Theta$.[4]

The proof of this proposition can be found in Appendix A. This proposition shows that if m increases at a rate faster than $n^{1/2}$, the limiting distribution of $\sqrt{n}(\hat{\theta}_u - \theta_0)$ is properly centered at zero; otherwise, it may not be so. The latter can happen because the second order term $Q_{n,m}(\theta)$ has the stochastic order $O_P(n^{1/2}m^{-1})$, which can be larger than the order $O_P(m^{-1/2})$ of $P_{n,m}(\theta)$ if m does not increase fast enough relative to n. The semiuniform convergence in probability is useful for the analysis when data generating processes were determined by a sequence of local alternatives.

COROLLARY 1. *Under the regularity assumptions, if* $\lim_{n\to\infty} n^{1/2}m^{-1} = c$ *a finite constant*,

$$S_{n,m}(\theta_0) \xrightarrow{D} N(c\mu, \mathcal{I}) \tag{2.5}$$

and

$$\sqrt{n}(\hat{\theta}_u - \theta_0) \xrightarrow{D} N(c\mathcal{I}^{-1}\mu, \mathcal{I}^{-1}), \tag{2.6}$$

where $\mathcal{I}$ *is the information matrix of the model and* $\mu = \mu(\theta_0)$ *for simplicity*.

The asymptotic distributions in (2.5) and (2.6) might be relevant distributions for consideration. In the following sections, we investigate possible consequences of (2.5) and (2.6) for various classical test statistics constructed with simulated likelihood functions.

3. SIMULATED TEST STATISTICS

Three classical test statistics for the testing of nested hypothesis are the LR, W, and SC statistics. Because the likelihood function of the model can only be simulated, corresponding statistics constructed with the exact likelihood function replaced by a simulated likelihood function are the objects of our investigation.

Let $r(\theta) = 0$ be the set of ν constraints, where $\nu < k$, on θ. The constrained simulated maximum likelihood estimator (CSMLE) $\hat{\theta}_c$ is derived from $\max_\theta \ln L_{n,m}(\theta)$ subject to $r(\theta) = 0$. A simulated likelihood ratio (SLR) test statistic can be intuitively defined as LR^{sl} where

$$LR^{sl} = 2[\ln L_{n,m}(\hat{\theta}_u) - \ln L_{n,m}(\hat{\theta}_c)]. \tag{3.1}$$

A simulated Wald (SW) test statistic can be

$$W^{sl} = nr'(\hat{\theta}_u)[R(\hat{\theta}_u)\hat{\mathcal{I}}^{-1}R'(\hat{\theta}_u)]^{-1}r(\hat{\theta}_u), \tag{3.2}$$

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

where $R(\theta) = \partial r(\theta)/\partial\theta'$ and $\hat{\mathcal{I}}$ is a consistent estimator of $\mathcal{I}$ evaluated at $\hat{\theta}_u$. A corresponding simulated score (SSC) test statistic is

$$S^{sl} = \frac{1}{n} \frac{\partial \ln L_{n,m}(\hat{\theta}_c)}{\partial\theta'} \hat{\mathcal{I}}^{-1} \frac{\partial \ln L_{n,m}(\hat{\theta}_c)}{\partial\theta}, \tag{3.3}$$

where $\hat{\mathcal{I}}$ is evaluated at $\hat{\theta}_c$. The asymptotic distributions of these statistics will depend on the asymptotic distributions of $\hat{\theta}_u$, $\hat{\theta}_c$, and/or the simulated score vector $S_{n,m}(\theta_0)$ in (2.3).

PROPOSITION 2. *Under the assumed regularity conditions and the null hypothesis* $H_0: r(\theta_0) = 0$, *if* $\lim_{n\to\infty} n^{1/2}m^{-1} = c$ *a finite constant, the* SW, SSC, *and* SLR *test statistics* W^{sl}, S^{sl}, *and* LR^{sl} *are asymptotically* $\chi^2_\nu(\delta_0)$ *distributed with* ν *degrees of freedom and noncentrality parameter* δ_0 *where* $\delta_0 = c^2\mu' P\mu$ *with* $P = \mathcal{I}^{-1}R'(R\mathcal{I}^{-1}R')^{-1}R\mathcal{I}^{-1}$.

The detailed proofs of all our results appear in Section 3 of Appendix B. From Proposition 2, the limiting distribution of the simulated statistics can be asymptotically distributed as a regular central χ^2 only if $c = 0$, which requires m to increase at a rate faster than the square root of n. In practice, one has a sample of finite size and it might be realistic to regard $\sqrt{n}/m$ as a finite but positive c for approximation. The noncentrality parameter δ_0 in these test statistics occurs because of the asymptotic biases in $S_{n,m}(\theta_0)$, $\hat{\theta}_u$, and $\hat{\theta}_c$. It can be small if $\sqrt{n}/m$ is small.

It is known that the classical W, LR, and SC test statistics without simulation are asymptotically χ^2_ν distributed. They are asymptotically equivalent in the sense that the limiting distributions of these statistics are identical under both the null hypothesis and a sequence of local alternative hypotheses (Engle, 1984). A sequence of local alternative hypotheses refers usually to local Pitman-type alternatives of the form

$$H_n: \theta_n = \theta_0 + \Delta/\sqrt{n}, \tag{3.4}$$

where Δ is an arbitrary finite constant vector (Gallant and Holly, 1980). Under the sequence $\{H_n\}$, the data generating process is determined by θ_n. The θ_0 is the limit of the sequence $\{\theta_n\}$. The limit point θ_0 would satisfy the constraint $r(\theta_0) = 0$ but the sequence θ_n might not. The asymptotic distributions of the three simulated test statistics under the local alternatives are in the following proposition.

PROPOSITION 3. *Assume that the regularity conditions are satisfied. Under the sequence of local alternatives in* (3.4), *if* $\lim_{n\to\infty} n^{1/2}m^{-1} = c$ *a finite constant, the* SW, SSC, *and* SLR *test statistics* W^{sl}, S^{sl}, *and* LR^{sl} *are asymptotically* $\chi^2_\nu(\delta_1)$ *where* $\delta_1 = (\mathcal{I}\Delta + c\mu)'P(\mathcal{I}\Delta + c\mu)$.

Because the SW, SSC, and SLR statistics have the same limiting distributions under both the null and local alternative hypotheses, these statistics are asymp-

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

totically equivalent. The noncentrality parameter δ_1 is caused by two sources—the nonzero Δ in H_n and the asymptotic biases in $\hat{\theta}_u$, $\hat{\theta}_c$, and $S_{n,m}(\theta_0)$.[5]

4. BIAS-ADJUSTED SMLE'S AND TEST STATISTICS

A bias-adjusted SMLE (BASMLE) has been suggested in Lee (1995) by removing the asymptotic bias in $\hat{\theta}_u$. The $E(h^2(v,z_i,\theta_0)|z_i)$ in μ can be estimated by $s_m(\hat{\theta}_u,z_i)$ where $s_m(\theta,z_i) = 1/m\sum_{j=1}^m h^2(v_i^{(j)},z_i,\theta)$. The $E(h(v,z_i,\theta_0)(\partial h(v,z_i,\theta_0)/\partial\theta)|z_i)$ can be estimated by $c_m(\hat{\theta}_u,z_i)$ where $c_m(\theta,z_i) = 1/m\sum_{j=1}^m h(v_i^{(j)},z_i,\theta)(\partial h(v_i^{(j)},z_i,\theta)/\partial\theta)$. The μ can be estimated by

$$\mu_{n,m}(\hat{\theta}_u) = 1/n\sum_{i=1}^n \frac{1}{f_m^2(\hat{\theta}_u,z_i)}\left(\frac{\partial \ln f_m(\hat{\theta}_u,z_i)}{\partial\theta}s_m(\hat{\theta}_u,z_i) - c_m(\hat{\theta}_u,z_i)\right).$$

A BASMLE of θ_0 is $\hat{\theta}_A = \hat{\theta}_u + H_{n,m}^{-1}(\hat{\theta}_u)\mu_{n,m}(\hat{\theta}_u)/m$, or, alternatively,

$$\hat{\theta}_A = \hat{\theta}_u - \left\{\frac{1}{n}\sum_{i=1}^n \frac{\partial \ln f_{n,m}(\hat{\theta}_u,z_i)}{\partial\theta}\frac{\partial \ln f_{n,m}(\hat{\theta}_u,z_i)}{\partial\theta'}\right\}^{-1}\frac{\mu_{n,m}(\hat{\theta}_u)}{m}. \tag{4.1}$$

It has been shown in Lee (1995) that $\sqrt{n}(\hat{\theta}_A - \theta_0) = \mathcal{I}^{-1}\{S_n(\theta_0) + P_{n,m}(\theta_0) + (Q_{n,m}(\theta_0) + E(Q_{n,m}(\theta_0)))\} + o_P(1) = \mathcal{I}^{-1}S_n(\theta_0) + o_P(1)$. Thus the bias $E(Q_{n,m}(\theta_0))$ has been removed from $\hat{\theta}_u$ and $\sqrt{n}(\hat{\theta}_A - \theta_0) \xrightarrow{D} N(0,\mathcal{I}^{-1})$. The basic idea behind this adjustment uses the property that $S_{n,m}(\theta) - n^{1/2}\mu(\theta)/m = S_n(\theta) + P_{n,m}(\theta) + (Q_{n,m}(\theta) - E(Q_{n,m}(\theta))) + R_{n,m}(\theta) = S_n(\theta) + o_P(1)$. The $\hat{\theta}_A$ is an unconstrained BASMLE.

For constrained estimation, it is possible to construct a bias-adjusted estimator that satisfies parameter constraints. For that purpose, it is desirable to rewrite the ν constraints $r(\theta) = 0$ in its constraint formulation into the equivalent freedom equation formulation $\theta = w(\alpha)$ where w is a vector-valued function from $R^{k-\nu}$ to R^k and α represents a free ν-dimensional parameter vector. With the freedom equation formation, define the constrained simulated likelihood (CSL) in terms of α as $L_{n,m}^{(c)}(\alpha) = L_{n,m}(w(\alpha))$. The true parameter α_0 can be estimated by the maximization of $L_{n,m}^{(c)}(\alpha)$ with respect to α. Let $\hat{\alpha}$ be the SMLE of α_0. Then $\hat{\theta}_c = w(\hat{\alpha})$ is a constrained SMLE (CSMLE) of θ_0. A BASMLE $\hat{\alpha}_A$ can be $\hat{\alpha}_A = \hat{\alpha} + H_{\alpha,n,m}^{-1}(\hat{\alpha})\mu_{\alpha,n,m}(\hat{\alpha})/m$, where $H_{\alpha,n,m}^{-1}(\hat{\alpha}) = 1/n\sum_{i=1}^n(\partial^2 \ln f_{n,m}(w(\hat{\alpha}),z_i)/\partial\alpha\partial\alpha')$ and $\mu_{\alpha,n,m}(\alpha)$ has a similar expression as $\mu_{n,m}(\theta)$ with $\partial h(v,z,\theta)/\partial\alpha$ and $\partial f_m(\theta,z)/\partial\alpha$ replacing the corresponding partial derivatives with respect to θ. A BASMLE for θ_0 can be constructed as $\hat{\theta}_{c,A} = w(\hat{\alpha}_A)$. The $\hat{\theta}_{c,A}$ satisfies the constraints $r(\hat{\theta}_{c,A}) = 0$ and can be regarded as a constrained BASMLE (CBASMLE) of θ_0. As $\hat{\theta}_{c,A} = \hat{\theta}_u + [w(\hat{\alpha}_A) - w(\hat{\alpha})]$, the difference within the bracket represents the bias adjustment.

The BASMLE's may be useful for the construction of test statistics. Instead of $r(\hat{\theta}_u)$, one may use $r(\hat{\theta}_A)$ in a possible bias-adjusted SW statistic W^{bc} where

$$W^{bc} = nr'(\hat{\theta}_A)[\tilde{R}\tilde{\mathcal{I}}^{-1}\tilde{R}']^{-1}r(\hat{\theta}_A), \tag{4.2}$$

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

where $\tilde{R}$ is a consistent estimator of R and $\tilde{\mathcal{I}}$ is a consistent estimator of $\mathcal{I}$. For a possible bias-adjusted SC test statistic, one may consider the use of $\partial L_{n,m}(\hat{\theta}_{c,A})/\partial\theta$ in the score test formulation. The $\hat{\theta}_{c,A}$ in the estimated simulated score is able to correct the asymptotic bias of the CSMLE $\hat{\theta}_c$ but the asymptotic bias in $\partial \ln L_{n,m}(\theta_0)/\partial\theta$ remains. A modified SSC statistic that adjusts these two sources of asymptotic biases is S^{bc} where

$$S^{bc} = \left(\frac{1}{\sqrt{n}}\frac{\partial \ln L_{n,m}(\hat{\theta}_{c,A})}{\partial\theta} - n^{1/2}\hat{\mu}/m\right)'\tilde{\mathcal{I}}^{-1}\left(\frac{1}{\sqrt{n}}\frac{\partial \ln L_{n,m}(\hat{\theta}_{c,A})}{\partial\theta} - n^{1/2}\hat{\mu}/m\right), \tag{4.3}$$

where $\hat{\mu}$ is a consistent estimate of μ. For the SLR test statistic, a simple adjustment by replacing SMLE's with their bias-adjusted counterparts is not available.[6]

PROPOSITION 4. *Assume that the regularity conditions hold and* $\lim_{n\to\infty} n^{1/2}m^{-1} = c$ *a finite constant. Then, under* H_0, *the modified* SW *and* SSC *test statistics* W^{bc} *and* S^{bc} *are asymptotically* χ^2_ν. *On the other hand, under the local alternative* H_n *in* (3.4), *their limiting distributions are* $\chi^2_\nu(\delta_2)$ *where* $\delta_2 = \Delta' R'(R\mathcal{I}^{-1}R')^{-1}R\Delta$.

Because the limiting distributions of these bias-corrected statistics under the null and alternative hypotheses are identical, they are asymptotically equivalent. These test statistics are bias-adjusted because the asymptotic biases in the original statistics under H_0 have been eliminated.

5. SIMULATED TEST STATISTICS WITH VARIOUS VARIANCES

The formulation of various simulated test statistics depends on variances of simulated estimates of θ_0. These variances need be estimated. For the classical test statistics without simulation, an information matrix can be estimated by using the negative Hessian matrix of a log likelihood function. For panel data models, a computationally simpler estimator of the information matrix can be constructed by an outer product of gradient vectors of the log likelihood (Berndt, Hall, Hall, and Hausman, 1974). These suggest that for simulated test statistics, the $\mathcal{I}$ in (2.5) and (2.6) can be estimated by similar matrices with simulation. The asymptotic variance of the SMLE can be estimated by $V_{h,n,m}(\theta)$ or $V_{p,n,m}(\theta)$ where

$$V_{h,n,m}(\theta) = -H^{-1}_{n,m}(\theta)/n \tag{5.1}$$

and

$$V_{p,n,m}(\theta) = \left\{\sum_{i=1}^{n}\frac{\partial \ln f_m(\theta, z_i)}{\partial\theta}\frac{\partial \ln f_m(\theta, z_i)}{\partial\theta'}\right\}^{-1}, \tag{5.2}$$

evaluated at a consistent estimate $\hat{\theta}$ of θ_0. From quasi-likelihood theory (see, e.g., White, 1982), a more robust variance estimate of a maximum quasi-likelihood estimator involves a three-term product of the inverse of the Hessian matrix and

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

the outer-product matrix of the log likelihood function. This suggests another possible variance estimate,

$$V_{r,n,m}(\theta) = V_{h,n,m}(\theta)V_{p,n,m}^{-1}(\theta)V_{h,n,m}(\theta). \tag{5.3}$$

As n and m go to infinity, these three variances multiplied by n converge in probability to $\mathcal{I}^{-1}$. In a finite sample setting, the performance of various test statistics could be influenced by the variance estimates used.[7]

In a Wald-type test statistic for $r(\theta_0) = 0$ based on a simulated estimator, for example, $\hat{\theta}_u$, the quadratic matrix in its formulation is the variance matrix of $r(\hat{\theta}_u)$. With the variance estimates in (5.1)–(5.3), there can be three slightly different versions of the SW test statistics. For example, with the SMLE $\hat{\theta}_u$, the three SW statistics are W_p^{sl}, W_h^{sl}, and W_r^{sl} where

$$W_x^{sl} = r'(\hat{\theta}_u)[R(\hat{\theta}_u)V_{x,n,m}(\hat{\theta}_u)R'(\hat{\theta}_u)]^{-1}r(\hat{\theta}_u) \tag{5.4}$$

for $x = p, h, r$. With the BASMLE $\hat{\theta}_c$, the corresponding bias-adjusted SW statistics will be denoted by W_p^{bc}, W_h^{bc}, and W_r^{bc}.

For the SSC and SLR tests, formulations with $V_{r,n,m}$ need further consideration. While both $nV_{h,n,m}$ and $nV_{p,n,m}$ are consistent estimates of $\mathcal{I}^{-1}$, $V_{r,n,m}$ is constructed as a robust estimate of the variance of the SMLE of θ_0 but not as a direct estimate of $\mathcal{I}^{-1}$. Because $\sqrt{n}(\hat{\theta}_u - \theta_0) = \mathcal{I}^{-1}S_{n,m} + o_P(1)$ from (2.2), $nV_{r,n,m}$ is an estimate of the variance matrix of $A_{n,m}$ where $A_{n,m} = \mathcal{I}^{-1}S_{n,m}$. From Proposition B.2.7 of Appendix B, $(1/\sqrt{n})(\partial \ln L_{n,m}(\hat{\theta}_c)/\partial\theta) = R'(R\mathcal{I}^{-1}R')^{-1}RA_{n,m} + o_P(1)$ under H_0. It follows that $R\mathcal{I}^{-1}(1/\sqrt{n})(\partial \ln L_{n,m}(\hat{\theta}_c)/\partial\theta) = RA_{n,m} + o_P(1)$. This motivates the following modified SSC statistic S_r^{sl}, which incorporates the robust variance, where

$$S_r^{sl} = \frac{\partial \ln L_{n,m}(\hat{\theta}_c)}{\partial\theta'}(n\hat{\mathcal{I}})^{-1}\hat{R}'[\hat{R}V_{r,n,m}\hat{R}']^{-1}\hat{R}(n\hat{\mathcal{I}})^{-1}\frac{\partial \ln L_{n,m}(\hat{\theta}_c)}{\partial\theta}, \tag{5.5}$$

where $\hat{\mathcal{I}}$ and $\hat{R}$ are, respectively, some consistent estimates of $\mathcal{I}$ and R.[8] A corresponding bias-adjusted SSC test statistic is

$$S_r^{bc} = \left(\frac{\partial \ln L_{n,m}(\hat{\theta}_{c,A})}{\partial\theta} - \frac{n^{1/2}}{m}\hat{\mu}\right)'(n\hat{\mathcal{I}})^{-1}\hat{R}'[\hat{R}V_{r,n,m}\hat{R}']^{-1}\hat{R}(n\hat{\mathcal{I}})^{-1} \times \left(\frac{\partial \ln L_{n,m}(\hat{\theta}_{c,A})}{\partial\theta} - \frac{n^{1/2}}{m}\hat{\mu}\right), \tag{5.6}$$

where $\hat{\mu}$ is μ evaluated at $\hat{\theta}_c$ or $\hat{\theta}_{A,c}$. The robust variance can also be incorporated into SLR tests. From Appendix B, $2[\ln L_{n,m}(\hat{\theta}_u) - \ln L_{n,m}(\hat{\theta}_c)] = -n(\hat{\theta}_c - \hat{\theta}_u)'H_{n,m}(\hat{\theta}_u)(\hat{\theta}_c - \hat{\theta}_u) + o_P(1)$. As $\sqrt{n}(\hat{\theta}_u - \hat{\theta}_c) = \mathcal{I}^{-1}R'(R\mathcal{I}^{-1}R')^{-1}RA_{n,m} + o_P(1)$ under H_0 from Proposition B.2.5, it follows that $R\sqrt{n}(\hat{\theta}_u - \hat{\theta}_c) = RA_{n,m} + o_P(1)$. Hence $(\hat{\theta}_u - \hat{\theta}_c)'R'[RV_{n,m}R']^{-1}R(\hat{\theta}_u - \hat{\theta}_c)$ is asymptotically (noncentral) chi-square distributed. This suggests a possible SLR statistic LR_r^{sl} where

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

$$LR_r^{sl} = 2[\ln L_{n,m}(\hat{\theta}_u) - \ln L_{n,m}(\hat{\theta}_c)] + (\hat{\theta}_c - \hat{\theta}_u)'\{nH_{n,m}(\hat{\theta}_u) + \hat{R}'[\hat{R}V_{n,m}\hat{R}']^{-1}\hat{R}\}(\hat{\theta}_c - \hat{\theta}_u). \quad \textbf{(5.7)}$$

PROPOSITION 5. *Assume that the regularity conditions are satisfied. Under the null hypothesis* $H_0: r(\theta_0) = 0$, *if* $\lim_{n\to\infty} n^{1/2}m^{-1} = c$ *a finite constant, the* W_r^{sl}, S_r^{sl}, *and* LR_r^{sl} *test statistics are asymptotically* $\chi_\nu^2(\delta_0)$. *On the other hand, under the sequence of local alternatives in* (3.4), *these statistics are asymptotically* $\chi_\nu^2(\delta_1)$.

PROPOSITION 6. *Assume that the regularity conditions hold and* $\lim_{n\to\infty} n^{1/2}m^{-1} = c$ *a finite constant. Then, under* H_0, *the modified* W_r^{bc} *and* S_r^{bc} *test statistics are asymptotically* χ_ν^2. *On the other hand, under the local alternative* H_n *in* (3.4), *their limiting distributions are* $\chi_\nu^2(\delta_2)$.[9]

6. MONTE CARLO RESULTS

A general dynamic discrete choice model was introduced in Heckman (1981). It can incorporate true state dependence, unknown heterogeneity, and serially correlated disturbances. It contains the Markov (MA), Polya (PO), renewal (RN), and habit persistence (HA) (or latent Markov) dynamic models as special cases. An interesting question concerns the possible use of the general dynamic model to distinguish various specific dynamic effects. In this experiment, we investigate the use of simulated classical test statistics for testing various specific dynamic structures within the general model framework.

The general model specified in Monte Carlo experiments is a dynamic panel model:

$$y_{it}^* = \beta x_{it} + \lambda_{ma} d_{i,t-1} + \lambda_{po} \sum_{j=1}^{t} \delta^{j-1} d_{i,t-j} + \lambda_{rn} \sum_{j=1}^{t} \delta^{j-1} \prod_{l=1}^{j} d_{i,t-l} + \lambda_{ha} y_{i,t-1}^* + \sigma\xi_i + \epsilon_{it},$$

where $\epsilon_{it} = \rho\epsilon_{i,t-1} + w_{it}$, $\epsilon_{i0} = 0$, $d_{i0} = 0$, $y_{i0}^* = 0$, and w_{it} are independent and identically distributed (i.i.d.) $N(0,1)$ for all i and t. The error component ξ_i is $N(0,1)$ for all i. The discount rate δ is assumed to be known and is set to be 0.7. The exogenous variables x_{it} are generated as $x_{it} = (1/\sqrt{2})r_{it} + \sqrt{6}s_i$ where r_{it} are i.i.d. truncated normal $N(0,1)$ with support $[-2,2]$ and s_i are independent uniform variables with support on $[-\frac{1}{2},\frac{1}{2}]$. The variance of x is about 1 and its correlation coefficient over time is about 0.5. The y_{it}^* is latent but is related to the observed dichotomous dependent variable d_{it} by $d_{i,t} = 1$ if $y_{it}^* > 0$, and $d_{it} = 0$ otherwise. Sample data are generated with $\beta = 1$, $\rho = 0.4$, and various combinations of λ's. With $\lambda_{ma} = 0.2$, $\lambda_{po} = 0$, $\lambda_{rn} = 0$, and $\lambda_{ha} = 0$, the specific dynamic model is an MA model with unknown heterogeneity and serially correlated disturbances. A PO model is generated with $\lambda_{ma} = 0$, $\lambda_{po} = 0.2$, $\lambda_{rn} = 0$, and $\lambda_{ha} = 0$. An RN model is generated with $\lambda_{ma} = 0$, $\lambda_{po} = 0$, $\lambda_{rn} = 0.2$, and $\lambda_{ha} = 0$. An

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

HA model can be generated with $\lambda_{ma} = 0$, $\lambda_{po} = 0$, $\lambda_{rn} = 0$, and $\lambda_{ha} = 0.2$. The summary statistics reported are all based on 300 replications. The size of cross-sectional units n is 200 and time periods T are 8 and 15 for each unit. The numbers of random draws m for $T = 8$ are from 15 to 100. For $T = 15$, m ranges from 50 to 100.

The model can be estimated by the method of simulated likelihood. The choice probabilities can be effectively simulated by the GHK simulator. For convenient reference, we briefly describe the simulation procedure and the simulated likelihood function. The details can be found in Lee (1997). For each i, generate $u_{i1}^{(j)}, \ldots, u_{i,T-1}^{(j)}$ independent uniform [0,1] random variables and generate $\xi_i^{(j)}$ from an $N(0,1)$ random variable generator, where $j = 1, \ldots, m$. These random draws are kept fixed during an optimization procedure. Let $D_{it} = (2d_{i,t} - 1)$,

$$q_{it}(\theta) = \beta x_{it} + \lambda_{ma} d_{i,t-1} + \lambda_{po} \sum_{j=1}^{t} \delta^{j-1} d_{i,t-j} + \lambda_{rn} \sum_{j=1}^{t} \delta_{j-1} \prod_{l=1}^{j} d_{i,t-l} + \lambda_{ha} y_{i,t-1}^{*} + \sigma \xi_i + \rho \epsilon_{i,t-1},$$

and Φ denotes the standard normal distribution. With given initial conditions, $y_{i1}^{*}, \ldots, y_{i,T-1}^{*}$ can be generated recursively from the following steps, from $t = 1$ to $T - 1$.

(1) Compute $w_{it} = -D_{it} \Phi^{-1}[u_{it} \Phi(D_{it} q_{it}(\theta))]$.
(2) Generate $y_{it}^{*} = q_{it}(\theta) + w_{it}$. The simulated log likelihood function for this model is

$$\ln L_{n,m}(\theta) = \sum_{i=1}^{n} \ln \left\{ \frac{1}{m} \sum_{j=1}^{m} \prod_{t=1}^{T} \Phi \left[D_{it} \left(\beta x_{it} + \lambda_{ma} d_{i,t-1} + \lambda_{po} \sum_{j=1}^{t} \delta^{j-1} d_{i,t-j} + \lambda_{rn} \sum_{j=1}^{t} \delta^{j-1} \prod_{l=1}^{j} d_{i,t-l} + \lambda_{ha} y_{i,t-1}^{*(j)} + \sigma \xi_i^{(j)} + \rho \epsilon_{i,t-1}^{(j)} \right) \right] \right\}.$$

Table 1 reports the actual levels of significance of various SW test statistics with 5% and 10% nominal levels of significance. The critical values for the 5% and 10% nominal levels of significance are from the (central) χ_3^2 distribution. For small or moderate amounts of random draws $m = 15$ or 50, the actual levels of significance can be larger than the nominal levels of significance. For the cases that $T = 8$ with $m = 15$ and $T = 15$ with $m = 50$, the bias-corrected statistics have some improvements upon the bias-unadjusted statistics. However, the differences of the various SW statistics are, in general, small. The agreement between actual and nominal significance levels of the tests is still poor even after the bias correction. In the case of $T = 8$ and $m = 50$, the bias correction may make matters

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

TABLE 1. Significant levels of Wald tests

H_0	T	m	W_p^{sl}	W_h^{sl}	W_r^{sl}	W_p^{bc}	W_h^{bc}	W_r^{bc}
			10% nominal level of significance					
MA	8	15	0.2267	0.2067	0.1800	0.1867	0.1767	0.1633
PO	8	15	0.2433	0.2267	0.1833	0.2667	0.2400	0.2100
RN	8	15	0.2500	0.2000	0.1633	0.2400	0.1900	0.1600
HA	8	15	0.3633	0.3133	0.2600	0.3033	0.2567	0.1900
MA	8	50	0.1467	0.1333	0.1400	0.1567	0.1533	0.1667
PO	8	50	0.1767	0.1533	0.1567	0.2033	0.1933	0.2033
RN	8	50	0.1500	0.1433	0.1267	0.1567	0.1500	0.1467
HA	8	50	0.2100	0.1767	0.1600	0.1967	0.1633	0.1533
MA	15	50	0.1367	0.1433	0.1467	0.1467	0.1433	0.1567
PO	15	50	0.1833	0.1567	0.1467	0.1600	0.1400	0.1400
RN	15	50	0.1667	0.1600	0.1500	0.1567	0.1467	0.1300
HA	15	50	0.3033	0.2667	0.2533	0.1900	0.1700	0.1400
MA	8	100	0.1033	0.0967	0.0967	0.1033	0.1067	0.0967
PO	8	100	0.0733	0.0667	0.0833	0.0733	0.0800	0.0867
RN	8	100	0.0933	0.0867	0.1000	0.0933	0.0967	0.1033
HA	8	100	0.1167	0.1233	0.1133	0.1200	0.1233	0.1167
MA	15	100	0.0903	0.0836	0.0803	0.0769	0.0836	0.0903
PO	15	100	0.0833	0.0833	0.0833	0.0767	0.0867	0.0933
RN	15	100	0.0767	0.0900	0.0800	0.0833	0.0967	0.0900
HA	15	100	0.0970	0.0903	0.0903	0.0903	0.0803	0.0635
			5% nominal level of significance					
MA	8	15	0.1400	0.1300	0.1100	0.1233	0.1200	0.0933
PO	8	15	0.1700	0.1367	0.1100	0.1933	0.1567	0.1433
RN	8	15	0.1500	0.1200	0.1167	0.1433	0.1167	0.1000
HA	8	15	0.2600	0.2167	0.1467	0.2133	0.1633	0.1200
MA	8	50	0.0900	0.0867	0.0800	0.0867	0.0867	0.0800
PO	8	50	0.1067	0.1000	0.0900	0.1233	0.1000	0.0933
RN	8	50	0.1000	0.0733	0.0600	0.0900	0.0800	0.0667
HA	8	50	0.1000	0.0767	0.0700	0.0967	0.0767	0.0667
MA	15	50	0.0733	0.0900	0.0967	0.0667	0.0800	0.0733
PO	15	50	0.0833	0.0833	0.0833	0.0867	0.0967	0.0900
RN	15	50	0.0867	0.0933	0.0767	0.0867	0.0733	0.0800
HA	15	50	0.1833	0.1600	0.1367	0.1000	0.0767	0.0600
MA	8	100	0.0533	0.0567	0.0567	0.0567	0.0600	0.0567
PO	8	100	0.0433	0.0433	0.0500	0.0433	0.0467	0.0533
RN	8	100	0.0600	0.0533	0.0500	0.0600	0.0567	0.0500
HA	8	100	0.0533	0.0500	0.0533	0.0533	0.0533	0.0467
MA	15	100	0.0435	0.0468	0.0569	0.0468	0.0468	0.0569
PO	15	100	0.0433	0.0533	0.0600	0.0500	0.0500	0.0533
RN	15	100	0.0433	0.0333	0.0367	0.0467	0.0333	0.0467
HA	15	100	0.0435	0.0301	0.0301	0.0334	0.0268	0.0201

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

worse as often as not. For a better performance, m needs to be larger. With $m = 100$, the actual levels of significance seem reasonably accurate. The message of Table 1 indicates that only when a large number of draws are used can the agreement between actual and nominal levels for the SW test statistics improve, and with that many random draws the bias correction no longer makes any difference. Table 2 reports results of the SSC tests. The performance of various SSC statistics is better than that of the SW statistics. With $m = 15$ for $T = 8$ and $m = 50$ for $T = 15$, the actual levels of significance are larger than the nominal levels, but the bias-corrected statistics can be greatly improved for those cases. Comparing the various versions with variance estimates, the statistics with robust variance are slightly better than those with outer-product gradients, which, in turn, are slightly better than those with the Hessian. With moderate or large numbers of simulation draws, the actual levels of significance of various SSC test statistics seem reasonably accurate. Table 3 reports the significance of the SLR statistics for the conventional version LR^{sl} and the version LR_r^{sl} that incorporates the robust statistics. With small $m = 15$ for $T = 8$, the actual levels of significance for LR^{sl} are also larger than the nominal levels. The LR_r^{sl} are better for this case. Even with as few as 50 draws, the actual levels of significance are reasonably accurate.

Table 4 reports performances of the SW, SSC, and SLR test statistics in terms of power with $m = 50$ or 100 random draws for $T = 15$. Because the actual levels of significance seem reasonably accurate with moderate or large amounts of simulation draws, the resulting powers can be meaningful. We consider two sets of situations. In the first circumstance, we consider the testing of $(\lambda_{po}, \lambda_{rn}, \lambda_{ha}) = 0$; i.e., the null hypothesis is the MA model, but the true data generating process is one of the other specific dynamic models. The second design is that the true data generating process is the MA process but the null hypothesis is PO, RN, or HA. Some common features emerge from this table. The three types of simulated classical statistics have similar power.[10] These simulated test statistics have strong power to reject the MA model within the general dynamic model framework when the true data generating process is either PO, RN, or HA. On the other hand, the power to reject PO, RN, or HA is weaker when the data generating process is MA.

7. CONCLUSIONS

The main concern of this article is statistical inference based on simulated statistics. Asymptotic properties of various simulated classical test statistics, namely, the Wald, efficient score (Lagrange multiplier), and likelihood ratio statistics, constructed from a simulated likelihood function are investigated. Monte Carlo experiments are conducted to investigate finite sample performance of various simulated classical test statistics.

It is shown that the simulated test statistics can be asymptotically χ^2 distributed and are asymptotically equivalent under a null hypothesis and a sequence of local alternative hypotheses. But, because simulated score vectors with a small or moderate amount of simulation draws can be asymptotically biased, the simu-

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

TABLE 2. Significant levels of simulated efficient score tests

H_0	T	m	S_p^{sl}	S_h^{sl}	S_r^{sl}	S_p^{bc}	S_h^{bc}	S_r^{bc}
			10% nominal level of significance					
MA	8	15	0.1600	0.2300	0.1200	0.1133	0.1733	0.0933
PO	8	15	0.1333	0.1900	0.1033	0.1100	0.1633	0.1000
RN	8	15	0.1367	0.2167	0.1200	0.1167	0.1633	0.0967
HA	8	15	0.2667	0.3167	0.2067	0.1367	0.1967	0.1100
MA	8	50	0.1167	0.1233	0.1067	0.1267	0.1367	0.1133
PO	8	50	0.1133	0.1067	0.0967	0.1133	0.1233	0.0933
RN	8	50	0.1067	0.1100	0.0933	0.1033	0.1133	0.0933
HA	8	50	0.1600	0.1700	0.1467	0.1233	0.1300	0.1033
MA	15	50	0.1300	0.1300	0.1133	0.1100	0.1100	0.0933
PO	15	50	0.1233	0.1267	0.1200	0.1033	0.1100	0.1033
RN	15	50	0.1200	0.1233	0.1033	0.1000	0.1067	0.0933
HA	15	50	0.1833	0.2200	0.1700	0.0900	0.0967	0.0733
MA	8	100	0.1100	0.1100	0.0967	0.1000	0.1133	0.0900
PO	8	100	0.1233	0.1167	0.1067	0.1233	0.1033	0.1100
RN	8	100	0.1133	0.1133	0.1067	0.1067	0.1167	0.1000
HA	8	100	0.1433	0.1567	0.1167	0.1333	0.1433	0.1167
MA	15	100	0.1338	0.1371	0.1204	0.1304	0.1003	0.1037
PO	15	100	0.1267	0.1233	0.1133	0.1067	0.1133	0.1033
RN	15	100	0.1167	0.1133	0.1033	0.1033	0.1167	0.0900
HA	15	100	0.1304	0.1438	0.1104	0.0936	0.1070	0.0769
			5% nominal level of significance					
MA	8	15	0.0967	0.1567	0.0767	0.0633	0.1033	0.0500
PO	8	15	0.0767	0.1067	0.0700	0.0700	0.0900	0.0500
RN	8	15	0.0800	0.1367	0.0667	0.0633	0.0967	0.0467
HA	8	15	0.1433	0.2333	0.1133	0.0800	0.1300	0.0533
MA	8	50	0.0833	0.0767	0.0733	0.0767	0.0733	0.0567
PO	8	50	0.0667	0.0800	0.0567	0.0667	0.0733	0.0600
RN	8	50	0.0700	0.0800	0.0533	0.0667	0.0800	0.0567
HA	8	50	0.0633	0.0967	0.0533	0.0633	0.0600	0.0367
MA	15	50	0.0767	0.0767	0.0700	0.0600	0.0600	0.0467
PO	15	50	0.0733	0.0733	0.0667	0.0667	0.0700	0.0567
RN	15	50	0.0567	0.0533	0.0433	0.0500	0.0567	0.0400
HA	15	50	0.1033	0.1300	0.0933	0.0367	0.0500	0.0300
MA	8	100	0.0600	0.0633	0.0467	0.0533	0.0600	0.0500
PO	8	100	0.0633	0.0500	0.0600	0.0600	0.0500	0.0600
RN	8	100	0.0733	0.0700	0.0533	0.0733	0.0667	0.0567
HA	8	100	0.0700	0.0900	0.0500	0.0533	0.0700	0.0433
MA	15	100	0.0602	0.0602	0.0502	0.0535	0.0502	0.0502
PO	15	100	0.0700	0.0733	0.0633	0.0700	0.0700	0.0700
RN	15	100	0.0567	0.0600	0.0500	0.0633	0.0733	0.0533
HA	15	100	0.0468	0.0602	0.0368	0.0301	0.0368	0.0201

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

TABLE 3. Significant levels of simulated likelihood ratio tests

			10% nominal level		5% nominal level	
H_0	T	m	LR^{sl}	LR_r^{sl}	LR^{sl}	LR_r^{sl}
MA	8	15	0.1633	0.1333	0.0867	0.0567
PO	8	15	0.1467	0.1067	0.0800	0.0567
RN	8	15	0.1567	0.1000	0.0900	0.0600
HA	8	15	0.2600	0.1667	0.1867	0.1000
MA	8	50	0.1033	0.0967	0.0633	0.0567
PO	8	50	0.1000	0.1033	0.0633	0.0467
RN	8	50	0.0967	0.0900	0.0733	0.0567
HA	8	50	0.1433	0.0933	0.0600	0.0467
MA	15	50	0.1000	0.0967	0.0733	0.0667
PO	15	50	0.1100	0.1033	0.0700	0.0533
RN	15	50	0.0833	0.0733	0.0433	0.0433
HA	15	50	0.1767	0.1367	0.1033	0.0833
MA	8	100	0.1033	0.0800	0.0533	0.0567
PO	8	100	0.0800	0.0867	0.0467	0.0533
RN	8	100	0.0933	0.0967	0.0567	0.0533
HA	8	100	0.1300	0.1033	0.0600	0.0433
MA	15	100	0.1003	0.0769	0.0468	0.0569
PO	15	100	0.0867	0.0800	0.0667	0.0667
RN	15	100	0.0933	0.0767	0.0467	0.0433
HA	15	100	0.1070	0.1003	0.0368	0.0301

TABLE 4. Powers of simulated tests ($T = 15$)

			10% level of significance			5% level of significance		
H_0	Data	m	S_p^{sl}	W_p^{sl}	LR^{sl}	S_p^{sl}	W_p^{sl}	LR^{sl}
MA	PO	50	0.9400	0.9300	0.9300	0.8800	0.8667	0.8967
MA	RN	50	0.8967	0.8467	0.8767	0.8067	0.7500	0.7933
MA	HA	50	0.7667	0.8367	0.7467	0.6700	0.7533	0.6500
MA	PO	100	0.9322	0.9695	0.9322	0.8847	0.9458	0.8915
MA	RN	100	0.8796	0.8829	0.8495	0.7793	0.8027	0.7525
MA	HA	100	0.8662	0.8595	0.8328	0.7692	0.7926	0.7525
PO	MA	50	0.5167	0.5833	0.4900	0.4000	0.4333	0.3533
RN	MA	50	0.3733	0.4700	0.3233	0.2600	0.3333	0.2400
HA	MA	50	0.6700	0.7467	0.6600	0.5667	0.6567	0.5700
PO	MA	100	0.4950	0.3746	0.4448	0.3645	0.2140	0.3211
RN	MA	100	0.3658	0.2215	0.3389	0.2617	0.1443	0.2148
HA	MA	100	0.5652	0.5284	0.5351	0.4415	0.4080	0.4080

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

lated classic statistics may have a noncentral parameter in their χ^2 limiting distribution under the null hypothesis. The noncentrality parameter can be removed with proper bias-adjustment procedures. Bias-adjusted simulated statistics, which can incorporate various estimates of variances of simulated maximum likelihood estimators, are considered. They are shown to be asymptotically equivalent and their limiting distributions are central χ^2 distributed under the null hypothesis.

Monte Carlo results demonstrate finite sample performance of various simulated test statistics. The performance of the bias-adjusted simulated efficient score statistics can be better than that of the unadjusted statistics in that the actual levels of significance can be more accurate. The versions that incorporate the robust variance formulae can be slightly better. However, for adequate matching of actual levels with nominal levels of significance, the use of a larger number of simulation draws is more effective.[11] The Monte Carlo experiments are conducted for the testing of specific dynamic structures in a general dynamic discrete choice framework originated in Heckman (1981). As a by-product, it is found that the various simulated test statistics have strong power to reject the popular Markov specification if true state dependence is from the Polya, renewal, or habit persistence models. The power to reject the Polya, renewal, or habit persistence structures is weaker when the data generating process is a Markov process.

NOTES

1. Some of the components of z may be (strict) exogenous variables. The term $f(\theta, z)$ would be the density or probability functions of the dependent variables conditional on the set of exogenous variables.

2. Simulation methods with identical random draws for all sample observations can be found in Lee (1992).

3. The present result allows semiuniform convergence in probabilities under slightly stronger assumptions.

4. A sequence of random functions $g_n(z,\theta)$ is said to converge to 0 in probability semiuniformly in θ if $\lim_{n\to\infty} \inf_\theta P[|g_n(z,\theta)| < \epsilon] = 1$ for any ϵ (Amemiya, 1985, p. 106).

5. The possible asymptotic biases in $\hat{\theta}_u$, $\hat{\theta}_c$, and $S_{n,m}(\theta_0)$ may increase or decrease powers of these test statistics under H_n depending on the value of Δ. If $\Delta = -\mathcal{I}^{-1}c\mu$, the biases from two sources would cancel each other and the resulting asymptotic limiting distribution under H_n could be a central χ^2.

6. A possible bias-adjusted SLR statistic can be LR^{bc} where $LR^{bc} = 2[\ln L_{n,m}(\hat{\theta}_A) - \ln L_{n,m}(\hat{\theta}_{c,A}) + (\partial \ln L_{n,m}(\hat{\theta}_A)/\partial\theta')(\hat{\theta}_{c,A} - \hat{\theta}_A)]$. This LR^{bc} statistic has the properties given in Proposition 4. However, our Monte Carlo results with this statistic are not better than those of bias-unadjusted SLR statistics. So it is not emphasized in the text.

7. Monte Carlo results for the use of Hessian matrix and the outer-product-gradient matrices for the classical score test can be found in Davidson and MacKinnon (1983) and Bera and McKenzie (1986). From our Monte Carlo experiments, there is evidence that the robust variance formula may provide better variance estimates in terms of smaller biases but they have much larger standard deviations than the outer-product or Hessian estimates.

8. The outer-product form has been used for the estimate of $\mathcal{I}$ for the subsequent Monte Carlo results.

9. A bias-corrected SLR statistic that incorporates a robust variance can be $LR_r^{bc} = 2[\ln L_{n,m}(\hat{\theta}_A) - \ln L_{n,m}(\hat{\theta}_{c,A})] + (\partial \ln L_{n,m}(\hat{\theta}_A)/\partial\theta')(\hat{\theta}_{c,A} - \hat{\theta}_A) + (\hat{\theta}_A - \hat{\theta}_{c,A})'\{nH_{n,m}(\hat{\theta}_u) + \hat{R}'[\hat{R}V_{n,m}\hat{R}']^{-1}\hat{R}\}(\hat{\theta}_A - \hat{\theta}_{c,A})$. It can be shown that this SLR has the asymptotic property given in Proposition 6. Again, Monte Carlo results with it are not better.

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

10. Powers for the various versions with different variance estimates are also similar. For easy exposition, they are not reported here.

11. From the Monte Carlo results, it seems that the amount of random draws for hypothesis testing needs to be larger than the amount of random draws for parameter estimation. For the latter, see Lee (1997).

REFERENCES

Amemiya, T. (1985) *Advanced Econometrics*. Cambridge, Massachusetts: Harvard University Press.

Bera, A.K. & C.R. McKenzie (1986) Alternative forms and properties of the score test. *Journal of Applied Statistics* 13, 13–25.

Berndt, E.R., B.H. Hall, R.E. Hall, & J.A. Hausman (1974) Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement* 3, 653–666.

Bickel, P.J., C.A.J. Klaassen, Y. Ritov, & J.A. Wellner (1993) *Efficient and Adaptive Estimation for Semiparametric Models.* Baltimore: Johns Hopkins University Press.

Borsch-Supan, A. & V.A. Hajivassiliou (1993) Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. *Journal of Econometrics* 58, 347–368.

Davidson, R. & J.G. MacKinnon (1983) Small sample properties of alternative forms of the Lagrange multiplier test. *Economics Letters* 12, 269–275.

Engle, R.F. (1984) Wald, likelihood ratio and Lagrange multiplier tests in econometrics. In Z. Griliches & M.D. Intriligator (eds.), *Handbook of Econometrics*, vol. II, pp. 775–826. Amsterdam: North-Holland.

Gallant, A.R. & A. Holly (1980) Statistical inference in an implicit, nonlinear, simultaneous equation model in the context of maximum likelihood estimation. *Econometrica* 48, 697–720.

Geweke, J. (1991) Efficient simulation from the multivariate normal and Student-t distributions subject to linear constraints. In *Computer Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface*, pp. 571–578. Alexandria, Virginia: American Statistical Association.

Godfrey, L.G. (1988) *Misspecification Tests in Econometrics.* New York: Cambridge University Press.

Gourieroux, C. & A. Monfort (1993) Simulation-based inference: A survey with special reference to panel data models. *Journal of Econometrics* 59, 5–33.

Hajek, J. & Z. Sidak (1967) *Theory of Rank Tests.* New York: Academic Press.

Hajivassiliou, V.A. (1993) Simulation estimation methods for limited dependent variable models. In G.S. Maddala, C.R. Rao, & H.D. Vinod (eds.), *Handbook of Statistics 11: Econometrics*, pp. 519–543. Amsterdam: North-Holland.

Hajivassiliou, V.A. & D. McFadden (1990) The Method of Simulated Scores, with Application to Models of External Debt Crises. Cowles Foundation Discussion paper 967, Cowles Foundation, Yale University.

Heckman, J.J. (1981) Statistical models for discrete panel data. In C.F. Manski & D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp. 114–178. Cambridge, Massachusetts: MIT Press.

Hendry, D.F. & J.F. Richard (1992) Likelihood evaluation for dynamic latent variable models. In H.M. Amman, D.A. Belsley, & L.F. Pau (eds.), *Computational Economics and Econometrics*, pp. 3–17. Amsterdam: Kluwer.

Keane, M.P. (1993) Simulation estimation for panel data models with limited dependent variables. In G.S. Maddala, C.R. Rao, & H.D. Vinod (eds.), *Handbook of Statistics 11: Econometrics*, pp. 545–571. Amsterdam: North-Holland.

Keane, M.P. (1994) A computationally practical simulation estimator for panel data. *Econometrica* 62, 95–116.

Laroque, G. & B. Salanie (1993) Simulation-based estimation of models with lagged latent variables. *Journal of Applied Econometrics* 8, S119–S133.

Laroque, G. & B. Salanie (1994) Estimating the canonical disequilibrium model: Asymptotic theory and finite sample properties. *Journal of Econometrics* 62, 165–210.

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

Le Cam, L. (1960) Locally asymptotically normal families of distributions. *University of California Publications in Statistics* 3, 37–98.
Lee, L.F. (1992) On efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete response models. *Econometric Theory* 8, 518–552.
Lee, L.F. (1995) Asymptotic bias in simulated maximum likelihood estimation of discrete choice models. *Econometric Theory* 11, 437–483.
Lee, L.F. (1997) Simulated maximum likelihood estimation of dynamic discrete choice statistical models—some Monte Carlo results. *Journal of Econometrics* 82, 1–35.
Lerman, S. & C. Manski (1981) On the use of simulated frequencies to approximate choice models. In C. Manski & D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp. 305–319. Cambridge, Massachusetts: MIT Press.
McFadden, D. (1989) A method of simulated moments of estimation of discrete choice models without numerical integration. *Econometrica* 57, 995–1026.
Pakes, A. & D. Pollard (1989) Simulation and the asymptotics of optimization estimators. *Econometrica* 54, 755–785.
Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics*. New York: Wiley.
White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1–25.

APPENDIX A

1. *Model Assumptions*

The following regularity conditions are assumed for analysis in this article.

Assumption 1.

(1) The sample observations z_i, $i = 1, \ldots, n$ are i.i.d. with a compact support Z.
(2) The parameter space Θ is a compact convex subset of a k-dimensional Euclidean space and the true parameter vector θ_0 is in the interior of Θ.
(3) The mapping $r : R^k \to R^\nu$ for the constraints $r(\theta_0) = 0$ is continuously differentiable in a neighborhood of θ_0. The derivative matrix $\partial r(\theta_0)/\partial\theta$ evaluated at θ_0 has full rank ν.
(4) The function $f(\theta, z)$ is continuous and bounded away from zero on $\Theta \times Z$.
(5) The $f(\theta, z)$ is differentiable in θ up to the second order. Those derivatives are continuous on $\Theta \times Z$.
(6) The θ_0 is the unique maximizer of $E\{\ln f(\theta, z)\}$.
(7) The information matrix $\mathcal{I}$, where $\mathcal{I} = E\{[\partial \ln f(\theta_0, z)/\partial\theta][\partial \ln f(\theta_0, z)/\partial\theta']\}$, is nonsingular.

Assumption 2.

(1) The random draws of v are from a common distribution and are independent of z and θ.
(2) The function $h(v, z, \theta)$ is a conditionally unbiased estimator of $f(\theta, z)$ conditional on z, for each θ.
(3) The $h(v, z, \theta)$ is twice differentiable in θ. Those derivatives are continuous on $\Theta \times Z$.

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

(4) The absolute values of $h(v,z,\theta)$ and its first and second order derivatives with respect to θ are dominated by square integrable functions of v uniformly in z and θ.
(5) The first six order moments of $h(v,z,\theta)$ and the first four order moments of $\partial h(v,z,\theta)/\partial\theta$ exit and are bounded functions on Θ.

The regularity conditions in Assumption 1 contain familiar regularity conditions for maximum likelihood estimation for panel data models. Condition 6 provides the identification of θ_0. Condition 7 excludes irregular singular information matrix cases. Because the asymptotic expansion of the simulated score involves the log-simulated likelihood function and its derivatives, simulation errors up to certain orders need to be analyzed. The moment conditions and domination by integrable functions in Assumption 2 guarantee the existence, boundedness, and continuity of these moments in θ. These properties are useful for the analysis of test statistics under a sequence of local Pitman-type alternatives. The compactness assumptions in conditions 1 and 2 in Assumption 1 combined with condition 5 guarantee the uniform convergence of the simulated likelihood function and its derivatives to the true likelihood function and derivatives (Amemiya, 1985, Theorem 4.2.1). Condition 4 in Assumption 1 is for technical reasons. It guarantees the uniform convergence of the logarithm of the simulated likelihood and the stochastic boundedness of $1/f_m(\theta,z)$ uniformly in θ and z. This condition is technically difficult to be relaxed. A similar issue occurs in semiparametric estimation, which does not have convenient treatment without the introduction of trimming. For discrete choice models and limited dependent variable models with double truncations, a bounded regressor assumption is sufficient for this condition to hold.

2. *Basic Asymptotic Expansions and Moments*
The following lemma is Lemma A in Serfling (1980, p. 304).

LEMMA A.1. *Let y_i, $i = 1,\ldots,m$ be independent random variables with zero mean. Let r be an even integer. Then, $E|\sum_{j=1}^m y_j|^r \le A_r m^{(r/2)-1}\sum_{j=1}^m E|y_j|^r$, where A_r is a universal constant dependent only on r.*

The following proposition generalizes a proposition in Lee (1995, Proposition A.3) to incorporate semiuniform convergence in probability. It is the key result to establish the asymptotic expansion of simulated score in Proposition 1 of the text.

PROPOSITION A.2. *Suppose that $(v_i^{(1)},\ldots,v_i^{(m)},z_i)$, $i = 1,\ldots,n$ are identically distributed random vectors. Let $q_{mi}(\theta) = s(z_i,\theta)[1/m\sum_{j=1}^m p(v_i^{(j)},z_i,\theta)]^{r_1}\times[1/m\sum_{j=1}^m t(v_i^{(j)},z_i,\theta)]^{r_2}$, where r_1 and r_2 are nonnegative integers and s, p, and t are measurable functions of (v,z) for each θ such that, conditional on z, $E(p(v,z,\theta)|z) = E(t(v,z,\theta)|z) = 0$ for all θ. If $r_1 > 0$ but $r_2 = 0$, assume that $E(s^2(z,\theta)p^{2r_1}(v,z,\theta))$ exists and is bounded in θ. Otherwise, assume that $E(|s(z,\theta)|^a p^{2r_1 a}(v,z,\theta))$ and $E(|s(z,\theta)|^b t^{2r_2 b}(v,z,\theta))$ exist and are bounded, where $(1/a)+(1/b)=1$, some $a>1$.*

Then $1/n\sum_{i=1}^n|q_{mi}(\theta)| = O_P(m^{-[(r_1+r_2)/2]})$ semiuniformly in θ. Furthermore, if $(v_i^{(1)},\ldots,v_i^{(m)},z_i)$, $i = 1,\ldots,n$ are i.i.d. random vectors, then $1/\sqrt{n}\sum_{i=1}^n[q_{mi}(\theta) - \mu_{q_m}(\theta)] = O_P(m^{-[(r_1+r_2)/2]})$ semiuniformly in θ, where $\mu_{q_m}(\theta) = E(q_{mi}(\theta))$ is the mean of $q_{mi}(\theta)$.

Proof. By the Holder inequality,

$$E(q_{mi}^2(\theta)) \le E^{1/a}\left\{|s(z,\theta)|^a\left[\frac{1}{m}\sum_{j=1}^m p(v_j,z,\theta)\right]^{2r_1 a}\right\} \times E^{1/b}\left\{|s(z,\theta)|^b\left[\frac{1}{m}\sum_{j=1}^m t(v_j,z,\theta)\right]^{2r_2 b}\right\}.$$

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

Lemma A.1 implies that

$$E\left(\left[\frac{1}{m}\sum_{j=1}^{m} p(v_j,z,\theta)\right]^{2r_1a}|z\right) \le \frac{c}{m^{r_1a}}E(p^{2r_1a}(v,z,\theta)|z)$$

and $E([1/m\sum_{j=1}^{m} t(v_j,z,\theta)]^{2r_2b}|z) \le (c/m^{r_2b})E(t^{2r_2b}(v,z,\theta)|z)$ where c is a constant. Therefore, $E(q_{mi}^2(\theta)) \le (cd/m^{r_1+r_2})$ for all θ, where $d = \sup_\theta E^{1/a}\{|s(z,\theta)|^a p^{2r_1a} \times (v,z,\theta)\}E^{1/b}\{|s(z,\theta)|^b t^{2r_2b}(v,z,\theta)\}$. By the Markov and absolute moments inequalities,

$$P\left(n^{-1}m^{(r_1+r_2)/2}\sum_{i=1}^{n}|q_{mi}(\theta)| \ge \epsilon\right) \le \epsilon^{-1}m^{(r_1+r_2)/2}E(|q_{mi}(\theta)|)$$
$$\le \epsilon^{-1}m^{(r_1+r_2)/2}E^{1/2}(|q_{mi}(\theta)|^2) \le \epsilon^{-1}c^{1/2}d^{1/2}$$

for all θ. Because ϵ is arbitrary and c and d are finite constants, $n^{-1}m^{(r_1+r_2)/2} \times \sum_{i=1}^{n}|q_{mi}(\theta)| = O_P(1)$ semiuniformly in θ. This proves the first part of the results.

When the random vectors $(v_i^{(1)},\ldots,v_i^{(m)},z_i)$ are i.i.d. for all i, the Chebyshev inequality implies

$$P\left(\left|n^{-1/2}m^{(r_1+r_2)/2}\sum_{i=1}^{n}(q_{mi}(\theta)-\mu_{q_m}(\theta))\right| \ge \epsilon\right) \le \epsilon^{-2}m^{r_1+r_2}E(q_{mi}(\theta)-\mu_{q_m}(\theta))^2$$
$$\le \epsilon^{-2}cd$$

for all θ. Because ϵ is arbitrary, $n^{-1/2}m^{(r_1+r_2)/2}\sum_{i=1}^{n}(q_{mi}(\theta)-\mu_{q_m}(\theta)) = O_P(1)$ semiuniformly in θ. This proves the second part of the results. The proof of the case with $r_2 = 0$ can be simplified without using the Holder inequality. ■

Proof of Proposition 1. Equation (2.3) is simply an identity by expanding the difference of the ratios $[\partial f_m(\theta,z)/\partial\theta]/f_m(\theta,z)$ and $[\partial f(\theta,z)/\partial\theta]/f(\theta,z)$. The means of $P_{n,m}(\theta)$ and $Q_{n,m}(\theta)$ are from direct calculations. Proposition A.2 provides the stochastic orders and their semiuniform convergence in θ. For the remainder term $R_n(\theta)$, one has to be careful about $f_m(\theta,z)$ as a denominator. Because $f(\theta,z)$ is bounded away from zero, the uniform convergence of $f_m(\theta,z)$ to $f(\theta,z)$ on $\Theta\times Z$ implies that $f_m(\theta,z)$ is bounded away from zero in probability for large m and, hence, $\sup_{\Theta\times Z}\{1/f_m(\theta,z)\} = O_P(1)$. Therefore,

$$|R_{n,m}(\theta)| \le O_P(1)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\|\left(\frac{\partial f_m(\theta,z_i)}{\partial\theta}-\frac{\partial f(\theta,z_i)}{\partial\theta}\right)(f_m(\theta,z_i)-f(\theta,z_i))^2\right.$$
$$\left.+\frac{\partial \ln f(\theta,z_i)}{\partial\theta}(f_m(\theta,z_i)-f(\theta,z_i))^3\right\| \le O_P(n^{1/2}m^{-3/2}),$$

semiuniformly in (θ,z), by using the first part of the results in Proposition A.2. ■

APPENDIX B

1. *Some Basic Asymptotic Properties under* H_n

The concept of contiguity of probability measures is useful. Suppose that $\{P_n\}$ and $\{Q_n\}$ are two sequences of probability measures. The sequence $\{Q_n\}$ is said to be contiguous to $\{P_n\}$ if for any sequence of random variables (or events) T_n for which $T_n \to 0$ in P_n-probability,

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

$T_n \to 0$ in Q_n-probability (Le Cam, 1960; Hajek and Sidak, 1967). The contiguity of $\{Q_n\}$ to $\{P_n\}$ means that Q_n is asymptotically absolutely continuous with respect to P_n in the sense of domination of measures. A useful result of contiguity can be derived from the induced log likelihood ratio. Let $f_n(z)$ be the density function corresponding to P_n and $g_n(z)$ be the one corresponding to Q_n. Le Cam's first lemma states that if the sequence of distributions of the log likelihood ratio $\ln(g_n(z)/f_n(z))$ under P_n converges to a normal distribution $N(-\frac{1}{2}\sigma^2, \sigma^2)$ for some σ^2, then $\{Q_n\}$ is contiguous to $\{P_n\}$ (Le Cam, 1960; Bickel, Klaassen, Ritov, and Wellner, 1993, Appendix A.9). For our purpose, the sequence of $\{P_n\}$ corresponds to the distribution of model under H_0 with the true parameter θ_0, and $\{Q_n\}$ corresponds to the sequence of local alternatives H_n with $\theta_n = \theta_0 + \Delta/\sqrt{n}$. The contiguity of $\{Q_n\}$ to $\{P_n\}$ is useful in the following lemma.

LEMMA B.1.1. *Let T_n be a statistic constructed with a sample of size n for the model under our assumptions. If $T_n \xrightarrow{p} 0$ under H_0, $T_n \xrightarrow{p} 0$ under H_n.*

Proof. The log density function under H_0 is $\ln L_n(\theta_0) = \sum_{i=1}^n \ln f(\theta_0, z_i)$, and the log density function under H_n is $\ln L_n(\theta_n) = \sum_{i=1}^n \ln f(\theta_n, z_i)$. Because $\theta_n = \theta_0 + \Delta/\sqrt{n}$, $\ln L_n(\theta_n) - \ln L_n(\theta_0) = (\partial \ln L_n(\theta_0)/\partial\theta')(\Delta/\sqrt{n}) + \frac{1}{2}((1/n)\Delta'(\partial^2 \ln L_n(\bar{\theta}_n)/\partial\theta\partial\theta')\Delta)$, where $\bar{\theta}_n$ lies between θ_n and θ_0. Under the assumed regularity conditions, a uniform law of large numbers for i.i.d. sample implies that $(1/n)(\partial^2 \ln L_n(\bar{\theta}_n)/\partial\theta\partial\theta') \xrightarrow{p} -\mathcal{I}$, and the central limit theorem for i.i.d. random variables implies that $(1/\sqrt{n})(\partial \ln L_n(\theta_0)/\partial\theta) \xrightarrow{D} N(0, \mathcal{I})$ under H_0. Hence $\ln L_n(\theta_n) - \ln L_n(\theta_0) \xrightarrow{D} N(-\frac{1}{2}\Delta'\mathcal{I}\Delta, \Delta'\mathcal{I}\Delta)$ under H_0. The result follows from Le Cam's lemma. ■

An immediate consequence of this lemma is that the consistency of SMLE's and uniform convergence for statistics under H_n follow from those under H_0.

PROPOSITION B.1.1. *The* SMLE $\hat{\theta}_u$ *and its* BASMLE $\hat{\theta}_A$ *are consistent under H_n as its data generating process.*

Proof. The consistency of $\hat{\theta}_u$ under H_0 follows as when both n and m go to infinity, $\ln L_{n,m}(\theta)/n$ converges in probability uniformly in $\theta \in \Theta$ to $E(\ln f(\theta, z))$, which attains its unique global maximum at θ_0. The $\hat{\theta}_A$ convergences to θ_0 under H_0 because the bias adjustment term goes to zero as m goes to infinity. The consistency of these estimators under H_0 implies their consistency under H_n from Lemma B.1.1. ■

PROPOSITION B.1.2. *Suppose that $T_n(\theta)$ converges in probability uniformly in θ to a continuous limit function $T(\theta)$ under H_0, and $\hat{\theta}$ is a consistent estimator of θ_0 under H_0. Then $T_n(\hat{\theta}) \xrightarrow{p} T(\theta_0)$ under H_n.*

Proof. These conditions imply that $T_n(\hat{\theta}) - T(\theta_0) \xrightarrow{p} 0$ under H_0. The result follows by the contiguity of H_n to H_0. ■

For convergence in distribution under H_n, the classical Lindeberg–Levy central limit theorem (CLT) is useful.

LEMMA B.1.2. *Let $\{z_i\}$ be i.i.d. random variables with mean ξ_n and variance σ_n under H_n. Then, under H_n, $1/\sqrt{n}\sum_{i=1}^n((z_i - \xi_n)/\sigma_n) \xrightarrow{D} N(0,1)$.*

Proof. The result follows by using the expansion of characteristic function as in the proof of the classical result under H_0. ■

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

2. *Asymptotic Distributions of Simulated Likelihood Estimators and Simulated Scores*
Simulated test statistics depend on simulated likelihood estimators. This section lists asymptotic distributions of simulated likelihood estimators of θ_0 under H_n. Corresponding results under H_0 can be obtained by setting Δ to zero. The regularity conditions in Section 1 of Appendix A are assumed for all the following propositions and lemmas.

LEMMA B.2.1. *Under* H_n, $S_{n,m}(\theta_n) \xrightarrow{D} N(c\mu,\mathcal{I})$.

Proof. By Proposition 1, $S_{n,m}(\theta_n) = S_n(\theta_n) + n^{1/2}m^{-1}\mu(\theta_n) + P_{n,m}(\theta_n) + [Q_{n,m}(\theta_n) - E(Q_{n,m}(\theta_n))] + R_{n,m}(\theta_n)$. Because $P_{n,m}(\theta) = O_P(1)$, $Q_{n,m}(\theta) - E(Q_{n,m}(\theta)) = O_P(1)$ and $R_{n,m}(\theta) = O_P(1)$ semiuniformly in θ, it follows that $P_{n,m}(\theta_n) = O_P(1)$, $Q_{n,m}(\theta_n) - E(Q_{n,m}(\theta_n)) = O_P(1)$ and $R_{n,m}(\theta_n) = O_P(1)$ under H_0 as n and m go to infinity and $\lim_{n\to\infty} n^{1/2}/m$ is finite. By Lemma B.1.1, $S_{n,m}(\theta_n) = S_n(\theta_n) + n^{1/2}m^{-1}\mu(\theta_n) + o_P(1)$ under H_n. By the Lindeberg–Levy CLT, $S_n(\theta_n) \xrightarrow{D} N(0,\mathcal{I})$. By the continuity of $\mu(\theta)$, $\lim_{n\to\infty}\mu(\theta_n) = \mu$. Hence $S_{n,m}(\theta_n) \xrightarrow{D} N(c\mu,\mathcal{I})$. ■

PROPOSITION B.2.1. *Under* H_n, *the* SMLE $\hat{\theta}_u$ *has the following asymptotic expansion* $\sqrt{n}(\hat{\theta}_u - \theta_n) = \mathcal{I}^{-1}S_{n,m}(\theta_n) + o_P(1)$. *Consequently, under* H_n, $\sqrt{n}(\hat{\theta}_u - \theta_0) \xrightarrow{D} N(c\mathcal{I}^{-1}\mu + \Delta,\mathcal{I}^{-1})$.

Proof. Because $S_{n,m}(\hat{\theta}_u) = 0$, the mean value theorem implies that $\sqrt{n}(\hat{\theta}_u - \theta_n) = -[H_{n,m}(\bar{\theta}_n)]^{-1}S_{n,m}(\theta_n)$, where $\bar{\theta}_n$ lies between $\hat{\theta}_u$ and θ_n. By Proposition B.1.2, $H_{n,m}(\bar{\theta}_n)$ converges in probability to $-\mathcal{I}$ under H_n. The results follow from Lemma B.2.1. ■

PROPOSITION B.2.2. *Under* H_n, $\sqrt{n}(\hat{\theta}_A - \theta_n) = \mathcal{I}^{-1}[S_{n,m}(\theta_n) - c\mu] + o_P(1) = \mathcal{I}^{-1}S_n(\theta_n) + o_P(1)$, *and* $\sqrt{n}(\hat{\theta}_A - \theta_o) \xrightarrow{D} N(\Delta,\mathcal{I}^{-1})$.

Proof. From the construction of $\hat{\theta}_A$, $\sqrt{n}(\hat{\theta}_A - \hat{\theta}_u) = (n^{1/2}/m)H_{n,m}^{-1}(\hat{\theta}_u)\mu_m(\hat{\theta}_u)$. By Proposition B.1.2, $H_{n,m}^{-1}(\hat{\theta}_u) \xrightarrow{p} -\mathcal{I}^{-1}$ and $\mu_{n,m}(\hat{\theta}_u) \xrightarrow{p} \mu$ under H_n. It follows that $\sqrt{n}(\hat{\theta}_A - \hat{\theta}_u) = -c\mathcal{I}^{-1}\mu + o_P(1)$. Hence $\sqrt{n}(\hat{\theta}_A - \theta_0) = \sqrt{n}(\hat{\theta}_u - \theta_0) - c\mathcal{I}^{-1}\mu + o_P(1)$. The results follow from Proposition B.2.1. ■

For estimation under constraints $r(\theta_0) = 0$, asymptotic distributions of CSMLE $\hat{\theta}_c$ and CBASMLE $\hat{\theta}_{c,A}$ are in the following propositions.

PROPOSITION B.2.3. *Under* H_n, $\sqrt{n}(\hat{\theta}_c - \theta_n) = GS_{n,m}(\theta_n) - P\mathcal{I}\Delta + o_P(1)$, *where* P *is defined in Proposition* 2, *and* $G = \mathcal{I}^{-1} - P$, *and* $\sqrt{n}(\hat{\theta}_c - \theta_0) \xrightarrow{D} N(G[c\mu + \mathcal{I}\Delta],G)$.

Proof. $\hat{\theta}_c$ satisfies the first order conditions of a Lagrange function $(\partial \ln L_{n,m}(\hat{\theta}_c)/\partial\theta) - R'(\hat{\theta}_c)\lambda = 0$ and $r(\hat{\theta}_c) = 0$ where λ is a vector of Lagrange multipliers. By the mean value theorem applied to the first condition at θ_n and the constraint condition at θ_0 and by using $r(\theta_0) = 0$, $S_{n,m}(\theta_n) + H_{n,m}(\bar{\theta}_c)\sqrt{n}(\hat{\theta}_c - \theta_n) - R'(\hat{\theta}_c)\lambda/\sqrt{n} = 0$ and $R(\tilde{\theta}_c)(\hat{\theta}_c - \theta_0) = 0$, where $\bar{\theta}_c$ lies between $\hat{\theta}_c$ and θ_n and $\tilde{\theta}_n$ lies between $\hat{\theta}_c$ and θ_0. The $\sqrt{n}(\hat{\theta}_c - \theta_n)$ and $\lambda/\sqrt{n}$ can be derived from this equation system in terms of $R(\hat{\theta}_c)$, $R(\tilde{\theta}_c)$, $H_{n,m}(\bar{\theta}_c)$, and $\theta_n - \theta_0$ by an inverse formula for a partitioned matrix. Because $\hat{\theta}_c$ is consistent under H_0, $\bar{\theta}_c$, and $\tilde{\theta}_c$ converge also to θ_0. By Proposition B.1.2, it follows that $R(\hat{\theta}_c)$ and $R(\tilde{\theta}_c)$ converge in probability to R under H_n. The stochastic expansion in this proposition follows. The asymptotic distribution of $\sqrt{n}(\hat{\theta}_c - \theta_0)$ is a consequence of this expansion by using Lemma B.2.1 and $G\mathcal{I}G = G$. ■

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

LEMMA B.2.2. *The following matrix identity holds for the equivalence of the constraint formulation* $r(\theta) = 0$ *and the freedom equation formulation* $\theta = w(\alpha)$: $W(W'\mathcal{I}W)^{-1}W' = \mathcal{I}^{-1} - \mathcal{I}^{-1}R'(R\mathcal{I}^{-1}R')^{-1}R\mathcal{I}^{-1}$, *where* $W = \partial w(\alpha_0)/\partial\alpha'$ *and* $R = \partial r(\theta_0)/\partial\theta'$.

Proof. It is sufficient to show the identity in terms of projection operators:

$$I - \mathcal{I}^{1/2}W(W'\mathcal{I}W)^{-1}W'\mathcal{I}^{1/2} = \mathcal{I}^{-1/2}R'(R\mathcal{I}^{-1}R')^{-1}R\mathcal{I}^{-1/2}.$$

Define a matrix C as

$$C = [\mathcal{I}^{-1/2}R', \mathcal{I}^{1/2}W]\begin{pmatrix}(R\mathcal{I}^{-1}R')^{-1/2} & 0\\ 0 & (W'\mathcal{I}W)^{-1/2}\end{pmatrix}.$$

Because $r(w(\alpha)) = 0$, $RW = 0$. Therefore, $C'C = I$. Because C has full rank, C' is the inverse of C. Hence $CC' = I$. The matrix identity holds because, explicitly, $CC' = \mathcal{I}^{-1/2}R'(R\mathcal{I}^{-1}R')^{-1}R\mathcal{I}^{-1/2} + \mathcal{I}^{1/2}W(W'\mathcal{I}W)^{-1}W'\mathcal{I}^{1/2}$. ■

PROPOSITION B.2.4. *Under* H_n, $\sqrt{n}(\hat{\theta}_{c,A} - \theta_n) = G(S_{n,m}(\theta_n) - c\mu) - P\mathcal{I}\Delta + o_P(1) = GS_n(\theta_n) - P\mathcal{I}\Delta + o_P(1)$, *and* $\sqrt{n}(\hat{\theta}_{c,A} - \theta_0) \xrightarrow{D} N(G\mathcal{I}\Delta, G)$.

Proof. Because $\hat{\theta}_{c,A} = w(\hat{\alpha}_A)$ and $\hat{\theta}_c = w(\hat{\alpha})$, it follows that $\hat{\theta}_{c,A} = \hat{\theta}_c + [w(\hat{\alpha}_A) - w(\hat{\alpha})] = \hat{\theta}_c + (\partial w(\bar{\alpha})/\partial\alpha')(\hat{\alpha}_A - \hat{\alpha})$. As $\hat{\alpha}_A - \hat{\alpha} = H_{\alpha,n,m}(\hat{\alpha})\mu_{\alpha,n,m}(\hat{\alpha})/m$, it follows that $\sqrt{n}(\hat{\theta}_{c,A} - \theta_n) = \sqrt{n}(\hat{\theta}_c - \theta_n) + (\sqrt{n}/m)(\partial w(\bar{\alpha})/\partial\alpha')H^{-1}_{\alpha,n,m}(\hat{\alpha})\mu_{\alpha,n,m}(\hat{\alpha})$. Because $\mu_{\alpha,n,m}(\alpha) = (\partial w'(\alpha)/\partial\alpha)\mu_{n,m}(\theta)$, $\mu_{\alpha,n,m}(\hat{\alpha}) \xrightarrow{p} W\mu$. Similarly, $H_{\alpha,n,m}(\hat{\alpha}) \xrightarrow{p} W'\mathcal{I}W$. Therefore, $\sqrt{n}(\hat{\theta}_{c,A} - \theta_n) = \sqrt{n}(\hat{\theta}_c - \theta_n) - cW(W'\mathcal{I}W)^{-1}W'\mu + o_P(1) = \sqrt{n}(\hat{\theta}_c - \theta_n) - cG\mu + o_P(1)$, where the last equality follows from the matrix identity in Lemma B.2.2. The asymptotic expansion of $\sqrt{n}(\hat{\theta}_{c,A} - \theta_n)$ follows from the asymptotic expansion of $\sqrt{n}(\hat{\theta}_c - \theta_n)$ in Proposition B.2.3. Because $\sqrt{n}(\hat{\theta}_{c,A} - \theta_n) = \sqrt{n}(\hat{\theta}_{c,A} - \theta_0) + \Delta$, $\sqrt{n}(\hat{\theta}_{c,A} - \theta_0) = GS_n(\theta_n) + G\mathcal{I}\Delta + o_P(1)$. Because $S_n(\theta_n)$ is asymptotically $N(0,\mathcal{I})$ under H_n from Lemma B.2.1, the asymptotic distribution of $\sqrt{n}(\hat{\theta}_{c,A} - \theta_0)$ is $N(G\mathcal{I}\Delta, G)$. ■

PROPOSITION B.2.5. *Under* H_n, $\sqrt{n}(\hat{\theta}_u - \hat{\theta}_c) = P(S_{n,m}(\theta_n) + \mathcal{I}\Delta) + o_P(1)$, *and* $\sqrt{n}(\hat{\theta}_u - \hat{\theta}_c) \xrightarrow{D} N(P(c\mu + \mathcal{I}\Delta), P)$.

Proof. This follows from Propositions B.2.1 and B.2.3. The second part of the results follows from Lemma B.2.1 and $P'\mathcal{I}P = P$. ■

PROPOSITION B.2.6. *Under* H_n, $\sqrt{n}(\hat{\theta}_A - \hat{\theta}_{c,A}) = P(S_{n,m}(\theta_n) + \mathcal{I}\Delta - c\mu) + o_P(1) = P(S_n(\theta_n) + \mathcal{I}\Delta) + o_P(1)$, *and* $\sqrt{n}(\hat{\theta}_A - \hat{\theta}_{c,A}) \xrightarrow{D} N(P\mathcal{I}\Delta, P)$.

Proof. This follows from Propositions B.2.2 and B.2.4. ■

PROPOSITION B.2.7. *Under* H_n, $(1/\sqrt{n})(\partial \ln L_{n,m}(\hat{\theta}_c)/\partial\theta) = R'(R\mathcal{I}^{-1}R')^{-1} \times R\mathcal{I}^{-1}(S_{n,m}(\theta_n) + \mathcal{I}\Delta) + o_P(1)$.

Proof. By the mean value theorem, $(1/\sqrt{n})(\partial \ln L_{n,m}(\hat{\theta}_c)/\partial\theta) = S_{n,m}(\theta_n) + H_{n,m}(\bar{\theta})\sqrt{n}(\hat{\theta}_c - \theta_n) = S_{n,m}(\theta_n) - \mathcal{I}\sqrt{n}(\hat{\theta}_c - \theta_n) + o_P(1)$. Proposition B.2.3 implies the expansion. ■

PROPOSITION B.2.8. *Under* H_n, $(1/\sqrt{n})(\partial \ln L_{n,m}(\hat{\theta}_{c,A})/\partial\theta) - c\mu = R'(R\mathcal{I}^{-1}R')^{-1} \times R\mathcal{I}^{-1}(S_{n,m}(\theta_n) + \mathcal{I}\Delta - c\mu) + o_P(1) = R'(R\mathcal{I}^{-1}R')^{-1}R\mathcal{I}^{-1}(S_n(\theta_n) + \mathcal{I}\Delta) + o_P(1)$.

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

Proof. By the mean value theorem, $(1/\sqrt{n})(\partial \ln L_{n,m}(\hat{\theta}_{c,A})/\partial\theta) = S_{n,m}(\theta_n) + H_{n,m}(\bar{\theta})\sqrt{n}(\hat{\theta}_{c,A} - \theta_n) = S_{n,m}(\theta_n) - \mathcal{I}\sqrt{n}(\hat{\theta}_{c,A} - \theta_n) + o_P(1)$. With the expansion in Proposition B.2.4,

$$\frac{1}{\sqrt{n}}\frac{\partial \ln L_{n,m}(\hat{\theta}_{c,A})}{\partial\theta} = S_{n,m}(\theta_n) - \mathcal{I}G[S_{n,m}(\theta_n) - c\mu] + \mathcal{I}P\mathcal{I}\Delta + o_P(1)$$

and the result follows as $I - \mathcal{I}G = \mathcal{I}P = R'(R\mathcal{I}^{-1}R')^{-1}R\mathcal{I}^{-1}$. ■

3. *Proofs of Propositions in Text*

Proof of Proposition 3. Consider first the SLR statistic. By a Taylor expansion and Proposition B.2.5, $2[\ln L_{n,m}(\hat{\theta}_u) - \ln L_{n,m}(\hat{\theta}_c)] = -n(\hat{\theta}_u - \hat{\theta}_c)'H_{n,m}(\bar{\theta})(\hat{\theta}_u - \hat{\theta}_c) = (S_{n,m}(\theta_n) + \mathcal{I}\Delta)'P'\mathcal{I}P(S_{n,m}(\theta_n) + \mathcal{I}\Delta) + o_P(1) = (S_{n,m}(\theta_n) + \mathcal{I}\Delta)'P(S_{n,m}(\theta_n) + \mathcal{I}\Delta) + o_P(1)$, because $P'\mathcal{I}P = P$. Lemma B.2.1 implies that, under H_n, $\mathcal{I}^{-1/2}(S_{n,m}(\theta_n) + \mathcal{I}\Delta) \to N(\mathcal{I}^{-1/2}(c\mu + \mathcal{I}\Delta), I)$. The matrix Q where $Q = \mathcal{I}^{1/2}P\mathcal{I}^{1/2} = \mathcal{I}^{-1/2}R'(R\mathcal{I}^{-1}R')^{-1} \times R\mathcal{I}^{-1/2}$ is an idempotent matrix with rank ν. Hence $2[\ln L_{n,m}(\hat{\theta}_u) - \ln L_{n,m}(\hat{\theta}_c)] = [\mathcal{I}^{-1/2}(S_{n,m}(\theta_n) + \mathcal{I}\Delta)]'Q[\mathcal{I}^{-1/2}(S_{n,m}(\theta_n) + \mathcal{I}\Delta)] + o_P(1)$ is asymptotically $\chi^2_\nu((c\mu + \mathcal{I}\Delta)'P(c\mu + \mathcal{I}\Delta))$. For the SW statistic, expand $r(\hat{\theta}_u)$ at θ_0 by a Taylor expansion. Because $r(\theta_0) = 0$, $r(\hat{\theta}_u) = R(\bar{\theta})(\hat{\theta}_u - \theta_0) = R(\bar{\theta})[\hat{\theta}_u - \theta_n + \Delta/\sqrt{n}]$. Therefore, $\sqrt{n}r(\hat{\theta}_u) = R(\bar{\theta})[\sqrt{n}(\hat{\theta}_u - \theta_n) + \Delta] = R\mathcal{I}^{-1}[S_{n,m}(\theta_n) + \mathcal{I}\Delta] + o_P(1)$ from Proposition B.2.1. Hence $nr'(\hat{\theta}_u)[R\mathcal{I}^{-1}R']^{-1}r(\hat{\theta}_u) = [S_{n,m}(\theta_n) + \mathcal{I}\Delta]'P[S_{n,m}(\theta_n) + \mathcal{I}\Delta] + o_P(1)$, which has the same limiting distribution as the SLR test statistic. The feasible SW statistic has the same limiting distribution by the Slutsky lemma. The SSC test follows from Proposition B.2.7. ■

Proof of Proposition 2. The H_0 case can be regarded as a special case of H_n with $\Delta = 0$. The results in Proposition 2 follow from Proposition 3 by setting Δ in relevant expressions. ■

Proof of Proposition 4. Consider the bias-corrected SW statistic. By a Taylor expansion of $r(\hat{\theta}_A)$ at θ_0, because $r(\theta_0) = 0$, $\sqrt{n}r(\hat{\theta}_A) = R\sqrt{n}(\hat{\theta}_A - \theta_0) + o_P(1) = R[\sqrt{n}(\hat{\theta}_A - \theta_n) + \Delta] + o_P(1) = R\mathcal{I}^{-1/2}[\mathcal{I}^{-1/2}S_n(\theta_n) + \mathcal{I}^{1/2}\Delta] + o_P(1)$ by Proposition B.2.2. Hence

$$nr'(\hat{\theta}_A)[R\mathcal{I}^{-1}R']^{-1}r(\hat{\theta}_A) = [\mathcal{I}^{-1/2}S_n(\theta_n) + \mathcal{I}^{1/2}\Delta]'Q[\mathcal{I}^{-1/2}S_n(\theta_n) + \mathcal{I}^{1/2}\Delta] + o_P(1).$$

Because $\mathcal{I}^{-1/2}S_n(\theta_n) \xrightarrow{D} N(0, I)$ under H_n, the $nr'(\hat{\theta}_A)[R\mathcal{I}^{-1}R']^{-1}r(\hat{\theta}_A)$ is asymptotically $\chi^2_\nu(\delta_2)$ with $\delta_2 = \Delta'\mathcal{I}^{1/2}Q\mathcal{I}^{1/2}\Delta = \Delta'R'(R\mathcal{I}^{-1}R')^{-1}R\Delta$. For H_0, this statistic is asymptotically χ^2_ν. The feasible bias-corrected SW statistic has a similar limiting distribution by the Slutsky lemma. For the bias-corrected SSC statistic, Proposition B.2.8 implies that $[(1/\sqrt{n})(\partial \ln L_{n,m}(\hat{\theta}_{c,A})/\partial\theta) - c\mu]'\mathcal{I}^{-1}[(1/\sqrt{n})(\partial \ln L_{n,m}(\hat{\theta}_{c,A})/\partial\theta) - c\mu] = (S_n(\theta_n) + \mathcal{I}\Delta)'\mathcal{I}^{-1/2}Q\mathcal{I}^{-1/2}(S_n(\theta_n) + \mathcal{I}\Delta) + o_P(1)$, which has the same limiting distribution as the bias-corrected SW statistic. ■

Proof of Propositions 5 and 6. The modified SW statistic W_r^{sl} in (5.4) has the same limiting distribution of the SW statistic by the Slutsky lemma because $nV_{r,n,m}$ converges in probability to $\mathcal{I}^{-1}$. For the modified SSC statistic in (5.4), Proposition B.2.7 implies that

$$\begin{aligned} S_r^{sl} &= (S_{n,m}(\theta_n) + \mathcal{I}\Delta)'\mathcal{I}^{-1}R'(RnV_{r,n,m}R')^{-1}R\mathcal{I}^{-1}(S_{n,m}(\theta_n) + \mathcal{I}\Delta) + o_P(1) \\ &= (S_{n,m}(\theta_n) + \mathcal{I}\Delta)'P(S_{n,m}(\theta_n) + \mathcal{I}\Delta) + o_P(1) \end{aligned}$$

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press

under H_n, which has the same limiting distribution as the SSC statistic. For the LR_r^{sl}, Proposition B.2.6 implies that

$$n(\hat{\theta}_u - \hat{\theta}_c)'\{H_{n,m}(\hat{\theta}) + R'[RnV_{r,n,m}R']^{-1}R\}(\hat{\theta}_u - \hat{\theta}_c)$$
$$= (S_{n,m}(\theta_n) + \mathcal{I}\Delta)'P\{-\mathcal{I} + R'[R\mathcal{I}^{-1}R']^{-1}R\}P(S_{n,m}(\theta_n) + \mathcal{I}\Delta) + o_P(1).$$

Because $P\mathcal{I}P = P$ and $PR'[R\mathcal{I}^{-1}R']^{-1}RP = P$, the preceding term converges in probability to zero. Therefore, the modified SSC statistic has the same limiting distribution as the SLR statistics. The same arguments can be applied to the modified bias-corrected statistics W_r^{bc} and S_r^{bc}. ■

https://doi.org/10.1017/S0266466699153039 Published online by Cambridge University Press