UNBIASED SIMULATORS FOR ANALYTIC FUNCTIONS AND MAXIMUM UNBIASED SIMULATED LIKELIHOOD ESTIMATION

BY GREGORY M. DUNCAN[1]

The paper collects some old results and presents some news ones for unbiased estimation of analytic functions of probabilities where the probabilities must be simulated and applies these to Simulated Maximum Likelihood (SML) estimation. The results include unbiased estimation of finite degree polynomials and other analytic functions, unbiased simulation of the score and likelihood, and the asymptotic properties of SML using these simulators. The motivating application is estimation in the mixed logit model. The old results are spread throughout the literature, primarily in the non-parametric and sequential estimation literatures. Moreover, they do not seem known to simulation researchers, much less applied. So they are collected here and presented, in context, with the new results.

0. INTRODUCTION

ONE OBSTACLE to estimation by Simulated Maximum Likelihood (SML) has been the apparent impossibility of finding unbiased estimators of the log-likelihood and score when the likelihood involves expectations that must be simulated. This is because non-linear functions of averages are usually not unbiased estimates of the function applied to the expectation of the average, that is, $E\left(f\left(\bar{X}\right)\right) \neq f\left(E\left(\bar{X}\right)\right)$ unless *f* is linear (affine).[2] In the first part of this paper, we remove this obstacle when the function *f* is analytic; we do it by finding another function $f^*\left(X_1, \ldots, X_I\right)$ having the property that $E\left(f^*\left(X_1, \ldots, X_I\right)\right) = f\left(E\left(X_1\right)\right)$. Specifically, we develop unbiased simulators of analytic functions of expectations where the expectations themselves must be simulated, regardless the function *f* except it be analytic and regardless the simulation distribution. In the second part, we show the consistency and asymptotic normality of Simulated

[1] Thanks go to members of the Econometrics Workshops at the University of California, Berkeley and Washington State University. Special thanks go to Scott Cardell, Ron Mittlehammer, Jim Powell, and Paul Ruud.
[2] See Gouriéroux and Monfort (1996) or Lee (1996), the problem is also mentioned in Hajivassiliou and McFadden (1998).

Maximum Likelihood estimators based on these simulators.

Prominent examples of the problem occur in estimating the log-likelihood and the score in mixed logit or multinomial probit discrete choice models. There the probabilities are simulated and substituted into the likelihood function as though they were exact, ignoring the bias and placing a, perhaps, unwarranted reliance on asymptotic results in small sample/trial situations. For though the approach is consistent, many practitioners have needed tens of thousands of simulations for each observation to obtain good estimates of the economic parameters of interest. This paper presents a very general solution to the problem.

Here are the major results: First, and preliminarily, if $l(p) = \sum_{i=0}^{I} l_i (p - p_0)^i$ is a polynomial of degree I and $\{U_1, \ldots, U_I\}$ is a sequence of random variables satisfying $E(U_i) = (p - p_0)^i$, then $\sum_{i=0}^{I} l_i U_i$ is an unbiased estimator of $l(p)$.[3] Second, if $l(p) = \sum_{i=0}^{\infty} l_i (p - p_0)^i$ is an analytic function with a non-terminating expansion and circle of convergence $C(R) = \{p \mid |p - p_0| < R\}$, I is a discrete non-negative random variable with finite expectation with a known survivor function $G(i) = \Pr[I > i] = 1 - F(i)$[4] and $\{U_1, \ldots, U_I\}$ is a sequence of random variables satisfying $E(U_i) = (p - p_0)^i$, then the random degree polynomial, $\sum_{i=0}^{I} l_i U_i / G(i)$ is an unbiased estimator of $l(p)$. Third, if $p = p(J)$ depends on unknown parameters J, $s = S(z, J)$, and $E(s) = p - p_0$ then under various differentiability and smoothness conditions the various estimators of the score are unbiased and stochastically equicontinuous: the gradient, $\sum_{i=0}^{I} l_i (\nabla_J U_i) / G(i)$, for differentiable simulators; the numeric gradient, for all simulators; and, where the simulation density is log-differentiable, the indirect score defined by the function $\%(z, I) \sum_{i=1}^{I} \left(\nabla_J \ln\left(h(z_i, J)\right)\right)$

---

[3] U_i might be the U-statistic for estimating p-p_0

[4] For example, a Geometric random variable with known parameter w, and survivor function $\Pr[I > i] = (1 - w)^i$

where z has density $h(z,J)$, and $\tilde{l}(z,I)$ is an unbiased estimator of $l\left(p(J)\right)$. Finally, the Simulated Maximum-Likelihood (SML) estimator using the aforementioned scores and functions are consistent and asymptotically normal.

In Part One, the first section treats polynomials which are analytic functions with terminating expansions,. While approximate unbiasedness is not our goal, approximately unbiased estimates of any analytic functions can be obtained by finitely and non-randomly truncating infinite series representation or by using minimax polynomial approximations. We investigate approximations in preparation for the fully unbiased results of the second section. They may also turn out to be practically useful, as they require little computation and may be sufficiently accurate. For the log probability, minimax polynomial approximations, with relative errors smaller than 5%, are found that require no more than 16 trials when the underlying probabilities are larger than .001.

The second section considers unbiased estimation of analytic functions with non-terminating expansions. These estimators do not require approximation but are somewhat more complex than their finite expansion counterparts. Interestingly, these estimators do utilize a truncation, but one done randomly and in such a way as to leave the estimator completely unbiased.

Part Two formalizes the aforementioned ways of estimating the gradient of the log-likelihood, we show unbiasedness of the score estimators, and then the stochastic equicontinuity of the implied simulation residual processes. Using the framework of Hajivassiliou and McFadden (1998) we then show the consistency and asymptotic normality of maximum likelihood estimators with the simulated scores.

1. UNBIASED SIMULATORS FOR ANALYTIC FUNCTIONS

Finding unbiased estimators of the log-likelihood and its gradient when the probabilities are simulated mixed logits or multinomial probits motivated this research and the conditions of this paper are developed with those in mind.[5] Following begins the running example that illustrates of the ideas developed herein.

[5] This results is some loss of generality. Particularly, as I deal primarily with simulated probabilities, I will not address the issue that the logarithmic expansion diverges outside –1 and 1.

1.1. Example: Bernoulli and General Simulants in the Random Coefficient Logit with Sign Constraints[6]

Consider a standard logit model with J alternatives where the coefficients are continuous functions of multivariate normals. Let

$$U = W^T g(Z) + e$$
$$Z : \ Gaussian(m, \Omega)$$

and define $J = \begin{pmatrix} m \\ vech(\Omega) \end{pmatrix}$.[7] U is an unobserved vector of utilities. The i[th] row of W is the vector of characteristics for alternative i. The e is a vector of independent and identically distributed Extreme Value (0,1) errors is statistically independent of Z. $g(Z)$ is a vector valued function that depends continuously on Z; this allows imposition of sign restrictions on the coefficients, i.e. $-\exp(Z_k)$ is always negative. Each alternative has a utility U_i and alternative i is chosen if $U_i > U_j \ \ \forall i \neq j$. Without loss of generality, label alternatives so that alternative 1 is the one chosen.

A Bernoulli simulant d is defined as follows let

$$A = \{(e, Z) | U_1 > U_j, \ j = 2, \ldots, J\}.$$

and draw (e, Z) according with the distributions above, with fixed J, and set

(1) $d = 1$ if $(e, Z) \in A$.
$= 0$ otherwise

Then with the probability

(2) $$\Pr[d = 1 | W] = p(J) = \int_{R^p} \left\{ \exp W_1^T g(Z) / \sum_{j=1}^{J} \exp W_j^T g(Z) \right\} f(Z; m, \Omega) dZ$$

The simulant, d, is not differentiable in J.

For a general simulant, *s*, draw *Z*, only, according with the distribution above with fixed J, and define

[6] See Cardell and Dunbar (1980) or Train (2002))
[7] See Ruud(2000) for a complete definition of the vech operator; basically, it is the lower triangle of a matrix ordered lexicographically in a vector.

(3) $s(Z) = \exp W_1^T g\,(Z) / \sum_{j=1}^{J} \exp W_j^T g\,(Z)$

In this case, too, $E(s) = p(J)$, however, this simulant does not depend explicitly on J. However, it can often be written in a second form where J appears. Let $Z \sim N(0, I)$ and $K = \Omega^{1/2}$ is upper triangular then

(4) $s = \exp W_1^T g\,(m + KZ) / \sum_{j=1}^{J} \exp W_j^T g\,(m + KZ)$

1.2. Unbiased Polynomials in the Bernoulli Case

The Bernoulli simulator of a probability $p = \int_{z \in A} h(z; J)\,dz$ is the proportion of times random variables *Z* drawn independently from *h* fall in A. Let

$$d = 1 \text{ if } x \in A$$
$$\quad = 0 \text{ otherwise},$$

then $B = \sum_{i=1}^{I} d_i$, the number of times that *Z* falls in A, is a *Binomial*(*I*,*p*) random variable. Generally, the expectation $m'_{[i]} = E\big(B(B-1)\ldots(B-i+1)\big) = E\big((B)_i\big)$[8] is called the ith factorial moment of *B*. For a *Binomial*(*I*,*p*), $m'_{[r]} = p^r (I)_r$, $r \le I$ or, more usefully, $E\big((B)_r / (I)_r\big) = m'_{[r]} / (I)_r = p^r$, $r \le I$. Let $l(p) = \sum_{i=0}^{I} l_i p^i$ be a degree *I* polynomial and $\tilde{l}(B) = l_0 + \sum_{i=0}^{I} l_i \big((B)_i / (I)_i\big)$ be its estimator. This equation is nothing more than substituting $(B)_i / (I)_i$ for each power of *p*. Note if *B* is less than *j*, the subsequent terms are all zero so we can redefine $\tilde{l}$ making *B* the upper limit of indices in the sum. Our estimator becomes $\tilde{l}(B) = l_0 + \sum_{i=0}^{B} l_i (B)_i / (I)_i$. And so,

PROPOSITION 1: *Let B be Binomial(I,p), then* $E\big(\tilde{l}(B)\big) = \sum_{i=0}^{I} l_i p^i$.

---

[8] We shall use the notation $(B)_r = B(B-1)\ldots(B-r+1)$ for the falling factorial symbol.

PROOF: From the theory of the binomial (Johnson, Kotz, Kemp (1993)), $m'_{[r]} = p^r \left(I\right)_r$, $r \le I$. Substituting gives the result.

Q.E.D.

By the Lehmann-Scheffe` Theorem, these estimators are efficient because they are unbiased estimators of the polynomials and are functions of the complete sufficient statistic *Z*. To increase efficiency, we simply increase *I*.

1.3. A First Application to Logarithms

In this section we develop an approximate method for estimating the log-probability or likelihood. It will serve as an introduction to the expansion method required for exactly unbiased methods developed below. The logarithm has the analytic expansion

(5) $$\ln\left(p\right) = \ln\left(p_0\right) + \sum\nolimits_{i=0}^{\infty}\left(-1\right)^i\left(p - p_0\right)^i / \left(ip_0^i\right) = \ln\left(p_0\right) - \sum\nolimits_{i=0}^{\infty}\left(p_0 - p\right)^i / \left(ip_0^i\right)$$

with circle of convergence $C = \left\{p \,|\, 0 < 1 \le 2\,p_0\right\}$[9]. For p moderate small to unity, the series converges quite rapidly. If we truncate the series at *I* terms, we have a finite polynomial to which Theorem 1 applies. Its bias is a completely non-statistical approximation error dictated solely by the degree of the polynomial approximation. Let *B~Binomial(I,1-p)* then the truncated estimator of log(p) is given by $\tilde{l}\left(B\right) = -\sum\nolimits_{i=0}^{B}\left(\left(B\right)_i / i\left(I\right)_i\right)$.

COROLLARY 1: *Let B be Binomial(I ,1-p), then* $\tilde{l}\left(B\right)$ *is an unbiased estimate of the truncated logarithm* $-\sum\nolimits_{j=0}^{B}\left(1 - p\right)^j / j$.

The bias in estimating the logarithm using the truncated logarithm depends on both the degree *I* and the probability *p*. The bias decreases as I increases or as p decreases. The user can pick I whereas p is unknown. For probabilities larger than .25, simply

---

[9] The expansion converges at one endpoint, $p=2p_0$ but not the other.

truncating the polynomial expansion at 10 terms works well; but smaller probabilities require ever increasing numbers of terms and trials, around 30000 trials are required for probabilities around .0001 to obtain relative errors less than .001.

One can greatly reduce the required number of moments and therefore observations by finding better fitting polynomials, ones whose coefficients are not the Taylor coefficients; for example a minimax approximation to log(*p*) that minimized the relative error over p in [.001, .999] reduced the number of trials needed from up to 30000 to no more than 16 with a relative error of approximately 4.8% .[10] If the *p* is outside this range, the approximation is defined but may have larger relative errors. Larger ranges lead to larger errors and eventually a complete breakdown of the minimax algorithm. A minimax approximation is easily computed using Mathematica. For the truncated series approximations, the error increases as the probability decreases; for the minimax, the worst relative errors are spread periodically over the domain. The approximately five percent relative error is about the best that can be done with a polynomial that is numerically stable.

1.4. Unbiased Analytic Functions of Probabilities in the Bernoulli Case

In the Bernoulli case, to obtain unbiased estimates for any function $\mathrm{l}(p)$ analytic on (0,1], we use an inverse binomial sampling scheme and work with a slightly different expansion than we developed for the polynomial. Referring to the logarithm example, the problem with fixed sample estimators for analytic functions with non-terminating series representations is that they require either an infinite number of observations or acceptance of an approximation error that depends on the unknown probability.

For Bernoulli simulators, we handle the problem with the following change. Instead of taking a fixed sample, we sample until the first success is recorded. The random variable *I* that is the number of trials before the success occurs then has a *Geometric*(*p*) distribution. To derive our estimator, let us first assume that an unbiased estimator $\tilde{\mathrm{l}}(I)$

[10] See Judd (2001) p212 for example.

of $l(p)$ exists. If so, by definition of unbiasedness

(6) $E\left(\hat{l}(I)\right) = l(p)$

while by definition of an expectation of a function of a *Geometric*(*p*) random variable,

(7) $E\left(\hat{l}(I)\right) = \sum_{i=0}^{\infty} \hat{l}(i)\, p(1-p)^{i}$

If an unbiased estimator exists the right hand sides of (6) and (7) must be equal. The same is true if we divide both sides of both equations by p, so we can write

(8) $E\left(\hat{l}(I)\right)/p = \sum_{i=0}^{\infty} l(i)(1-p)^{i} = \sum_{i=0}^{\infty} (-1)^{i}\, l(i)(p-1)^{i}$.[11]

If the series expansion around p=1 of

(9) $l(p)/p = \sum_{i=0}^{\infty} b_i (p-1)^{i}$

then equating the coefficients of like terms in (8) and (9) gives $l(n) = (-1)^{n} b_n$ as and unbiased estimator of $l(p)$. This is a result from the sequential analysis literature due to DeGroot (1959). It seems unknown in the simulation literature.

1.5. A Second Application to Logarithms

The following result is new, but an application of the previous discussion.

PROPOSITION 2: *Let I be a Geometric(p) random variable, then the Nth term of the recursion*

$$
\begin{aligned}
l(i+1) &= (-1)^{i+1}\, l(i) + (-1)^{2i+1}\, i & \quad i > 1 \\
&= -1 & \quad i = 1 \\
&= 0 & \quad i = 0
\end{aligned}
$$

is an unbiased estimator of $\ln(p)$.

PROOF: Expanding $\ln(p)/p$ around *p*=1 gives a

(10) $f(p)/p = \sum_{i=1}^{\infty} (a_i / i!)(p-1)^{i}$.

[11] If $l(p)$ is analytic on (0,1] then so is $l(p)/p$ see Krantz and Parks(1991).

By induction it may be shown that $a_{i+1} = -(i+1)a_i + (-1)^i i!$. Assume

$$\frac{d^{n+1}\left(\ln(p)/p\right)}{dp^{n+1}} = a_{n+1}\frac{1}{p^{n+2}} + b_{n+1}\frac{\ln(p)}{p^{n+2}}$$

is true for all n. Differentiating gives

(11) $$\frac{d^n\left(a_n\frac{1}{p^{n+1}} + b_n\frac{\ln(p)}{p^{n+1}}\right)}{dp^n} = \left(-(n+1)\right)a_n\frac{1}{p^{n+2}} + \left(-(n+1)\right)b_n\frac{\ln(p)}{p^{n+2}} + b_n\frac{1}{p^{n+2}}$$
$$= \left(\left(-(n+1)a_n\right) + b_n\right)\frac{1}{p^{n+2}} + \left(-(n+1)\right)b_n\frac{\ln(p)}{p^{n+2}}$$

For n=1 we have

(12) $$\frac{d\left(\ln(p)/p\right)}{dp} = \frac{1}{p^2} + (-1)\frac{\ln(p)}{p^2}$$

So $a_1 = 1$ and $b_1 = -1$, equating coefficients gives the rest of the recursion

$$a_{n+1} = \left(\left(-(n+1)a_n\right) + b_n\right)$$
(13) $$b_{n+1} = \left(-(n+1)\right)b_n$$

$$\begin{aligned} b_{i+1} &= -\frac{a_{i+1}}{(i+1)!} \\ &= -\frac{(i+1)}{(i+1)!}a_i + (-1)^i\frac{i!}{(i+1)!} \\ &= -\frac{a_i}{i!} + (-1)^i i \\ &= -b_i + (-1)^i i \end{aligned}$$

so

$$\begin{aligned} l(i+1) &= (-1)^{i+1} b_{i+1} \\ &= (-1)^{i+1}(-1)^i b_i + (-1)^{2i+1} i \end{aligned}$$
(14) $$= (-1)^{i+1} l(i) + (-1)^{2i+1} i$$

with $l(1) = -1$.

Q.E.D.

1.6. Unbiased Polynomials in the General Simulator Case

More generally, simulators have the form $s = S(Z, \vartheta)$ where *Z*, perhaps a vector, has density function $h(z;\vartheta)$ and where $p(\vartheta) = \int (S(z,\vartheta)) h(z;\vartheta) dz$. Neither term in the integrand need be differentiable, though both can be, similarly, neither term need depend explicitly on ϑ although at least one must.[12] Cases of such simulators can be found in Genz (1992), Hajivassiliou and McFadden (1998) and include the mixed logit (Train (2002)) (a.k.a. random coefficients logit).

Let $E(s_i) = p - p_0$, where $\{s_1, \ldots s_I\}$ be independent and identically distributed simulators. The uniformly minimum variance unbiased U-statistic for estimating the parameters $\gamma_j = (E(s_1))^j \;\; j \le I$ is given by

(15) $$U_i(s,I) = \sum_{r_1 < \cdots < r_i} (s_{r_1} \cdots s_{r_i}) / C(I,i).$$ [13]

This suggests using $\tilde{\gamma}(s,I) = \lambda_0 + \sum_{i=0}^{I} \lambda_i U_i(s,I)$ to estimate polynomials of degree *I*. We state without proof the following proposition.

PROPOSITION 3**:** *Let* $\tilde{\gamma}(s,I) = \sum_{i=0}^{I} \lambda_i U_i(s,I)$ *then* $E(\tilde{\gamma}(s,I)) = \sum_{i=0}^{I} \lambda_i (p - p_0)^i$.

1.7. A Third Application to Logarithms

To estimate the logarithm, ln(*p*), define $t_i = p_0 - s_i$. Again we state without proof the obvious proposition.

---

[12] We will only treat the polar cases where only one of h or S depends on ϑ.

[13] $C(I,i)$ is the binomial coefficient "I choose i".

PROPOSITION 4: *Let* $\hat{l}(t,I)=\ln(p_0)-\sum_{i=0}^{I}\left(U_i(t,I)\big/\left(ip_0^i\right)\right)$ *then* $E\left(-\hat{l}(t,I)\right)=-\sum_{i=0}^{I}\left((1-p)^i\big/i\right)=\ln(p)+e(p,I)$ *where* $e(p,I)$ *is the error of terminating the polynomial expansions at I terms.*

As in Section 1, minimax approximations can be used instead of truncating the expansion. The bias properties and values are identical, as the source of the bias is a non-statistical truncation.

1.8. Unbiased Analytic Functions for General Simulators

In section 1, we showed that truncating a series expansion at a fixed non-random degree gives an unbiased estimator of the truncated polynomial. In section 2, we found if we truncated at a random degree, the polynomial estimated could be of infinite degree. What made the random truncation method work for the Bernoulli simulator was not so much changing to a new distribution, the Geometric, as it was the fact that the range of the Geometric random variable used to truncate terms was infinite, allowing the expected value to have an infinite number of terms so one could equate like terms, but having a finite number of terms in the expansion with probability 1. Something very similar works here.

1.9. A First Simple Estimator

Consider the following procedure, choose *I* according with a *Geometric*(*w*) distribution, then choose *I* independent simulants, $s_i \;\; i=1,\ldots,I$, with $E(s_i)=p-p_0$ and let $\hat{l}(s,I)=l(I)\prod_{i=1}^{I}s_i$ be the estimator of $l(p)$. By definition,

(16) $$E\left(\hat{l}(I)\right)=\sum_{i=0}^{\infty}l(i)w(1-w)^i(p-p_0)^i$$

if $l(p)$ is analytic then

(17) $$l(p)=\sum_{i=0}^{\infty}l^{(i)}(p_0)(p-p_0)^i/i!$$

If $\hat{l}(s,I)$ is unbiased then (16) equals (17) and since $l(p)$ is analytic, coefficients of like terms must be equal, hence $l(I)w(1-w)^I = (-1)^I 1^{(I)}(p_0)/I!$ or

$l(I) = (-1)^I 1^{(I)}(p_0)/\left(I!w(1-w)^I\right)$ which suggests the following proposition.

PROPOSITION 5: *If* $1(p)$ *is analytic and* $I \sim Geometric(w)$ *then*

$\tilde{1}(s,I) = (-1)^I 1^{(I)}(p_0)/\left(I!w(1-w)^I\right)\prod_{i=1}^{I} s_i$ *is an unbiased estimator of* $1(p)$.

This estimator seems strange. No outcome is close to $1(p)$ in any intuitive sense. It almost necessarily has a large variance. One simple correction is the following. Take the first *I** terms of the expansion with probability 1. As we discussed above this will have a bias that depends on the unknown p. Then correct the bias by taking on additional term that term will be randomly selected as in Proposition 5. More specifically let I* be fixed and let *I* be *Geometric*(*w*), we have as an estimator

$\tilde{1}(s,I) = \sum_{i=1}^{I^*} l(i)\prod_{i=1}^{i} s_i + l(I^*+I+1)\prod_{i=1}^{I^*+I+1} s_i$. Following the same approach as above we have $E\left(\tilde{1}(s,I)\right) = \sum_{i=1}^{I^*} l(i)\prod_{i=1}^{i} s_i + \sum_{I=0}^{\infty} l(I^*+I+1)\left(\prod_{i=1}^{I^*+I+1} s_i\right)w(1-w)^I$

Then equating coefficients we have

$$l(i) = (-1)^i 1^{(i)}(p_0)/i! \qquad i \le I^*$$
$$= (-1)^{I+I^*+1} 1^{(I+I^*+1)}(p_0)/\left((I+I^*+1)!w(1-w)^I\right) \qquad i = I+I^*+1$$

PROPOSITION 6: *If* $1(p)$ *is analytic, I* is fixed and* $I \sim Geometric(w)$ *then*

$\tilde{1}(s,I) = \sum_{i=1}^{I^*}\left((-1)^i 1^{(i)}(p_0)/i!\right)\prod_{j=1}^{i} s_j + \left[(-1)^{I^*+I+1} 1^{(I^*+I+1)}(p_0)/\left((I^*+I+1)!w(1-w)^I\right)\right]\prod_{i=1}^{I^*+I+1} s_i$ *is an unbiased estimator of* $1(p)$.

While this estimator is clearly more intuitive it suffers from the problem that it does not use all the available information. Specifically, it does not use optimal estimators to estimate the monomials $(p-p_0)^i$, replacing the simple products with the optimal U-statistic for estimating the monomial seem to be an intuitive improvement, similarly, filling in the gap between the *I** term and the *I*+I+1* term seems similarly intuitive. The next estimator does just that. However, since the U-statistics are not independent, it is not necessarily the case the following estimator is better however more intuitive.

1.10. A More Complicated Estimator

For expansions where we choose the degree to be a random variable *I* in [0,) having a finite expectation and having survival function $G(i)$, then, as we will show,

(18) $E\left(\sum_{i=0}^{I} a_i U_i(s, I)\right) = \sum_{i=0}^{\infty} a_i G(i)(p - p_0)^i$.[14]

So to obtain a series whose expectation is

(19) $\sum_{i=0}^{\infty} l_i (p - p_0)^i$

we simply equate coefficients on the right hand side of (18) with those of (19) and solve for the a_i. This means that the coefficients of the estimating expansion are those of the desired expansion, weighted by a survival function $G(i)$ that can be chosen by the user. This intuition is formalized by the following assumptions, lemmae and propositions.

ASSUMPTION 1: $l(p) = \sum_{i=0}^{\infty} l_i (p - p_0)^i$ *is an analytic function with circle of convergence* $C(R) = \{p \mid |p - p_0| < R\}$.

ASSUMPTION 2: *The* $\{s_1, \ldots, s_I\}$ *are independent and identically distributed for any I and* $|s_i| \le R - 2e$ *for arbitrary* $e > 0$[15] .

ASSUMPTION 3: *The random truncation variable I has survival function* $G(i)$*, finite expectation, and is independent of any s.*[16]

[14] Note that the limits in the sums in (18) are different.

[15] The seemingly superfluous factor 2 will simplify some analysis later.

[16] If *I* defined on $[0, \infty]$ has a finite expectation then $E[I] = \sum_{i=0}^{\infty} G(i) < \infty$.

ASSUMPTION 4: $E(s_i \mid I) = p - p_0$.

We will find it convenient to define related functions:

$$1^*(p) = \sum_{i=0}^{\infty} |l_i| |p - p_0|^i$$

$$1'(p) = \sum_{i=0}^{\infty} l_i i (p - p_0)^{i-1}$$

$$1'^*(p) = \sum_{i=0}^{\infty} |l_i| |p - p_0|^i .$$

By the analycity of $1(p)$ all are analytic and all have circle of convergence $C(R)$ [17] and all are bounded for $p \in C(R - e)$ for any $e > 0$.

If we naively chose our estimator to be $S_I = \sum_{i=0}^{I} l_i U_i(s, I)$, an expansion with coefficients α_i equal to the l_i in Assumption 4 and degree of truncation, *I*, as in Assumption 3, we would obtain a biased estimator, but one that suggests a correction to obtain an unbiased one. First, the expectation of the naive estimator,

LEMMA 1: *Under Assumptions 1 – 4* $E(S_I) = \sum_{i=0}^{\infty} l_i G(i)((p - p_0))^i$

Proof: Let $g_I = \sum_{i=0}^{I} l_i U_i(z, I)$ then

$$\begin{aligned}
E(|g_I| \mid I) &= E\left|\sum_{i=0}^{I} l_i U_i(z, I)\right| \\
&\le \sum_{i=0}^{I} |l_i| E|U_i(z, I)| \\
&= \sum_{i=0}^{I} |l_i| E\left|\sum_{r_1 < L < r_i} (s_{r_1} L \; s_{r_i}) / C(I, i)\right| \\
&\le \sum_{i=0}^{I} |l_i| \sum_{r_1 < L < r_i} E\left|(s_{r_1} L \; s_{r_i})\right| / C(I, i) \\
&= \sum_{i=0}^{I} |l_i| \sum_{r_1 < L < r_i} E|s_1|^i / C(I, i) \\
&= \sum_{i=0}^{I} |l_i| E|s_1|^i \\
&\le \sum_{i=0}^{\infty} |l_i| E|s_1|^i \\
&= 1¾^*\left(E|s_1|^i\right)
\end{aligned}$$

[17] See Parks and Krantz (1992).

Since $|s_1| < R - 2e$, $E|s_1|$ lies strictly in the circle of convergence of I%*. By Levi's Theorem (see Kolmogorov and Fomin (1975) Chapter 30, Section 8, Theorem 2.)

$$
\begin{aligned}
E(g_I) &= \sum_{I=0}^{\infty} \sum_{i=0}^{I} l_i E\left(U_i(z,I) \mid I\right) P[I] \\
&= \sum_{I=0}^{\infty} \sum_{i=0}^{I} l_i (p - p_0)^i P[I]
\end{aligned}
$$

moreover both the inner sum and outer converge absolutely, hence by the Weierstrauss Double Sum Theorem (Knopp) the order of the sums can be exchanged. Thus

$$
\begin{aligned}
\sum_{I=0}^{\infty} \sum_{i=0}^{I} l_i (p - p_0)^i P[I] &= \sum_{i=0}^{\infty} \sum_{I=i+1}^{\infty} l_i (p - p_0)^i P[I] \\
&= \sum_{i=0}^{\infty} l_i (p - p_0)^i \sum_{I=i+1}^{\infty} P[I] \\
&= \sum_{i=0}^{\infty} l_i (p - p_0)^i G(i)
\end{aligned}
$$

Q.E.D.

Each term of the expansion is off by a factor of $G(i)$. This suggests that to obtain unbiased estimates of an analytic function whose coefficients are λ_i we use weighted coefficients for the estimating expansion, in particular,

ASSUMPTION 5: I%$(z, I) = \sum_{i=1}^{I} l_i U_i\left(s(z), I\right) / G(i)$

However, if we stop here, we might have a new problem, as this expansion may not converge. For though the $U_i(s, I)$ are all in the circle of convergence, $U_i\left(s(z), I\right)/G(i)$ and more importantly, $E\left(U_i\left(s(z), I\right) \mid I\right)/G(i)$ can easily be outside the circle of convergence. Thus, a crucial condition in rearrangement the proof would be violated. So we additionally require that $E\left(U_i\left(s(z), I\right) \mid I\right)/G(i) \in C(R)\ \forall I \geq I_0 \geq 0$, that is, eventually the ratios all fall in the circle of convergence. Another way of writing this is, $(R-e)^i E\left(U_i\left(s(z), I\right) \mid I\right) / \left((R-e)^i G(i)\right) \in C(R)\ \forall I \geq I_0 \geq 0$, since the simulants satisfy Assumption 2 and since $R - e$ is in the circle of convergence, this means that a sufficient condition for convergence is the survival function must go to zero more slowly than $((R-2e)/(R-e))^i$. Thus the following guarantees convergence and existence of the expansion.

ASSUMPTION 6: *For all* $0 \le a < 1$, $a^i / G(i) < 1$.

This assumption further limits the distributions that can be used to generate the truncation term. In the last section, on computation, we derive a class of survival functions satisfying this assumption as well as supporting a finite expectation, an assumption we will need below.

PROPOSITION 7: *Under Assumptions 1-6* $E(\tilde{l}(z,I)) = l(p)$

PROOF: Using the approach of the last Proposition, we write

$g_I = \sum_{i=0}^{I} l_i U_i(z,I)/G(i)$ then

$E(\tilde{l}(z,I)) = E\sum_{I=0}^{\infty}\left(\sum_{i=0}^{I} l_i (U_i(z,I)/G(i)) P[I]\right) = E\sum_{I=0}^{\infty} g_I$. We show the sum converges absolutely.

$$
\begin{aligned}
|g_I| &= \left|\sum_{i=0}^{I} l_i U_i(z,I)/G(i)\right| P[I] \\
&\le \sum_{i=0}^{I} |l_i| E\left|U_i(z,I)\right| P[I]/G(i) \\
&= \sum_{i=0}^{I} |l_i| E\left|\sum_{r_1<\cdots<r_i}(s_{r_1}\cdots s_{r_i})/C(I,i)\right| P[I]/G(i) \\
&\le \sum_{i=0}^{I}\left(|l_i| \sum_{r_1<\cdots<r_i} E\left|(s_{r_1}\cdots s_{r_i})\right|/C(I,i)\right) P[I]/G(i) \\
&= \sum_{i=0}^{I}\left(|l_i| \sum_{r_1<\cdots<r_i} E|s_1|^i/C(I,i)\right) P[I]/G(i) \\
&= \sum_{i=0}^{I} |l_i| E|s_1|^i P[I]/G(i) \\
&\le \sum_{i=0}^{\infty} |l_i| (R-e)^i \left(E|s_1|^i/(R-e)^i\right) P[I]/G(i) \\
&\le \sum_{i=0}^{\infty} |l_i| (R-e)^i ((R-2e)/(R-e))^i P[I]/G(i) \\
&\le \sum_{i=0}^{\infty} |l_i| (R-e)^i P[I]\left[((R-2e)/(R-e))^i/G(i)\right] \\
&\le \tilde{l}^*(R-e) P[I]
\end{aligned}
$$

This latter converges to zero so by Levi (Kolmogorov and Fomin) the order of the continuous expectation and the sums can be exchanged as $I \to \infty$, moreover, the outer sum converges absolutely, so by the Weierstrauss rearrangement theorem the order of the sums can be rearranged.

Taking expectations and rearranging gives

$$\begin{aligned}
E\left(\tilde{l}(z,I)\right) &= \sum_{I=0}^{\infty}\sum_{i=0}^{I}\left(l_i\,(p-1)^i / G(i)\right)P[I] \\
&= \sum_{i=0}^{\infty}\sum_{i=I+1}^{\infty}\left(l_i\,(p-1)^i / G(i)\right)P[I] \\
&= \sum_{i=0}^{\infty} l_i\,(p-1)^i \sum_{i=I+1}^{\infty} P[I]/G(i) \\
&= \sum_{i=0}^{\infty} l_i\,(p-1)^i \\
&= 1\,(p)
\end{aligned}$$

Q.E.D.

We mention that this estimator is a Rao-Blackwell estimator since for each *I*, the order statistics are sufficient, hence by a theorem due to Fay(1950)[18], $\left\{x_{(1)},\ldots,x_{(I)},I\right\}$ is a sufficient statistic. In the next section we will need an additional property for this estimator.

DEFINITION 1: The simulation residual process is given by

$\sum_{t=1}^{T}\left(\tilde{l}_t-\tilde{\tilde{l}}_t-\tilde{E}\left(\tilde{l}_t\right)-\tilde{E}\left(\tilde{\tilde{l}}_t\right)\right)/\sqrt{T}$.[19]

Following Hajivassiliou and McFadden (1994), for purposes of using the simulants in estimation, we will need to show this and various other simulation residual processes are stochastically equicontinuous.

DEFINITION 2: A simulant $s = s(w,J)$ is said to be *stochastically equicontinuous* if for any h^* and e^* there is a d such that $\forall J\ \left|J - J'\right| < d$

$\Pr\left[\left|s(w,J) - s(w,J')\right| > e^*\right] \le h^*$.

---

[18] See Theorem 4.3.1 in Govindarajulu(1987).

[19] $\tilde{E}$ indicates the expectation with respect to the simulation process.

PROPOSITION 8: *Let the simulants* $\{s_{it}\}$ $i = 1,\ldots, I_t, t = 1,\ldots,T$ *and* $\sum_{t=1}^{T} \tilde{E}\left(\tilde{\ell}_t\right)/\sqrt{T}$ *be stochastically equicontinuous and let* $\tilde{\ell}_t$ *be the unbiased estimators of the analytic functions* $\tilde{\ell}_t$ *then the residual simulation process is stochastically equicontinuous.*

PROOF: Under the assumptions above residual simulation process is stochastically equicontinuous if $\sum_{t=1}^{T} \tilde{\ell}_t/\sqrt{T}$ is (see Newey (1991)). We will find a $d > 0$ such that $\forall J \; \|J - J'\| < d \;\; \Pr\left[\sum_{t=1}^{T}\left|\tilde{\ell}_t - \hat{\ell}_t\right|/\sqrt{T} > e\right] \le h$. By the boundedness of analytic functions and Chebychev's inequality we have

$$\Pr\left[\sum_{t=1}^{T}\left|\tilde{\ell}_t - \hat{\ell}_t\right|/\sqrt{T} > e\right] \le \Pr\left[\sum_{t=1}^{T} \tilde{\ell}^{*} \max_{i \le I_t}\left|s_{it} - s'_{it}\right|/\sqrt{T} > e\right]$$

$$\le \sum_{t=1}^{T}\left(\tilde{\ell}^{*}\right)^2 V\left(\max_{i\le I_t}\left|s_{it} - s'_{it}\right|\right)/\left(Te^2\right)$$

Now $V\left(\max_{i\le I_t}\left|s_{it} - s'_{it}\right|\right) \le E\left(\max_{i\le I_t}\left|s_{it} - s'_{it}\right|^2\right)$

$$\le E\left(I_t\right)E\left(\left|s_{it} - s'_{it}\right|^2\right)$$

$$= E\left(I\right)E\left(\left|s_{it} - s'_{it}\right|^2\right)$$

By stochastic equicontinuity for each t, and any h^* and e^* there is a d_t such that $\forall J \; \|J - J'\| < d_t \;\; \Pr\left[\left|s_{it} - s'_{it}\right| > e^*\right] \le h^*$. Let $z = \left|s_{it} - s'_{it}\right|$ and f be its density then $E\left(\left|s_{it} - s'_{it}\right|^2\right) = \int_0^R z^2 f(z)\,dz$. Let

$I(z;e^*)$ be the indicator for the set $\{z \mid z \le e^*\}$ then

$$\int_0^R z^2 f(z)\,dz = \int_0^R I(z;e^*)\,z^2 f(z)\,dz + \int_0^R \left(1 - I(z;e^*)\right) z^2 f(z)\,dz$$

$$\le (e^*)^2 \Pr[z \le e^*] + \left(R^2 - (e^*)^2\right)\left(1 - \Pr[z \le e^*]\right)$$

$$= (e^*)^2 + \left(R^2\right)\left(1 - \Pr[z \le e^*]\right)$$

$$\le (e^*)^2 + \left(R^2\right)h^*$$

so

$$\Pr\left[\sum_{t=1}^{T}\left|\tilde{\ell}_t - \hat{\ell}_t\right| / \sqrt{T} > e\right] \le \sum_{t=1}^{T}\left(\ell^{*}\right)^2 V\left(\max_{i \le I_t}\left|s_{it} - s'_{it}\right|\right) / \left(Te^2\right)$$

$$\le E\left(I\right)\left(\ell^{*}\right)^2\left(\left(e^{*}\right)^2 + \left(R^2\right)h^{*}\right)/e^2$$

$$= h$$

Now choose any h^{*} and e^{*} so that $h = E\left(I\right)\left(\ell^{*}\right)^2\left(\left(e^{*}\right)^2 + \left(R^2\right)h^{*}\right)/e^2$

and choose $d = \min_{t \le T}\left\{d_t\right\}$.

Q.E.D.

1.11. Final Applications to the Logarithm

For the log-likelihood, again expand ln(*p*) around p_0 to obtain

(20) $\ln\left(p\right) = \ln\left(p_0\right) + \sum_{i=1}^{\infty}\left(-1\right)^i\left(p - p_0\right)^i / \left(ip_0^i\right)$

and let $I \sim Geometric\left(w\right)$, which has survival function,[20] $G\left(i\right) = \left(1-w\right)^i$.

$\left|E\left(U_i\left(s\left(z\right), I\right) \mid I\right)\right| / \left(1-w\right)^i = \left|\left(p - p_0\right)^i\right| / \left(1-w\right)^i$ so if

(21) $\left|\left(p - p_0\right) / \left(1-w\right)\right| \le M < 2p_0$

then by Proposition 6 $l_i = \left(-1\right)^i / ip_0^i\left(1-w\right)^i$, $i = 1, \ldots, I$ are the appropriate weights. Without knowing p it is hard to guarantee (21). Alternatively, we can use survival functions that go to zero more slowly (eventually) than $\left|\left(p_0 - p\right)\right|^i$. As we will see below, the random truncation point must also have a finite mean for use in estimation. As shown in an appendix, a simple practical class of survival functions satisfying all requirements has the form $G\left(i; B, a\right) = B^{i^a}$ $0 < a < 1$, $0 < B < 1$.

PROPOSITION 9: *Let* $G\left(i\right) = B^{i^a}$ $0 < a < 1$, $0 < B < 1$ *then for any* $a < 1$ *there is an I* such that for all* $i > I^{*}$, $a^i / G\left(i\right) \le M < 1$. *Moreover,* $\lim_{i \to \infty} a^i / G\left(i\right) = 0$.

General construction methods and random number generation for a truncation variable

---

[20] See Johnson, Kotz and Kemp(1993).

with this survival function are presented in the appendices.

2. Derivatives and the Score

In this section, we develop and examine three estimators for the score, the gradient of the log-likelihood. We are solely interested here in developing unbiased estimators of derivatives with respect to J when the simulants depend on J either explicitly or implicitly. The first estimator is simply the gradient of an unbiased analytic function estimator when the simulator depends explicitly on J and is continuously differentiable. We shall refer to this as a direct estimator; the mixed logit is a good example. The second estimator is the numerical gradient with fixed increment Δ. We shall call this the numeric estimator; we shall use this when the simulant does not depend differentiably on J. The final estimator will be used when the simulant does not depend on J explicitly but the density of the simulating process does, and the expected value of the log-likelihood is differentiable. This estimator seems new and will be developed fully below. However it is very simply described: it is the unbiased estimator of the log likelihood multiplied by the score of the simulation process itself. We shall call it the indirect estimator. For use in estimation we need versions that fit neatly into the framework of Hajivassiliou and McFadden (1998), particularly, we will need to show the three implied score residual simulation processes are stochastically equicontinuous. We will demonstrate stochastic equicontinuity of the simulated scores as a partial consequence of the stochastic equicontinuity of commonly used simulants, i.e. those demonstrated as such by Hajivassiliou and Mcfadden (1998).

2.1. The Direct Score

We begin with the direct estimator. For the mixed logit the simulant can be written as an explicit function of the parameter

$$(22)\qquad \begin{aligned} s &= S\left(m+\Omega^{1/2}Z\right) \\ &= \exp\left(W_1\left(m+\Omega^{1/2}Z\right)\right)/\left(\sum\nolimits_{j=1}^{p}\exp\left(W_j\left(m+\Omega^{1/2}Z\right)\right)\right), \end{aligned}$$

where $Z \sim N(0,I)$. Clearly, (22) is continuously differentiable and depends explicitly on the parameters. More generally, if $s = s(z;J)$ where s is differentiable, we use the gradient of the unbiased log-likelihood $\tilde{l}\left(s(z;J),I\right)$ given by

(23) $$\nabla_J \tilde{l}\left(s(z;J),I\right)=\sum_{i=1}^{I} l_i \nabla_J U_i\left(s(z;J),I\right)/G(i)$$

where

(24) $$\nabla_J U_i(s,I)=\sum_{1\le r_1<\cdots<r_i\le I} \nabla_J\left(s_{r_i}\right)\prod_{j\ne i} s_{r_j}/C(I,i)$$

as an unbiased estimator of the score. The formula may look complicated but is quickly calculated using a recursive algorithm developed in the Appendix. The next two propositions show that the direct score is unbiased and stochastically equicontinuous.

PROPOSITION 10: *Let* $E\left|\nabla_J s(Z;J)\right|^2 \le C^* < \infty$ *then*

1. *there exists a random variable, C, with finite second moments then* $\left|\nabla_J \tilde{l}\left(s(Z;J),I\right)\right| \le C$ *and*

2. $$E\left[\nabla_J \tilde{l}\left(s(Z;J),I\right)\right]=\sum_{I=1}^{\infty}\int\left[\sum_{i=1}^{I} l_i \nabla_J U_i\left(s(z;J),I\right)/G(i)\right]\prod_{i=1}^{I} h^*(z_i)dz_i P[I]$$
$$=\sum_{I=1}^{\infty} i\, l_i\, p(J)^{i-1}\nabla_J\left(s_1(J)\right)$$
$$=\nabla_J l\left(p(J)\right).$$

PROOF: The proof is an exercise in exchanging the order of various limiting operations. Let $g_I=\left[\sum_{i=1}^{I} l_i \nabla_J U_i\left(s(z;J),I\right)/G(i)\right]$ then

$$\left|g_I\right| \le \left[\sum_{i=1}^{I}\left|l\right|_i\left|\nabla_J U_i\left(s(z,J),I\right)\right|/G(i)\right]$$
$$=\left[\sum_{i=1}^{I}\left|l\right|_i\left|\sum_{r_1<\cdots<r_i\le I}\sum_{j=1}^{i}\nabla_J\left(s_{r_j}\right)\prod_{k\ne j,k\le i} s_{r_k}/C(I,i)\right|/G(i)\right]$$
$$\le \sum_{i=1}^{I}\left|l\right|_i \max_{i\le I}\left\{\left|s_i\right|\right\}^{i-1} i \max_{i\le I}\left\{\left|\nabla_J s_i\right|\right\}/G(i)$$
$$=\sum_{i=1}^{I}\left|l\right|_i (R-e)^{i-1}\left(\max_{i\le I}\left\{\left|s_i\right|\right\}^{i-1}/(R-e)^{i-1}\right) i \max_{i\le I}\left\{\left|\nabla_J s_i\right|\right\}/G(i)$$

$$
\begin{aligned}
&\leq \sum_{i=1}^{I} |l|_i (R-e)^{i-1}\left((R-2e)^{i-1}/(R-e)^{i-1}\right) i \max_{i\leq I}\{|\nabla_J s_i|\}/G(i) \\
&\leq \sum_{i=1}^{I} |l|_i (R-e)^{i-1}\, i \max_{i\leq I}\{|\nabla_J s_i|\}\left((R-2e)/(R-e)\right)^{i-1}/G(i) \\
&\leq \sum_{i=1}^{I} |l|_i (R-e)^{i-1}\, i \max_{i\leq I}\{|\nabla_J s_i|\}\left[(a)^{i-1}/G(i)\right] \\
&\leq \sum_{i=1}^{\infty} i|l|_i (R-e)^{i-1} \max_{i\leq I}\{|\nabla_J s_i|\} \\
&= \text{Ĩ⁄ð}*(R-e)\max_{i\leq I}\{|\nabla_J s_i|\}
\end{aligned}
$$

Let $C = \sup_{J\in\Theta}\max_{i\leq I}\{|\nabla_J s_i|\}$ and the first assertion is proved. For the second assertion, we have from the foregoing that

$$
E\left[|g_I| \mid I\right] \leq \text{Ĩ⁄ð}*(R-e)E\left[\max_{i\leq I}\{|\nabla_J s_i|\} \mid I\right] \leq \text{Ĩ⁄ð}*(R-e)IE\left[|\nabla_J s_1|\right] \leq \text{Ĩ⁄ð}*(R-e)IM*
$$

exists for each I. Since $E[I]<\infty$, $E(|g_I|) = \sum_{I=0}^{\infty} E(|g_I| \mid I)P[I]$ converges for all I. Thus, the sum $\text{Ĩ⁄ð}*(R-e)E\left[|\nabla_J s_1|\right]\sum_{I=0}^{\infty} IP[I] = \text{Ĩ⁄ð}*(R-e)E\left[|\nabla_J s_1|\right]E[I]$ exists. So the series $\sum_{I=0}^{\infty} g_I$ converges absolutely and the order of all limiting operations can be interchanged. Again by Levi's theorem that limit is

$$
\sum_{I=1}^{\infty}\int\left[\sum_{i=1}^{I} l_i \nabla_J U_i\left(s(z;J),I\right)/G(i)\right]\prod_{i=1}^{I} h(z_i)dz_i P[I]
$$

$$
\begin{aligned}
E[g_I] &= \sum_{I=0}^{\infty}\left[\sum_{i=0}^{I} l_i \sum_{r_1<\dots<r\leq I_i} E\left(\sum_{k=1}^{i}\nabla_J\left(s_{r_k}\right)\prod_{j\neq k} s_{r_j} \mid I\right)/\left(C(I,i)G(i)\right)\right]P[I] \\
&= \sum_{I=0}^{\infty}\left[\sum_{i=0}^{I} l_i \sum_{k=1}^{i}\nabla_J\left(Es_1(J)\right)\left(Es_1(J)\right)^{i-1}/G(i)\right]P[I] \\
&= \sum_{I=0}^{\infty}\left[\sum_{i=0}^{I} l_i ip(J)^{i-1}\nabla_J\left(p(J)\right)/G(i)\right]P[I]
\end{aligned}
$$

Exchanging sums and using definition of the survivor function $G(i) = \sum_{I=i+1}^{\infty} P[I]$ gives

$$
\begin{aligned}
E[g_I] &= \sum_{i=0}^{\infty}\Big[i\sum\nolimits_{I=i+1}^{\infty} l_i p(J)^{i-1}\nabla_J\left(p(J)\right)/G(i)\Big]P[I] \\
&= \sum_{i=0}^{\infty}\Big[i\,l_i p(J)^{i-1}\nabla_J\left(p(J)\right)/G(i)\sum\nolimits_{I=i+1}^{\infty} P[I]\Big] \\
&= \sum_{i=0}^{\infty}\Big[i\,l_i p(J)^{i-1}\nabla_J\left(p(J)\right)\Big]\Big[\sum\nolimits_{I=i+1}^{\infty} P[I]/G(i)\Big] \\
&= \sum_{i=0}^{\infty}\Big[i\,l_i p(J)^{i-1}\Big]\nabla_J\left(p(J)\right) \\
&= l'\left(p(J)\right)\nabla_J\left(p(J)\right) \\
&= \nabla_J l\left(p(J)\right)
\end{aligned}
$$

Q.E.D.

PROPOSITION 11: *Under the Assumptions of the previous Proposition and the assumptions that* $\nabla_J s(Z;J)$ *is differentiable, and* $\nabla_J l\left(p(J)\right)$ *is equicontinuous the simulation residual process for the direct score is stochastically equicontinuous.*

PROOF: Define g_I as above. Using Lemma 3 on differences of products, and Lemma 5 on derivatives of U-statistics,

$$
\begin{aligned}
g_I - g_I' &= \sum\nolimits_{i=0}^{I} l_i \sum\nolimits_{1\le r_1 < \dots < r_i \le I}\sum\nolimits_{k=1}^{i}\left[\nabla_J s_{r_k}\prod_{j\ne k} s_{r_j} - \nabla_J s_{r_k}'\prod_{j\ne k} s_{r_j}'\right]/\left(C(I,i)G(i)\right) \\
&= \sum\nolimits_{i=0}^{I} l_i \sum\nolimits_{1\le r_1 < \dots < r_i \le I}\sum\nolimits_{k=1}^{i}\left[\left\{\nabla_J s_{r_k} - \nabla_J s_{r_k}'\right\}\prod_{j\ne k} s_{r_j} - \nabla_J s_{r_k}'\left\{\prod_{j\ne k} s_{r_j}' - \prod_{j\ne k} s_{r_j}\right\}\right]/\left(C(I,i)G(i)\right) \\
&= \sum\nolimits_{i=0}^{I} l_i \sum\nolimits_{1\le r_1 < \dots < r_i \le I}\sum\nolimits_{k=1}^{i}\Bigg[\left\{\nabla_J s_{r_k} - \nabla_J s_{r_k}'\right\}\prod_{j\ne k} s_{r_j} \\
&\qquad -\nabla_J s_{r_k}'\left\{\sum\nolimits_{j\ne k}\left[s_{r_j}' - s_{r_j}\right]\prod\nolimits_{1\le i'<j, i'\ne k} s_{r_j}'\prod\nolimits_{j<j'<i, j'\ne k} s_{r_j}\right\}\Bigg]/\left(C(I,i)G(i)\right)
\end{aligned}
$$

Proceeding as before we have

$$\left|\left(g_I - g_I^{'}\right)\right| \le \sum_{i=0}^{I}\left[\left|l_i\right|\left(R-e\right)^{i-1}/G\left(i\right)\right]\left[i\left(\left\{\max_{i\le I}\left|\nabla_J s_i - \nabla_J s_i^{'}\right|\right\}\max_{i\le I}\left|s_i\right|^{i-1}\right)/\left(R-e\right)^{i-1}\right.$$

$$\left. +\left\{\max_{i\le I}\left|\nabla_J s_i^{'}\right|\right\}\left\{i\max_{i\le I}\left|s_i - s_i^{'}\right|\left(R-2e\right)^{i-2}/\left(R-e\right)^{i-1}\right\}\right]$$

$$\le 1'^*\left(R-e\right)\left(\max_{i\le I}\left|\nabla_J s_i - \nabla_J s_i^{'}\right| + \max_{i\le I}\left|\nabla_J s_i^{'}\right|\max_{i\le I}\left|s_i - s_i^{'}\right|\left(\left(R-e\right)/\left(R-2e\right)^2\right)\right)M$$

where $\left(\left(R-2e\right)/\left(R-e\right)\right)^i/G\left(i\right)\le M \;\; \forall i$

By assumption $\nabla_J s\left(Z;J\right)$ is differentiable, thus using Taylor's expansions we have $\left|\left(s_i - s_i^{'}\right)\right| = \left|\left(\nabla_J s_i^{*}\left(J-J'\right)\right)\right| \le \sup_{J^*\in\Theta}\left\|\nabla_J s_i^{*}\right\|\left\|\left(J-J'\right)\right\| = B_i\left\|\left(J-J'\right)\right\|$ and

$\left|\left(\nabla_J s_i - \nabla_J s_i^{'}\right)\right| \le C_i\left\|J-J'\right\|$ for some random $C_i \ge 0$.

Hence,

$$\left|g_I - g_I^{'}\right| \le 1'^*\left(R-e\right)\left(\max_{i\le I}C_i\left\|\left(J-J'\right)\right\| + \left(\max_{i\le I}B_i\right)^2\left\|\left(J-J'\right)\right\|\left(\left(R-e\right)/\left(R-2e\right)^2\right)\right)M$$

$$= 1'^*\left(R-e\right)\left(\max_{i\le I}C_i + \left(\max_{i\le I}B_i\right)^2\left(\left(R-e\right)/\left(R-2e\right)^2\right)\right)M\left\|\left(J-J'\right)\right\|$$

$$= A_i\left\|\left(J-J'\right)\right\|$$

So for all J, J' such that $\left\|\left(J-J'\right)\right\| < d$ we have

$$\Pr\left[\left|g_I - g_I^{'}\right| > e\right] \le \Pr\left[A_i\left\|\left(J-J'\right)\right\| > e\right] = \Pr\left[A_i > e/\left\|\left(J-J'\right)\right\|\right] \le \Pr\left[A_i > e/d\right] .$$

Now for any h and e, d can be chosen so small that

$\Pr\left[\left|g_I - g_I^{'}\right| > e\right] \le \Pr\left[A_i > e/d\right] < h$. Since $\nabla_J 1\left(p\left(J\right)\right)$ is equicontinuous, Lemma 1 below and Lemma A.1 of Newey (1991) hold and the proposition is proved.

Q.E.D.

2.2. Indirect Score

So when the simulant is a differentiable function of J, simply take the derivative. However, when the distribution of *s* depends on unknown parameters J but s does not, or does so, but is not differentiable, other approaches are required. The second of the estimators mentioned above, I call the indirect score; it is valid when the density of the simulating process is differentiable.

An example of such is the mixed logit written as in (4). There s does not depend on J , but its expectation,

(25) $$s\left(J\right)=E\left(s\right)=\int \exp\left(W_1 z\right)/\left(\sum_{j=1}^{p}\exp\left(W_j z\right)\right)h\left(z;J\right)dz$$

does. By differentiating both sides, we can discover another unbiased estimator of the score. More generally, let $\tilde{l}\left(s\left(z\right),I\right)$ be an unbiased estimator of $l\left(p\left(J\right)\right)$, where s does not depend explicitly on J and let $z=\left\{z_1,\ldots,z_I\right\}$ have density $h_I\left(z,J\right)=\prod_{i=1}^{I}h\left(z_i,J\right)$ conditional on *I*, then by unbiasedness,

(26) $$l\left(p\left(J\right)\right)=\sum_{i=0}^{\infty}\left[\int \tilde{l}\left(s\left(z\right),i\right)h_I\left(z,J\right)dx\right]\Pr[I=i].$$

Assuming we can freely differentiate both sides by è term by term if necessary

(27) $$\begin{aligned}\nabla_J l\left(p\left(J\right)\right) &= \nabla_J \sum_{i=0}^{\infty}\left[\int \tilde{l}\left(s\left(z\right),i\right)h_i\left(z,J\right)dx\right]\Pr[I=i]\\ &= \sum_{i=0}^{\infty}\left[\int \tilde{l}\left(s\left(z\right),i\right)\nabla_J h_i\left(z,J\right)dx\right]\Pr[I=i]\\ &= \sum_{i=0}^{\infty}\left[\int \tilde{l}\left(s\left(z\right),i\right)\left(\nabla_J h_i\left(z,J\right)/h_i\left(z,J\right)\right)h_i\left(z,J\right)dx\right]\Pr[I=i]\\ &= E\left(\tilde{l}\left(s\left(z\right),I\right)\cdot\left(\nabla_J h_I\left(z,J\right)/h_I\left(z,J\right)\right)\right)\\ &= E\left(\tilde{l}\left(s\left(z\right),I\right)\cdot\nabla_J \ln\left(h_I\left(z,J\right)\right)\right)\end{aligned}$$

So if $\tilde{l}\left(s\left(z\right),I\right)$ is an unbiased estimator of $l\left(p\left(J\right)\right)$ then the simple product

(28) $$\begin{aligned}\tilde{l}_J^{*} &= \tilde{l}\left(S\left(z\right),I\right)\cdot\nabla_J \ln\left(h_I\left(z,J\right)\right)\\ &= \sum_{i=0}^{I}\left(l_i U_i\left(z,I\right)/G\left(i\right)\right)\nabla_J \ln\left(h_I\left(z,J\right)\right)\end{aligned}$$

is unbiased for $\nabla_J l\left(p\left(J\right)\right)$. This is formalized by the following, ASSUMPTION 7: *Conditional on I,* $\{Z_1,\ldots,Z_I\}$ *are independent and identically distributed.*

ASSUMPTION 8: *Conditional on I,* $Z=\{Z_1,\ldots,Z_I\}$ *has density*

$h_I(z,J)=\prod_{i=1}^{I} h(z_i,J)$,.

ASSUMPTION 9: $s=S(Z)$ *is an integrable function of Z.*

ASSUMPTION 10: $p(J)=\int S(z)h_I(z,J)dz$

PROPOSITION 12: *Let* $\hat{\delta}_J=\tilde{\lambda}(S(z),I)\cdot\nabla_J \ln(h_I(z,J))$, *if A1-A10 hold then* $\left|\hat{\delta}_J\right|\le\tilde{\lambda}^*(R-e)I\max_{i\le I}\left|\nabla_J \ln(h(z_i,J))\right|$*; if, in addition,* $E(I)<\infty$ *holds, then* $E(\hat{\delta}_J)=\nabla_q\lambda(p(J))$.

PROOF: $\hat{\delta}_J=\tilde{\lambda}(S(z),I)\cdot\nabla_J \ln(h_I(z,J))$

$$=\left(\sum_{i=0}^{I}\lambda_i U_i(z,I)/G(i)\right)\sum_{i=0}^{I}\nabla_J \ln(h(z_i,J))$$

Using an argument we have used before

$$\left|\hat{\delta}_J\right|\le\left|\left(\sum_{i=0}^{I}\lambda_i U_i(z,I)/G(i)\right)\right|\left|\sum_{i=0}^{I}\nabla_J \ln(h(z_i,J))\right|$$

$$\le\left(\sum_{i=0}^{I}\left(|\lambda_i|\sum_{r_1<\ldots<r_i}\left|\left(s_{r_1}\ldots s_{r_i}\right)\right|/C(I,i)\right)/G(i)\right)\left|\sum_{i=0}^{I}\nabla_J \ln(h(z_i,J))\right|$$

$$\le\left(\sum_{i=0}^{I}\left(|\lambda_i|\max_{i\le I}\{|(s_i)|\}\right)/G(i)\right)\left|\sum_{i=0}^{I}\nabla_J \ln(h(z_i,J))\right|$$

$$\le\left(\sum_{i=0}^{I}\left(|\lambda_i|(R-e)^i\left(\max_{i\le I}\{|(s_i)|\}/(R-e)\right)^i\right)/G(i)\right)\left|\sum_{i=0}^{I}\nabla_J \ln(h(z_i,J))\right|$$

$$\le\left(\sum_{i=0}^{I}\left(|\lambda_i|(R-e)^i\left((R-2e)/(R-e)\right)^i\right)/G(i)\right)\left|\sum_{i=0}^{I}\nabla_J \ln(h(z_i,J))\right|$$

$$\le\left(\sum_{i=0}^{I}|\lambda_i|(R-e)^i\right)\sum_{i=0}^{I}\left|\nabla_J \ln(h(z_i,J))\right|$$

$$=\tilde{\lambda}^*(R-e)\sum_{i=0}^{I}\left|\nabla_J \ln(h(z_i,J))\right|$$

Thus using Wald's Equation[21]

[21] See Ross (1992) Theorem 3.6 p.38.

$$E\left|\tilde{s}_J\right| \le \tilde{s}^*\left(R-e\right)E\left(\sum_{i=0}^{I}\left|\nabla_J \ln\left(h\left(z_i,J\right)\right)\right|\right)$$

$$= \tilde{s}^*\left(R-e\right)E\left(I\right)E\left(\left|\nabla_J \ln\left(h\left(z_1,J\right)\right)\right|\right)$$

Thus by the Levi's theorem (27) holds since exchanging the infinite sum, differentiation and the integral is valid.

Q.E.D.

Thus an unbiased estimate of the score is $\tilde{s}$ times the score of the log-likelihood of the distribution of the underlying simulants. It is also stochastically equicontinuous.

PROPOSITION 13: *Provided $\nabla_J \ln\left(h\left(z,J\right)\right)$ is continuously differentiable and the simulant is stochastically equicontinuous, the indirect score and its residual simulation process are stochastically equicontinuous.*

Proof: Write $\tilde{s} = \tilde{s}\left(s\left(z\right),I\right)$ and $\tilde{s}' = \tilde{s}\left(s\left(z'\right),I\right)$ where $z \in R^I$ and the densities of the simulation processes z and z' are $h = h\left(z,J\right)$ and $h' = h\left(z,J'\right)$, respectively.

$$\Pr\left[\left|l\,h - l'h'\right| > h^*\right] = \Pr\left[\left|l\left(h-h'\right) - \left(l'-l\right)h'\right| > h^*\right]$$

$$\le \Pr\left[\left|l\left(h-h'\right)\right| > h^*/2\right] + \Pr\left[\left|\left(l'-l\right)h'\right| > h^*/2\right]$$

$$\le \Pr\left[l^*\left(R-e\right)\left|\left(h-h'\right)\right| > h^*/2\right] + \Pr\left[l'^*\left(R-e\right)\left|h'\right|\max_{i\le I}\left|s_i - s_i'\right| > h^*/2\right]$$

For the first term, since h is twice logarithmically differentiable, we may, for any e^* and h^*, find a d^* such that

$$\Pr\left[l^*\left|\left(h-h'\right)\right| > h^*/2\right] \le \Pr\left[l^* B\left|\left(J-J'\right)\right| > h^*/2\right]$$

$$\le \Pr\left[B > h^*/\left(2l^*\left|\left(J-J'\right)\right|\right)\right]$$

$$\le \Pr\left[B > h^*/\left(2l^* d^*\right)\right]$$

$$\le e^*/2 \qquad \forall \ \left|\left(J-J'\right)\right| \le d^*$$

For the second term, consider the set $B = \left\{\left(y,x\right): yx \ge h^*/2\right\} \cap \left\{\left(y,x\right): x \ge h\right\}$ where x corresponds to $\max_{i\le I}\left|s_i - s_i'\right|$ and y corresponds to $\left|h'\right|$, B is contained in the set

$A=\{(x,y) \mid x>h, y>h^*/(2h)\}$. Hence $B \subseteq A$ and

$$\Pr\left[|h'|\max_{i\le I}\left|s_i-s_i^{'}\right|>h^*/(2l'^*)\cap\max_{i\le I}\left|s_i-s_i^{'}\right|\le h\right]\le\Pr\left[|h'|\ge\left(h^*/(2l^*h)\right)\right]+\Pr\left[\max_{i\le I}\left|s_i-s_i^{'}\right|\ge h\right]$$

$$\le E[I]\left(\Pr\left[\left|h(z_1,J')\right|\ge\left(h^*/(2l^*h)\right)\right]+\Pr\left[\left|s_1-s_1^{'}\right|\ge h\right]\right)$$

For the first term of this latter, recall e^* and h^* are fixed arbitrarily , pick h so small that $E[I]\left(\Pr\left[\left|h(z_1,J')\right|\ge\left(h^*/(2l^*h)\right)\right]\right)<(e^*/4)$. For the second of these terms, since $s(z)$ is stochastically equicontinuous we can find a d^o such that

$$E[I]\left(\Pr\left[\left|s_1-s_1^{'}\right|\ge h\right]\right)<(e^*/4) \quad \forall \quad \left|(J-J')\right|\le d^o \quad \text{hence}$$

$$\Pr\left[\left|(l'-l)h'\right|>h^*\right]<e^*/2 \quad \forall \quad \left|(J-J')\right|\le d^o$$

now set $d=\min\left(d^*,d^o\right)$ so for each for any e^* and h^*, we have $\Pr\left[\left|l\,h-l'h'\right|>h^*\right]<e^* \;\; \forall \;\; \left|(J-J')\right|\le d$. Thus the indirect score is stochastically equicontinuous. Differentiability of the simulation likelihood means it is equicontinuous. By Lemma 1 below and Lemma A.1 of Newey (1991) the proposition is proved.

Q.E.D.

2.3. The Numeric Score

Finally, since the numerical derivative, $\left(\tilde{l}(J+\Delta)-\tilde{l}(J)\right)/\Delta$, is simply a linear combination of two unbiased estimators, it is automatically unbiased for any fixed Δ. For use in SML estimation it must also be shown to be stochastically equicontinuous, as in the next theorem.

PROPOSITION 14: *If $s(Z)$ is stochastically equicontinuous then the residual simulation process for the numerical gradient of the unbiased estimator of an analytic function is stochastically equicontinuous.*

PROOF: By Proposition 8 we found that the residual simulation process for $\tilde{l}_t$ was stochastically equicontinuous. We state without proof that a linear combination of such processes is also stochastically equicontinuous.

Q.E.D.

3. SML ESTIMATION

The results presented in this paper provide unbiased estimators for the log-probability, score and any other rational functions of expectations that have a radius of convergence equal to the range of the random variable used for simulation. It remains then to show SML estimators based on the score estimators are consistent and asymptotically normal. The conditions below are easy to check and cover all practical situations; they are not the weakest possible. Because of the unbiasedness, the structure of the problem is a bit simpler than that of Hajivassiliou and McFadden (1998). Indeed, the SML using any of the simulators above is a simple exercise in checking their conditions.

ASSUMPTION 11: *The true value* J^* *is in the interior of a compact parameter set* Θ.

ASSUMPTION 12: *The simulated score is an unbiased estimator of* $\nabla_J l(J)$.

ASSUMPTION 13: *The score* $\nabla_J l(J)$ *is continuously differentiable on* Θ .

ASSUMPTION 14: *The score, its derivatives, and the simulated score, are dominated by functions independent of* J *with finite first and second order moments, and the simulants, s, lie in a compact set* $C(R-2e)$ *in the circle of convergence* $C(R)$ *and the residual simulation process is stochastically equicontinuous.*

ASSUMPTION 15: $E\left[\nabla_J l_n(J)\right]=0$ *if and only if* $J=J^*$.

ASSUMPTION 16: $J=-E_n\left[\nabla_{JJ} l_n(J)\right]$ *is positive definite, where* E_n *denotes expectation with respect to the distribution of the observations*

ASSUMPTION 17: *Observations and simulators are independently identically distributed* *across observations*

ASSUMPTION 18: *The SML estimator solving* $0 = \nabla_J \tilde{\ell}(J)$ *exists for each N.*

THEOREM 1: *Under Assumptions 1-17 the SML estimator satisfies*

$$\hat{J}_N \xrightarrow{P} J^*$$

$$\sqrt{N}\left(\hat{J}_N - J^*\right) \xrightarrow{d} Z \sim N\left(0, J^{-1} - J^{-1}QJ^{-1}\right)$$

where $Q = E\left[\nabla_J \tilde{\ell}(J)\nabla_J \tilde{\ell}(J)^T\right]$.

PROOF: By construction the three scores developed satisfy Assumptions 12 and 13. The rest of the Assumptions depend on the data generating process (as opposed to the simulation process). The theorem then follows directly from Hajivassiliou and McFadden (1998).

Q.E.D.

4. COMPUTATIONAL ISSUES

To prevent chatter, for $n = 1,\mathrm{L}\ ,N$ one simulates I_n, then that number of standard normal vectors e_i each having the same dimension as Z_i and then calculates $Z_i = m + H^T e_i \;\; i = 1,\mathrm{L}\ , I_n$, reusing the same I_n and e_i each time.

A.1 A Useful Recursion

The sums of products for the U statistics and their gradients in can be difficult to calculate efficiently. The following recursion is useful in this regard. Define M and M' with regard to the U statistic and gradient as

$$M_j = C(I, j)U_j(s, I)$$
$$M'_j = C(I, j)\nabla_J U_j(s, I).$$

And let $M = \{M_1, \mathrm{L}\ \ M_I\}$ then the algorithm

$$
\begin{aligned}
& i = 1 \\
& M = s_i \\
& M' = \nabla_J s_i^T \\
& \textit{while } i < I \\
& \qquad i = i + 1 \\
& \qquad M' = \begin{bmatrix} M' \\ \cdot\cdot \\ 0 \end{bmatrix} + s_i \begin{bmatrix} 0 \\ \cdot\cdot \\ M' \end{bmatrix} + \begin{bmatrix} 1 \\ \cdot\cdot \\ M \end{bmatrix} \nabla_J s_i^T \\
& \qquad M = \begin{bmatrix} M \\ \cdot\cdot \\ 0 \end{bmatrix} + s_i \begin{bmatrix} 1 \\ \cdot\cdot \\ M \end{bmatrix} \\
& \textit{endwhile}
\end{aligned}
$$

generates the M's and their gradients.

4.1. Survival Functions

In this section we list some useful survival functions, a prove the function suggested above indeed has a finite mean. If $N \sim Poisson(l)$ then $G_P(i) = g(i+1, l)/\Gamma(i)$ where $g(i, l)$ is the incomplete gamma function. If $N \sim Geometric(w)$ then $S_G(i) = (1-w)^i$. To use at least n_0 terms in the expansion, we use a displaced survival function. Displaced survival functions are simply computed from the survival function. If I has survival function $G(i)$ and $Z = I + n_0$ where n_0 is fixed then the survival function for z is given by $G_Z(z) = S_N(0 \vee (z - n_0))$. For the domination result required by Hajivassiliou and McFadden (1998), we proposed a form for a survival function that would satisfy our needs without requiring us to know the unknown probability p. We provide the details here.

Survival functions for random variables with expectations have a simple structure

$$
\begin{aligned}
& G(0) = 1 \\
& G(i) \geq G(i+1) \geq 0 \quad i \geq 0 \\
& G(\infty) = 0
\end{aligned}
$$

For a finite expectation we need $E(I)=\sum_{i=0}^{\infty}G(i)\leq M<\infty$. We also want a survival function that goes to zero slower than $(p_0-p)^i$. If *I* is Geometric(w) with $w<p$ and we set $p_0=1$, then $(1-w)^i$ is such a survival function. We need one that works even if we have a poor idea as to the value of p. We shall construct a simple survival function of this form that satisfies all the conditions. The survival function $G(i)=B^{g(i)}$ with $B<1$ $g(0)=0$, $g(\text{g})$ increasing, $g(\infty)=\infty$, and $(1-p)^i\leq B^{g(i)}$ satisfies all the requirements. Taking logarithms we obtain

(29) $\quad i\ln(1-p)/\ln(B)\geq g(i)$.

The coefficient on *i* is positive so

(30) $\quad g(i)=i^a \;\; 0<a<1$

eventually satisfies (29). Showing that the random variable has a finite mean requires a little more finesse. The ratio and root tests for convergence are indeterminate, as were many of the usual comparison tests. Ermakoff's test worked. Ermakoff's test for convergence[22] says if $G(i)\geq 0$ and $\lim_{k\to\infty}\frac{e^kG(e^k)}{G(k)}=q<1$ then $\sum_{k=0}^{\infty}G(k)$ converges. Its application to the sum of the survivor function shows the expectation exists for all 0<*B*<1 and 0<α<1.

4.2. Generating Stopping Times

Generating random truncation points for a distribution with survival function $G(i)=B^{i^a}$ $0<a<1$ is straightforward. The distribution function is $F(i)=1-B^{i^a}$. So by a standard argument: Let U~U(0,1), set $U'=1-U=B^{I^a}$ where U' is uniform since U is. Solving gives $I=\left\lfloor\sqrt[a]{\ln(U')/\ln(B)}\right\rfloor$ where $\lfloor x\rfloor$ indicates the largest integer smaller than x. In simulations, $a=.3$ *B*=.8 has worked well. It is also the case that the larger is the implied expected value of *I*, the smaller the variance.

[22] See Knopp (1990).

4.3. Expansion Points and Circle of Convergence

The expansion point, p_0, for the logarithm, is also arbitrary, but is important because the circle of convergence is $\{p \mid 0 \leq p \leq 2p_0\}$. For completely unknown probabilities, this means taking $p_0 > .5$. In practice, I have found .51 works well, while anything less sometimes causes numeric problems, and anything greater increases the variance.

5. CONCLUSION

We have developed a general method for obtaining unbiased estimators of analytic functions of expectations when the expectations must be simulated. We then showed that three estimates of the gradient or score of these unbiased functions were also unbiased and that if the underlying simulants that are stochastically equicontinuous, the unbiased functions and scores are stochastically equicontinuous as well. We then showed how to incorporate these into the framework of Hajivassiliou and McFadden (1998) to obtain consistent and asymptotically normal SML estimates based on the unbiased score estimates. Finally, we detailed some computational methods needed to implement the methods.

Department of Economics, University of California, Berkeley, CA, 94720 USA
gmduncan@econ.berkeley.edu

APPENDIX A:

LEMMA 1: *Let* $s_t = s_t(w_t, J)$ *be independent and identically distributed and let* $E\left(|s_t|^2\right) < \infty$ *then* $\sum_{t=1}^{T} s_t / \sqrt{T}$ *is stochastically equicontinuous.*

PROOF:

Let $s_t' = s_t(w_t, J')$. By the independent and identically distributed assumption and Chebychev'e inequality we have

$$\Pr\left[\sum_{t=1}^{T}|s_t - s_t'|/\sqrt{T} > e\right] \le \sum_{t=1}^{T} V(s_t - s_t')/\left(Te^2\right) = V(s_1 - s_1')/e^2 \le E\left(|s_t - s_t'|^2\right)/e^2 .$$

By stochastic equicontinuity choose any h^* and e^* there is a d such that

$\forall J \; \|J - J'| < d_t \;\; \Pr\left[|s_1 - s_1'| > e^*\right] \le h^*$. Let $z = |s_t - s_t'|$ and *f* be its density

then $E\left(|s_t - s_t'|^2\right) = \int_0^\infty z^2 f(z)\,dz \le M$. Let $I(z;e^*)$ be the indicator for the set

$\{z \mid z \le e^*\}$ then

$$\begin{aligned}
\int_0^\infty z^2 f(z)\,dz &= \int_0^\infty I(z;e^*) z^2 f(z)\,dz + \int_0^\infty \left(1 - I(z;e^*)\right) z^2 f(z)\,dz \\
&\le e^{*2} \Pr[z \le e^*] + E\left(z^2 \mid z > e^*\right)\left(1 - \Pr[z \le e^*]\right) \\
&= e^{*2} + M\left(1 - \Pr[z \le e^*]\right) \\
&\le e^{*2} + Mh^*
\end{aligned}$$

where the conditional second moment exists and is bounded by the same assumption on the unconditional moment. So $\Pr\left[\sum_{t=1}^{T}|s_t - s_t'|/\sqrt{T} > e\right] \le \left(e^{*2} + Mh^*\right)/e^2$. Now choose any h^* and e^* so that $h = \left(e^{*2} + Mh^*\right)/e^2$ and with the d determined above.

Q.E.D.

LEMMA 2: *Let* $\{x_1, \ldots, x_I\}$ *be independent and identically distributed random variables and let I be a nonnegative random integer with* $E(I) < \infty$ *then*

$$\Pr\left[\max_{i \le I} |x_i| > e\right] \le E(I)\Pr\left[|x_1| > e\right].$$

PROOF:

$$\begin{aligned}
\Pr\left[\max_{i\le I}|x_i|>e\right] &= \sum_{i=0}^{\infty}\Pr\left[\max_{i\le I}|x_i|>e\mid I\right]\Pr[I] \\
&= \sum_{i=0}^{\infty}\left(1-\Pr\left[\max_{i\le I}|x_i|\le e\mid I\right]\right)\Pr[I] \\
&= \sum_{i=0}^{\infty}\left(1-\Pr\left[|x_1|\le e\right]^I\right)\Pr[I] \\
&= \sum_{i=0}^{\infty}\left(1-\left(1-\Pr\left[|x_1|>e\right]\right)^I\right)\Pr[I] \\
&\le \sum_{i=0}^{\infty} I\Pr\left[|x_1|>e\right]\Pr[I] \\
&= E(I)\Pr\left[|x_1|>e\right]
\end{aligned}$$

Q.E.D.

LEMMA 3: *Let* $\sum_{i=0}^{\infty}v(I)$ *and* $\sum_{I=0}^{\infty}\sum_{i=0}^{I}u(i)v(I)$ *converge absolutely then*

$$\sum_{I=0}^{\infty}\sum_{i=0}^{I}u(i)v(I)=\sum_{i=0}^{\infty}\sum_{I=i+1}^{\infty}u(i)v(I).$$

PROOF:

Let $Y(i,I)=1$ if $i\le I$

$=0$ otherwise

Then

$$\sum_{I=0}^{\infty}\sum_{i=0}^{I}u(i)v(I)=\sum_{I=0}^{\infty}\sum_{i=0}^{\infty}Y(i,I)u(i)v(I)=\sum_{i=0}^{\infty}\sum_{I=0}^{\infty}Y(i,I)u(i)v(I)=\sum_{i=0}^{\infty}\sum_{I=i+1}^{\infty}u(i)v(I)$$

where the exchange in sum order follows from absolute convergence of the sums.

Q.E.D.

LEMMA 4: *Let* $S=\prod_{i=1}^{I}s_i$ *and* $S'=\prod_{i=1}^{I}s_i^{'}$ *then*

$$S-S'=\prod_{i'=1}^{i}s_{i'}-\prod_{i'=1}^{i}s_{i'}^{'}=\sum_{i'=1}^{i}\left(s_{i'}-s_{i'}^{'}\right)\prod_{0\le j<i'}s_j\prod_{i'<j'\le i}s_{j'}^{'}$$

PROOF:

By induction: This is true for i=2, since

$$\begin{aligned}
s_1s_2-s_1^{'}s_2^{'} &= s_1s_2-s_1^{'}s_2+s_1^{'}s_2-s_1^{'}s_2^{'} \\
&= \left(s_1-s_1^{'}\right)s_2+s_1^{'}\left(s_2-s_2^{'}\right)
\end{aligned}$$

Assume it is true for i-1, so for i it is also true, since

$$
\begin{aligned}
s_1 \cdots s_i - s_1^{'} \cdots s_i^{'} &= s_1 \cdots s_i - s_1 \cdots s_{i-1} s_i^{'} + s_1 \cdots s_{i-1} s_i^{'} - s_1^{'} \cdots s_i^{'} \\
&= s_1 \cdots s_{i-1}\left(s_i - s_i^{'}\right) + \left(s_1 \cdots s_{i-1} - s_1^{'} \cdots s_{i-1}^{'}\right) s_i^{'} \\
&= s_1 \cdots s_{i-1}\left(s_i - s_i^{'}\right) + \left(\sum\nolimits_{i'=1}^{i-1}\left(s_{i'} - s_{i'}^{'}\right)\prod\nolimits_{j=1}^{i'-1} s_j \prod\nolimits_{j'=i'+1}^{i-1} s_{j'}^{'}\right) s_i^{'} \\
&= \sum\nolimits_{i'=1}^{i}\left(s_{i'} - s_{i'}^{'}\right)\prod\nolimits_{j=1}^{i'-1} s_j \prod\nolimits_{j'=i'+1}^{i} s_{j'}^{'}
\end{aligned}
$$

Q.E.D..

LEMMA 5: *Let* $U_i(s,I,J) = \sum_{r_1 < \cdots < r_i}\left(s\left(Z_{r_1},J\right)\cdots s\left(Z_{r_i},J\right)\right)/C(I,i)$ *then*

$$
\begin{aligned}
\left|U_i(s,I,J) - U_i(s,I,J')\right| &= \left|\sum_{1 \le r_1 < \cdots < r_i \le I}\left(s\left(Z_{r_1},J\right)\cdots s\left(Z_{r_i},J\right) - s\left(Z_{r_1},J'\right)\cdots s\left(Z_{r_i},J'\right)\right)/C(I,i)\right| \\
&\le i(R-2e)^{i-1}\max_{i \le I}\left|s\left(Z_i,J\right) - s\left(Z_i,J'\right)\right|
\end{aligned}
$$

PROOF: By Lemma 4

$$
\begin{aligned}
\left|s\left(Z_{r_1},J\right)\cdots s\left(Z_{r_i},J\right) - s\left(Z_{r_1},J'\right)\cdots s\left(Z_{r_i},J'\right)\right| &= \left|\sum\nolimits_{i'=1}^{i}\left(s_{r_{i'}} - s_{r_{i'}}^{'}\right)\prod\nolimits_{j=1}^{i'-1} s_{r_j}\prod\nolimits_{j'=i'+1}^{i} s_{r_{j'}}^{'}\right| \\
&\le \sum\nolimits_{i'=1}^{i}\left|s_{r_{i'}} - s_{r_{i'}}^{'}\right|\left|\prod\nolimits_{j=1}^{i'-1}\right|s_{r_j}\left|\prod\nolimits_{j'=i'+1}^{i}\right|s_{r_{j'}}^{'}\right| \\
&\le \sum\nolimits_{i'=1}^{i}\left|s_{r_{i'}} - s_{r_{i'}}^{'}\right|(R-2e)^{i-1} \\
&\le i\max_{i \le I}\left|s_i - s_i^{'}\right|(R-2e)^{i-1}
\end{aligned}
$$

for each i-tuple $(r_1, \cdots, r_i)$ there are $C(I,i)$ such i-tuples.

Q.E.D..

APPENDIX B:

In this appendix we present an exact variance calculation. $\tilde{l}(z,I)$ is a random sum of U-statistics of different orders. Using elements of the theory of U statistics and the law of iterated expectations as it applies to variances, we can calculate the exact variance of $\tilde{l}(z,I)$. Since the simulants are bounded, the variances necessarily exist. Let $s_1, \cdots, s_I$ be independent and identically distributed conditional on *I*, and let U_i be the U-statistic estimator for $m^i = E(s_1)^i$ based on the kernel $s_1 \cdot s_2 \cdots s_i$ conditional on *I* then from

Randles and Wolfe (1979) or Lehmann(1998).

Let $A(I,i,j,c)=C(i,c)C(I-j,i-c)$ then

(B.1) $$Cov(U_i,U_j \mid I)=\sum_{c=1}^{i} A(I,i,j,c)z_c^{i,j} / C(I,j)$$

where $i<j\leq I$, $m=E(s_1)$, $n_2=E(s_1^2)$ and

(B.2) $$z_c^{(i,j)}=E\left[(s_1 \mathrm{L}\ s_c s_{c+1} \mathrm{L}\ s_i)(s_1 \mathrm{L}\ s_c s_{i+1} \mathrm{L}\ s_{i+j-c})\right]-m^{i+j}=n_2^c m^{i+j-c}-m^{i+j}$$

or

(B.3) $$Cov(U_i,U_j \mid I)=\sum_{c=1}^{i} A(I,i,j,c)\left(n_2^c m^{i+j-c}-m^{i+j}\right)/C(I,j)$$

$$=\sum_{c=1}^{i} A(I,i,j,c)m^{i+j}\left(\left(\frac{n_2}{m}\right)^c-1\right)/C(I,j)$$

(B.4) $$V(\tilde{U}\mid I)=\sum_{j=1}^{N} l_j^2 \sum_{c=1}^{j} A(I,j,j,c)m^{2i}\left(\left(\frac{n_2}{m}\right)^c-1\right)/C(I,j)$$

$$+2\sum_{i=1}^{j-1} l_i l_j \sum_{c=1}^{i} A(I,i,j,c)m^{i+j}\left(\left(\frac{n_2}{m}\right)^c-1\right)/C(I,j)$$

From the well known identity $V(\tilde{U})=E(V(\tilde{U}\mid I))+V(E(\tilde{U}\mid I))$ we obtain the variance of $\tilde{U}$.

(B.5) $$E(V(\tilde{U}\mid N))=\sum_{I=1}^{\infty}\left[\sum_{j=1}^{I} l_j^2 \sum_{c=1}^{j} A(I,j,j,c)m^{2i}\left(\left(\frac{n_2}{m}\right)^c-1\right)/C(I,j)\right.$$

$$\left.+2\sum_{i=1}^{j-1} l_i l_j \sum_{c=1}^{i} A(I,i,j,c)m^{i+j}\left(\left(\frac{n_2}{m}\right)^c-1\right)/C(I,j)\right]$$

(B.6) $$E(\tilde{U}\mid I)=\sum_{i=0}^{I} l_i m^i$$

(B.7) $$E\left(E(\tilde{U}\mid I)^2\right)=\sum_{I=1}^{\infty}\left(\sum_{i=0}^{I} l_i m^i\right)^2 \Pr(I)$$

(B.8) $$E\left(\tilde{l}\right)^2 = \left(\sum_{I=1}^{\infty}\left(\sum_{i=0}^{I} l_i m^i\right)\Pr\left(I\right)\right)^2$$

(B.9) $$\begin{aligned}V\left(\tilde{l}\right) = \sum_{I=1}^{\infty}\Bigg[\sum_{j=1}^{I} l_j^2 \sum_{c=1}^{j} A\left(I, j, j, c\right) m^{2i}\left(\left(\frac{n_2}{m}\right)^c - 1\right)\Pr\left(I\right) \\ +2\sum_{i=1}^{j-1} l_i l_j \sum_{c=1}^{i} A\left(I, i, j, c\right) m^{i+j}\left(\left(\frac{n_2}{m}\right)^c - 1\right)\Pr\left(I\right)\Bigg] / C\left(I, j\right) \\ +\sum_{I=1}^{\infty}\left(\sum_{i=0}^{I} l_i m^i\right)^2 \Pr\left(I\right) - \left(\sum_{I=1}^{\infty}\left(\sum_{i=0}^{I} l_i m^i\right)\Pr\left(I\right)\right)^2\end{aligned}$$

APPENDIX C:

The following detail the gradient formulae. As before, let $Z_1, \ldots, Z_I$ be independent and identically distributed *Gaussian*(m,W), then from Ruud (2000) (pp. 928-930)

(C.1) $$\ln\left(h\left(Z;m,\Omega\right)\right) = -\frac{I}{2}\ln\left(2p\right) - \frac{1}{2}\left(I\ln\left(\det\left(\Omega\right)\right) + \left(\sum_{i=1}^{I}\left(Z_i - m\right)^T \Omega^{-1}\left(Z_i - m\right) \right)\right)$$

(C.2) $$w = vech\left(\Omega\right) = \left[w_{11}, w_{12}, \ldots, w_{1I}, w_{22}, w_{23}, \ldots, w_{2I}, \ldots, w_{II}\right]^T$$

(C.3) $$W = \sum_{i=0}^{I}\left(\left(Z_i - m\right)\left(Z_i - m\right)^T\right)$$

(C.4) $$\frac{\partial \ln\left(h\left(Z;m,\Omega\right)\right)}{\partial m} = \Omega^{-1}\sum_{i=0}^{I}\left(Z_i - m\right)$$

(C.5) $$\frac{\partial \ln\left(h\left(Z;m,\Omega\right)\right)}{\partial w} = -\frac{1}{2} vech\left(\Omega^{-1} - \Omega^{-1}W\Omega^{-1}\right)$$

Alternatively, let $Z_i = m + Ke_i$ where K is upper triangular so

$$\begin{aligned} K_{ij} &= k_{ij} \quad j \le i \le p \\ &= 0 \text{ otherwise.} \end{aligned}$$

and consider the gradient of $g\left(WZ_i\right)$ for a differentiable function g

if $W = \left[W_1, \ldots, W_p\right]^T$ and $J = \left[m, k_{11}, \ldots, k_{1p}, k_{22}, \ldots, k_{2p}, \ldots, k_{(p-1)(p-2)}, k_{(p-1)(p-1)}, k_{pp}\right]^T$

then

$$\nabla_J g\left(W^T Z_i\right) = g'\left(W^T Z_i\right)\left[W, W_1 e_1, \ldots, W_1 e_p, W_2 e_2, \ldots, W_2 e_p, \ldots, W_{(p-1)} e_{(p-2)}, W_{(p-1)} e_{(p-1)}, W_p e_p\right]^T$$

REFERENCES

Abramowitz, Milton and Irene A. Stegun, 1970, Handbook of Mathematical Functions, Ninth Printing. Government Printing Office.

Apostol, Tom, M., 1974, Mathematical Analysis, Second Edition. (Addison-Wesley Publishing, Reading, Massachusetts.)

Cardell, S. and F. Dunbar 1980, Measuring the societal impacts of automobile downsizing, Tranportation Research A 14, 423–434.

DeGroot, M. H., 1959, Unbiased Binomial Sequential Estimation, Annals of Mathematical Statistics, 30, 80-101

Genz, Alan, 1992, Numerical Computation of Multivariate Normal Probabilities, Journal Computational and Graphical Statistics, 1, 141-149

Gouriéroux, C. and A. Monfort, 1996, Simulation Based Econometric Methods, (Oxford University Press, London).

Govindarajulu, Zakkulu, 1987, The Sequential Statistical Analysis of Hypothesis Testing, Point and Interval Estimation, and Decision Theory, (American Series in Mathematical and Management Sciences, American Sciences Press, Inc., Columbus, OH)

Hajivassiliou, Vassilis A. and Daniel L. McFadden, 1998, The Method of Simulated Scores for the Estimation of LDV Models. Econometrica, Vol. 66, No., 863-896.

Hardy, G. H. Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed. New York: Chelsea,1999.

Hogg, Robert V., and Allen T. Craig, 1978, Introduction to Mathematical Statistics, Fourth Edition, (Macmillan Publishing Company Inc., New York).

Johnson, Norman L., Samuel Kotz, and Adrienne W.Kemp, 1993,Univariate Discrete Distributions. (John Wiley and Sons Inc., New York).

Judd, Kenneth, 1998, Numerical Methods in Economics. (MIT Press, Cambridge).

Kendall, Maurice and Alan Stuart, 1977, The Advanced Theory of Statistics, Volume 1, Fourth Edition. (Macmillian Publishing Company Inc., New York).

Kolmogorov, A.N. and S.V.Fomin, 1975, Introductory Real Analysis. (Dover, New York,NY.)

Knopp, Konrad, 1990, Theory and Application of Infinite Series. (Dover, Mineola, NY).

Krantz, Steven G. and Harold R. Parks, 1992, A Primer of Real Analytic Functions. (Basel, Boston).

Lee, Myoung-jae, 1996, Method of Moments and Semiparametric Econometrics for Limited Dependent Variable Models. (Springer-Verlag, New York).

Lehmann, E. L., 1998, Elements of Large-Sample Theory. (Springer, New York).

Lehmann, E. L. and George Casella, 1998, Theory of Point Estimation. Second Edition. (Springer, New York).

McFadden, Daniel, 1989, A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration, Econometrica 57, 995-1026.

Mood, Alexander M., Franklin A. Graybill and Duane C. Boes, 1974, Introduction to the Theory of Statistics, Third Edition, (McGraw-Hill, New York).
Newey, Whitney K., 1991, Uniform Convergence in Probability and Stochastic Equicontinuity, Econometrica, 59, 1161-1167.

Randles, Ronald H. and Douglas A. Wolfe, 1979, Introduction to the Theory of Nonparametric Statistics, (John Wiley and Sons, Inc., New York).

Rosen, Kenneth H., 1995, Discrete Mathematics and Its Applications, Third Edition, (McGraw-Hill, Inc., New York).

Ross, Sheldon, M., 1992, Applied Probability Models with Optimization Applications, Dover Publications, New York.

Ruud, P. A., 2000, An Introduction to Classical Econometric Theory, Oxford University Press, New York.

Stern, Steven, Simulation Based Inference in Econometrics: Motivation and Methods, 2000, in Simulation-Based Inference: Econometrics: Methods and Applications, eds., Roberto S. Mariano, Melvyn Weeks, and Til Schuermann, (Cambridge University Press, Cambridge) 9-40.

Train, Kenneth, 2002, Discrete Choice Methods with Simulation, (Cambridge University Press, Cambridge).

Weisstein, Eric W. Ermakoff's Test, (from MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/ErmakoffsTest.html)

Wolfram Research, Inc., 2001, Mathematica, Version 4.1, (Wolfram Research, Inc., Champaign, IL).