On Unification of the Asymptotic Theory of Nonlinear Econometric Models

by

Jose F. Burguete and A. Ronald Gallant*

Institute of Statistics, North Carolina State University,

Raleigh, NC 27650 USA

*Presented at the 1980 World Congress of the Econometric Society. This work was supported in part by North Carolina Agricultural Experiment Station Project No. NC03641.

ABSTRACT

The majority of the estimators which have been proposed for nonlinear econometric models are obtained as the solution of an optimization problem. Examples include single equation nonlinear least squares, minimum distance and maximum likelihood estimators for nonlinear multivariate regression, and two-stage least squares, three-stage least squares, and maximum likelihood estimators for nonlinear simultaneous systems of equations. In the paper, an optimization problem is proposed which encomposes all of these examples. The almost sure limit and the asymptotic distribution of the solution of the optimization problem are found for data generated according to the general nonlinear model $q(y,x,\gamma) = e$. The majority of the inference procedures used with nonlinear models may be obtained by treating the objective function of the optimization problem as if it were the likelihood and deriving the Wald test statistic, the likelihood ratio test statistic, and Rao's efficient score test statistic. The null and non-null asymptotic distributions of these statistics are derived. To obtain an asymptotic theory for a nonlinear model with these results, the appropriate objective function is identified and the asymptotic theory obtains at once by direct computation; several examples are included in the paper. Since the model which motivated the optimization problem need not be the same as the model which generates the data, these results may be used to obtain the asymptotic behavior of inference procedures under model misspecification.

The Hartley-Booker (1965) estimator is, to our best knowledge, the first use of the method of moments per se in nonlinear statistical models. Their method was proposed for the univariate response nonlinear model

$$y_t = f(x_t, \theta^*) + e_t$$

where θ^* is an unknown p-vector. The space $\mathcal{X}$ of possible values for the sequence $\{x_t\}$ is divided into p disjoint sets $\mathcal{X}_i$. The moment equations

$$\Sigma_{x_t \in \mathcal{X}_i} y_t = \Sigma_{x_t \in \mathcal{X}_i} f(x_t, \theta) \qquad i = 1, 2, \ldots, p$$

are computed and solved to obtain an estimator $\hat{\theta}$. They used it as the first step of a scoring method but we consider it as an estimator in its own right.

From our point of view, a handier notation results by letting

$$z_t = e_i \qquad \text{if} \quad x_t \in \mathcal{X}_i$$

where e_i is the i-th elementary p-vector. The moment equations are now written as

$$m_n(\theta) = (1/n) \Sigma_{t=1}^{n} z_t[y_t - f(x_t, \theta)] \ .$$

The Hartley-Booker estimator is, then, the solution of $m_n(\theta) = 0$.

A problem with this approach is that the equations $m_n(\theta) = 0$ may not have a solution. This problem is eliminated by defining $\hat{\theta}$ to be the maximum of

$$s_n(\theta) = -\tfrac{1}{2} m_n'(\theta) \, m_n(\theta) \ .$$

That is, redefine the estimator as the solution of an optimization problem whose first order conditions imply $m_n(\theta) = 0$ when the moment equations can be solved.

This formulation of the Hartley-Booker estimator eliminates the need to restrict the number of disjoint subsets of $\mathcal{X}$ to exactly p. The vectors z_t

of the moment equations

$$m_n(\theta) = (1/n) \Sigma_{t=1}^{n} z_t[y_t - f(x_t,\theta)]$$

may have length greater than p . But in this case, one can argue by analogy to generalized least squares that an optimization problem with objective function

$$s_n(\theta) = -\tfrac{1}{2} m_n'(\theta) [(1/n) \Sigma_{t=1}^{n} z_t z_t']^{-1} m_n(\theta)$$

will yield more efficient estimators. One notes that this is the optimization problem which defines the two-stage nonlinear least-squares estimator (Amemiya, 1974). Only the restriction that z_t be chosen according as $x_t \in \mathfrak{X}_i$ or not prevents the modified Hartley-Booker estimator from being properly considered a two-stage nonlinear least-squares estimator.

These remarks motivate a general definition of the method of moments estimator. To permit consideration of iteratively rescaled estimators such as three-stage nonlinear least squares, both the moment equations

$$m_n(\lambda) = (1/n) \Sigma_{t=1}^{n} m(y_t, x_t, \hat{\tau}_n , \lambda)$$

and the objective function

$$s_n(\lambda) = d[m_n(\lambda), \hat{\tau}_n]$$

of the optimization problem are permitted to depend on a random variable $\hat{\tau}_n$ via the argument τ in $m(y,x,\tau,\lambda)$ and in the distance function $d[m,\tau]$.

In this paper, the asymptotic distribution of an estimator defined as that $\hat{\lambda}_n$ which maximizes $s_n(\lambda)$ is found for data generated according to the multivariate nonlinear model

$$q(y_t, x_t, \gamma_n^{\circ}) = e_t \quad .$$

Then $s_n(\lambda)$ is treated as if it were the likelihood for the purpose of deriving the Wald test statistic, the likelihood ratio test statistic, and Rao's efficient score test statistic. The null and non-null asymptotic distributions of these statistics are derived.

Estimators which are properly thought of as method of moment estimators, in the sense that they can be posed no other way, are: The Hartley-Booker estimator - Hartley and Booker (1965). Scale invariant M-estimators - Ruskin (1978). Two-stage least-squares estimators - Amemiya (1974). Three-stage least-squares estimators - Jorgenson and Laffont (1974), Amemiya (1977), Gallant and Jorgenson (1979).

A second group of estimators, termed M-estimators here, are of the form

$$\hat{\lambda}_n \text{ maximizes } s_n(\lambda) = (1/n)\Sigma_{t=1}^{n} s(y_t, x_t, \hat{\tau}_n, \lambda) \quad .$$

They can be cast into the form of method of moments estimators by putting

$$m_n(\lambda) = (1/n)\Sigma_{t=1}^{n}(\partial/\partial\lambda)s(y_t, x_t, \hat{\tau}_n, \lambda)$$

and $d[m,\tau] = -\frac{1}{2}\, m'm$. This second group is: Single equation nonlinear least-squares - Jennrich (1969), Malinvaud (1970), Gallant (1973, 1975a, 1975b). Multivariate least-squares - Malinvaud (1970b), Gallant (1975c), Holly (1978). Single equation and multivariate maximum likelihood - Malinvaud (1970b), Barnett (1976), Holly (1978). Maximum likelihood for simultaneous systems - Amemiya (1977), Gallant and Holly (1980). M-estimators - Balet-Lawrence (1975), Grossman (1976), Ruskin (1978). Iteratively rescaled M-estimates - Souza and Gallant (1979).

If one's only interest is to find the asymptotic distribution of the estimator, then posing the problem as a method of moments estimator is the more convenient approach. One pays two penalties. The first, the problem is no longer posed in a way that permits the use of the likelihood ratio test. The second, the consistency results are weaker. With the method of moments approach one can prove the existence of a consistent estimator which solves $(\partial/\partial\lambda)s_n(\lambda) = 0$. With the M-estimator approach, one can prove that that $\hat{\lambda}_n$ which maximizes $s_n(\lambda)$ converges almost surely. For the reader's convenience, we include a précis of these stronger results.

2. PRELIMINARIES

The M-variate responses y_t are generated according to

$$q(y_t, x_t, \gamma_n^\circ) = e_t \qquad t = 1, 2, \ldots, n$$

with $x_t \in \mathcal{X}$, $y_t \in \mathcal{Y}$, $e_t \in \mathcal{E}$, and $\gamma_n^\circ \in \Gamma$. The sequence $\{y_t\}$ is actually doubly indexed as $\{y_{tn}\}$ due to the drift of γ_n° with n; the sequences $\{e_t\}$ and $\{x_t\}$ are singly indexed and the analysis is conditional on $\{x_t\}$ throughout.

<u>Assumption 1.</u> The errors are independently and identically distributed with common distribution $P(e)$.

Obviously, for the model to make sense, some measure of central tendency of $P(e)$ ought to be zero but no formal use is made of such an assumption. If $P(e)$ is indexed by parameters, they cannot drift with sample size as may γ_n°.

The models envisaged here are supposed to describe the behavior of a physical, biological, economic, or social system. If so, to each value of (e, x, γ°) there should correspond one and only one outcome y. This condition and continuity are imposed.

<u>Assumption 2</u>. For each $(x, \gamma) \in \mathcal{X} \times \Gamma$ the equation $q(y, x, \gamma) = e$ defines a one-to-one mapping of $\mathcal{E}$ onto $\mathcal{Y}$ denoted as $Y(e, x, \gamma)$. Moreover, $Y(e, x, \gamma)$ is continuous on $\mathcal{E} \times \mathcal{X} \times \Gamma$.

It should be emphasized that it is not necessary to have a closed form expression for $Y(e, x, \gamma)$, or even to be able to compute it using numerical methods, in order to use the statistical methods set forth here.

Repeatedly, in the sequel, the uniform limit of a Cesaro sum such as $(1/n)\Sigma_{t=1}^{n} f(y_t, x_t, \gamma)$ is required. In the nonlinear regression literature much attention has been devoted to finding conditions which insure this behavior yet are plausible and can be easily recognized as obtaining or not obtaining in an application (Jennrich, 1969; Malinvaud, 1970a; Gallant, 1977; Gallant and Holly, 1980). Details and examples may be found in these references; we follow Gallant and Holly (1980).

Definition. (Gallant and Holly, 1980) A sequence $\{v_t\}$ of points from a Borel set $\mathcal{V}$ is said to be a Cesaro sum generator with respect to a probability measure ν defined on the Borel subsets of $\mathcal{V}$ and a dominating function $b(v)$ with $\int b \, d\nu < \infty$ if

$$\lim_{n\to\infty}(1/n) \Sigma_{t=1}^{n} f(v_t) = \int f(v) \, d\nu(v)$$

for every real valued, continuous function f with $|f(v)| \leq b(v)$.

Assumption 3. (Gallant and Holly, 1980) Almost every realization of $\{v_t\}$ with $v_t = (e_t, x_t)$ is a Cesaro sum generator with respect to the product measure $\nu(A) = \int_{\mathcal{X}}\int_{\mathcal{E}} I_A(e,x) \, dP(e) \, d\mu(x)$ and a dominating function $b(e,x)$. The sequence $\{x_t\}$ is a Cesaro sum generator with respect to μ and $b(x) = \int_{\mathcal{E}} b(e,x) dP(e)$. For each $x \in \mathcal{X}$ there is a neighborhood N_x such that $\int_{\mathcal{E}} \sup_{N_x} b(e,x) \, dP(E) < \infty$.

Theorem 1. (Gallant and Holly, 1980) Let Assumptions 1 through 3 hold. Let $f(y,x,\rho)$ be continuous on $\mathcal{Y} \times \mathcal{X} \times K$ where K is compact. Let $|f(y,x,\rho)| \leq b[q(y,x,\gamma),x]$ for all $(y,x) \in \mathcal{Y} \times \mathcal{X}$ and all (ρ,γ) in $K \times \Lambda$ where Λ is compact. Then both $(1/n) \Sigma_{t=1}^{n} f(y_t, x_t, \rho)$ and $(1/n) \Sigma_{t=1}^{n} \int_{\mathcal{E}} f[Y(e,x_t,\gamma),x_t,\rho] \, dP(e)$ converge uniformly to

$$\int_{\mathcal{X}}\int_{\mathcal{E}} f[Y(e,x,\gamma),x,\rho] \, dP(e) \, d\mu(x)$$

except on the event E with $P^*(E) = 0$ given by Assumption 3.

In typical applications, a density $p(e)$ and a Jacobian

$$J(y,x,\gamma^\circ) = (\partial/\partial y')q(y,x,\gamma^\circ)$$

are available. With these in hand, the conditional density

$$p(y|x,\gamma^\circ) = |\det J(y,x,\gamma^\circ)| p[q(y,x,\gamma^\circ)]$$

may be used for computing limits since

$$\int_{\mathcal{X}}\int_{\mathcal{E}} f[Y(e,x,\gamma^{\circ}),x,\gamma]\ dP(e)\ d\mu(x) = \int_{\mathcal{X}}\int_{\mathcal{Y}} f(y,x,\gamma)\ p(y|x,\gamma^{\circ})\ dy\ d\mu(x)\ .$$

The choice of integration formulas is dictated by convenience.

3. METHOD OF MOMENTS ESTIMATORS

Consider the moment equations

$$m_n(\lambda) = (1/n)\Sigma_{t=1}^{n} m(y_t, x_t, \hat{\tau}_n, \lambda)$$

where $\hat{\tau}_n$ is a random variable with almost sure limit τ^*. Suppose that there is a natural association of λ to γ, say $\lambda = g(\gamma)$, which solves

$$\int_{\mathcal{E}} m[Y(e,x,\gamma),x,\tau^*,\lambda]\, dP(e) = 0 ,$$

for all x. The classical method of moments procedure is to equate sample moments to their expectation

$$m_n(\lambda) = 0$$

and solve the equations for λ. These equations may not have a solution. To eliminate this problem one may reason by analogy with regression methods and maximize, say, $-\frac{1}{2} m_n'(\lambda)\, m_n(\lambda)$ to find an estimator. In general, consider $\hat{\lambda}_n$ maximizing

$$d[m_n(\lambda), \hat{\tau}_n]$$

where $d[m,\tau]$ is some measure of distance with $d[o,\tau] = 0$ and $d[m,\tau] < 0$ for $m \neq 0$. The constrained method of moments estimator $\tilde{\lambda}_n$ is the solution of the optimization problem

$$\text{Maximize:} \quad d[m_n(\lambda), \hat{\tau}_n] \quad \text{subject to } h(\lambda) = 0 \quad .$$

The assumptions are somewhat abstract due to the scope of applications envisaged. As a counterbalance, an example is carried throughout this section. The best choice of an example seems to be a robust, scale-invariant, M-estimator for the univariate model

$$y_t = f(x_t, \gamma_n^\circ) + e_t$$

due to both its intrinsic interest and freedom from tedious notational details.

The error distribution $P(e)$ for the example is assumed to be symmetric with $\int_{\mathcal{E}} |e| dP(e)$ finite and $\int_{\mathcal{E}} e^2 dP(e) > 0$. The reduced form is

$$Y(e,x,\gamma) = f(x,\gamma) + e \ .$$

Proposal 2 of Huber (1964) leads to the moment equations

$$m_n(\lambda) = (1/n)\Sigma_{t=1}^{n} \begin{pmatrix} \Psi\{[y_t - f(x_t,\theta)]/\sigma\}\,(\partial/\partial\theta)\, f(x_t,\theta) \\ \Psi^2\{[y_t - f(x_t,\theta)]/\sigma\} - \beta \end{pmatrix}$$

with $\lambda = (\theta',\sigma)'$. For specifity let

$$\Psi(u) = \tfrac{1}{2} \tanh\,(u/2) \ ,$$

a bounded odd function with bounded even derivative and let

$$\beta = \int \Psi^2(e)\ d\,\Phi(e) \ .$$

There is no previous estimator $\hat{\tau}_n$ with this example so the argument τ of $m(y,x,\tau,\lambda)$ is suppressed to obtain

$$m(y,x,\lambda) = \begin{pmatrix} \Psi\{[y - f(x,\theta)]/\sigma\}(\partial/\partial\theta)\ f(x,\theta) \\ \Psi^2\{[y - f(x,\theta)]/\sigma\} - \beta \end{pmatrix} \ .$$

The distance function is

$$d(m) = -\tfrac{1}{2}m'm ,$$

again suppressing the argument τ , whence the estimator $\hat{\lambda}_n$ is defined as that value of λ which maximizes

$$s_n(\lambda) = -\tfrac{1}{2}\, m_n'(\lambda)\, m_n(\lambda) .$$

Notation

$$m_n(\lambda) = (1/n) \textstyle\sum_{t=1}^{n} m(y_t, x_t, \hat{\tau}_n, \lambda)$$

$$\bar{m}(\gamma,\tau,\lambda) = \int_{\mathcal{X}}\int_{\mathcal{E}} m[Y(e,x,\gamma),x,\tau,\lambda]\; dP(e)\; d\mu(x)$$

$$s_n(\lambda) = d[m_n(\lambda), \hat{\tau}_n]$$

$$\bar{s}(\gamma,\tau,\lambda) = d[\bar{m}(\gamma,\tau,\lambda),\tau]$$

The identification condition is

Assumption 4. The sequence γ_n° converges to a point γ^* . The sequence $\hat{\tau}_n$ converges almost surely to a point τ^* and $\sqrt{n}(\hat{\tau}_n - \tau^*)$ is bounded in probability. There is an association of λ to γ , denoted as $\lambda = g(\gamma)$, which satisfies

$$\bar{m}[\gamma, \tau^*, g(\gamma)] = 0 .$$

The sequence $\lambda_n^\circ = g(\gamma_n^\circ)$ has $\lim_{n\to\infty}\sqrt{n}(\lambda_n^\circ - \lambda^*) = \delta$ where $\lambda^* = g(\gamma^*)$ and δ is finite. The constraint $h(\lambda) = 0$ is satisfied at λ^* .

For the example, let σ^* solve $\int_{\mathcal{E}} \Psi^2(e/\sigma)\, dP(e) = \beta$, a solution exists since $G(\sigma) = 1 - \int_{\mathcal{E}} \Psi(e/\sigma)\, dP(e)$ is a continuous distribution function if $P(e)$ does not put all its mass at zero. Define $g(\gamma) = (\gamma, \sigma^*)$. Then

$$\int_{\mathcal{E}} m[e + f(x,\gamma),x,\ (\gamma,\sigma^*)]\; dP(e)$$

$$= \begin{pmatrix} \int_{\mathcal{E}} \Psi(e/\sigma^*)\, dP(e)\ (\partial/\partial\theta)\, f(x,\gamma) \\ \int_{\mathcal{E}} \Psi^2(e/\sigma^*)\, dP(e) - \beta \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ 0 \end{pmatrix} .$$

As the integral is zero for every x , integration over $\mathcal{X}$ with respect to μ must yield

$$\bar{m}[\gamma, g(\gamma)] = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

as required by Assumption 4.

Notation

$$S = \int_{\mathcal{X}}\int_{\mathcal{E}} m[Y(e,x,\gamma^*),x,\tau^*,\lambda^*]\ m'[Y(e,x,\gamma^*),x,\tau^*,\lambda^*]\ dP(e)\ d\mu(x)$$

$$M = \int_{\mathcal{X}}\int_{\mathcal{E}} (\partial/\partial\lambda')m[Y(e,x,\gamma^*),x,\tau^*,\lambda^*]\ dP(e)\ d\mu(x)$$

$$D = (\partial^2/\partial m \partial m')\ d(0,\tau^*)$$

$$S_n(\lambda) = (1/n)\Sigma_{t=1}^{n} m(y_t,x_t,\hat{\tau}_n,\lambda)\ m'(y_t,x_t,\hat{\tau}_n,\lambda)$$

$$M_n(\lambda) = (1/n)\Sigma_{t=1}^{n} (\partial/\partial\lambda')m(y_t,x_t,\hat{\tau}_n,\lambda)$$

$$D_n(\lambda) = (\partial^2/\partial m \partial m')\ d[m_n(\lambda),\hat{\tau}_n]$$

$$\mathcal{I} = M'\ D\ S\ D\ M$$

$$\mathcal{J} = -M'\ D\ M$$

$$\mathcal{I}_n(\lambda) = M_n'(\lambda)\ D_n(\lambda)\ S_n(\lambda)\ D_n(\lambda)\ M_n(\lambda)$$

$$\mathcal{J}_n(\lambda) = -M_n'(\lambda)\ D_n(\lambda)\ M_n(\lambda)$$

$$H = (\partial/\partial\lambda')\ h(\lambda^*)$$

$$H(\lambda) = (\partial/\partial\lambda')\ h(\lambda)$$

For the example, direct computation yields

$$S = \begin{pmatrix} \int_{\mathcal{E}} \Psi^2(e/\sigma^*)\, dP(e)\, F'F & 0 \\ 0 & \int_{\mathcal{E}} [\Psi^2(e/\sigma^*) - \beta]^2\, dP(e) \end{pmatrix}$$

$$M = \begin{pmatrix} -(1/\sigma^*)\int_{\mathcal{E}} \Psi'(e/\sigma^*)\, dP(e)\, F'F & 0 \\ 0 & -2(1/\sigma^*)^2 \int_{\mathcal{E}} \Psi(e/\sigma^*)\Psi'(e/\sigma^*)\, e\, dP(e) \end{pmatrix}$$

$$D = -I$$

where

$$F'F = \int_{\mathcal{X}} (\partial/\partial\theta)\, f(x,\theta)\ (\partial/\partial\theta')\, f(x,\theta)\, d\mu(x)\Big|_{\theta=\gamma^*} .$$

This computation exploits the fact that $\Psi(e/\sigma^*)$, e are odd and $\Psi'(e/\sigma^*)$, $\Psi^2(e/\sigma^*)$ are even. If $P(e)$ does not put all its mass at zero and $F'F$ is non-singular then S, M, and D have full rank by inspection.

Assumption 5. There are bounded, open spheres Γ, T, Λ containing γ^*, τ^*, λ^* for which the elements of $m(y,x,\tau,\lambda)$, $(\partial/\partial\lambda_i)\, m(y,x,\tau,\lambda)$, $(\partial^2/\partial\lambda_i\partial\lambda_j)\, m(y,x,\tau,\lambda)$ are continuous and dominated by $b[q(y,x,\gamma),x]$ on $\mathcal{Y} \times \mathcal{X} \times \bar{T} \times \bar{\Lambda} \times \bar{\Gamma}$; $b(e,x)$ is that of Assumption 3 and the overbar indicates closure of a set. The distance function $d(m,\tau)$ and derivatives $(\partial/\partial m)\, d(m,\tau)$, $(\partial^2/\partial m\, \partial m')\, d(m,\tau)$ are continuous on $\bar{\mathfrak{S}} \times \bar{T}$ where $\mathfrak{S}$ is some open sphere containing the zero vector. The constraining function $h(\lambda)$ and its derivative $H(\lambda)$ are continuous on $\bar{\Lambda}$. The matrix D is negative definite, $(\partial/\partial m)\, d(0,\tau) = 0$ for all τ , and M, H have full rank.

To illustrate the construction of $b(e,x)$, consider for the example

$$\|m_{(1)}(y,x,\lambda)\| = |\Psi\{[y - f(x,\theta)]/\sigma\}| \cdot \|(\partial/\partial\theta)\, f(x,\theta)\|$$
$$\le \|(\partial/\partial\theta)\, f(x,\theta)\|$$

because $|\Psi(u)| = |\tfrac{1}{2} \tanh\, (u/2)| \le \tfrac{1}{2}$. What is required then is that

$\sup_\theta \|(\partial/\partial\theta)\, f(x,\theta)\|$ be integrable with respect to μ. Or, since $\bar{\Lambda}$ is compact, $(\partial/\partial\theta)\, f(x,\theta)$ continuous in (x,θ) and $\mathfrak{X}$ compact would bound $\|(\partial/\partial\theta) f(x,\theta)\|$ in which case $b_i(e,x) = \text{const.}$ One accumulates $b_i(e,x)$ in this fashion to satisfy the assumptions. Then $b(e,x)$ of Assumption 3 is $b(e,x) = \Sigma\, b_i(e,x)$. Because $\Psi(u)$ and its derivatives are bounded, this construction of $b(e,x)$ is not very interesting. More interesting, and detailed, constructions are given in Gallant and Holly (1980).

Theorem 2. (Consistency) Let Assumptions 1 through 5 hold. There is a sequence $\{\hat{\lambda}_n\}$ such that for almost every realization of $\{e_t\}$, $\lim_{n\to\infty}\hat{\lambda}_n = \lambda^*$ and there is an N such that $(\partial/\partial\lambda)\; s_n(\hat{\lambda}_n) = 0$ for $n > N$. Similarly, there is a sequence $\tilde{\lambda}_n$ and associated Lagrange multipliers $\tilde{\theta}_n$ such that $\lim_{n\to\infty}\tilde{\lambda}_n = \lambda^*$ and $(\partial/\partial\lambda)[s_n(\tilde{\lambda}_n) + \tilde{\theta}_n' h(\tilde{\lambda}_n)] = 0$, $h(\tilde{\lambda}_n) = 0$ for $n > N$.

Proof: The result will be proved for $\tilde{\lambda}_n$. Fix a sequence $\{e_t\} \notin E$, this fixes $\hat{\tau}_n$.

$$(\partial/\partial\lambda_i)\; s_n(\lambda) = \Sigma_\alpha (\partial/\partial m_\alpha)\; d[m_n(\lambda),\hat{\tau}_n](\partial/\partial\lambda_i) m_{\alpha n}(\lambda)\,,$$

$$(\partial^2/\partial\lambda_i\partial\lambda_j)\; s_n(\lambda) = \Sigma_\alpha\Sigma_\beta (\partial^2/\partial m_\alpha \partial m_\beta)\; d[m_n(\lambda),\hat{\tau}_n](\partial/\partial\lambda_i) m_{\alpha n}(\lambda)(\partial/\partial\lambda_j) m_{\beta n}(\lambda)$$

$$+ \Sigma_\alpha (\partial/\partial m_\alpha) d[m_n(\lambda),\hat{\tau}_n](\partial^2/\partial\lambda_i\partial\lambda_j) m_{\alpha n}(\lambda) \quad .$$

The assumptions suffice for an application of Theorem 1 and the conclusion that $m_n(\lambda)$, $(\partial/\partial\lambda_i)\; m_n(\lambda)$, and $(\partial^2/\partial\lambda_i\partial\lambda_j) m_n(\lambda)$ converge uniformly on Λ to $\bar{m}(\gamma^*,\tau^*,\lambda)$, $(\partial/\partial\lambda_i)\bar{m}(\gamma^*,\tau^*,\lambda)$, and $(\partial^2/\partial\lambda_i\partial\lambda_j)\bar{m}(\gamma^*,\tau^*,\lambda)$; the domination required to apply Theorem 1 permits the interchange of differentiation and integration as needed. Since $\bar{m}(\gamma^*,\tau^*, \lambda^*) = 0$, one can shrink the radius of Λ to Λ' so that $m_n(\lambda) \in \mathfrak{S}$ for all $\lambda \in \Lambda'$ and n suitably large whence $s_n(\lambda)$,

$(\partial/\partial\lambda)s_n(\lambda)$ and $(\partial^2/\partial\lambda\partial\lambda')s_n(\lambda)$ converge uniformly on Λ' to $\bar{s}(\gamma^*,\tau^*,\lambda)$, $(\partial/\partial\lambda)\ \bar{s}(\gamma^*,\tau^*,\lambda)$, and $(\partial^2/\partial\lambda\partial\lambda')\ \bar{s}(\gamma^*,\tau^*,\lambda)$ respectively. As $(\partial/\partial m)\ d[0,\tau^*] = 0$ and $(\partial^2/\partial m\partial m')\ d[0,\tau^*]$ is negative definite, $(\partial/\partial\lambda)\bar{s}(\gamma^*,\tau^*,\lambda^*) = 0$ and $(\partial^2/\partial\lambda\partial\lambda')\bar{s}(\gamma^*,\tau^*,\lambda^*)$ is negative definite. Thus, one may shrink the radius of Λ' to Λ'' so that $\bar{s}(\gamma^*,\tau^*,\lambda)$ has a unique maximum at $\lambda = \lambda^*$ on Λ''.

Let $\tilde{\lambda}_n$ maximize $s_n(\lambda)$ subject to $h(\lambda) = 0$ and $\lambda \in \Lambda''$. Now $h(\lambda^*) = 0$ and $s_n(\lambda)$ converges uniformly to $\bar{s}(\gamma^*,\tau^*,\lambda)$ on Λ'' so that for large n the solution $\tilde{\lambda}_n$ cannot lie on the boundary of Λ''. The existence of the Lagrange multipliers and satisfaction of the first order conditions follows.

As Λ'' is compact, $\tilde{\lambda}_n$ has at least one limit point $\dot{\lambda}$; let $\tilde{\lambda}_{n_m}$ converge to $\dot{\lambda}$. Then, by uniform convergence,

$$\begin{aligned}\bar{s}(\gamma^*,\tau^*,\dot{\lambda}) &= \lim_{n\to\infty} s_{n_m}(\gamma^o_{n_m}, \hat{\tau}_{n_m}, \tilde{\lambda}_{n_m}) \\ &\geq \lim_{n\to\infty} s_{n_m}(\gamma^o_{n_m}, \hat{\tau}_{n_m}, \lambda^*) \\ &= \bar{s}(\gamma^*, \tau^*, \lambda^*) .\end{aligned}$$

But λ^* is the unique maximum of $\bar{s}(\gamma^*, \tau^*, \lambda)$ on Λ'' whence $\dot{\lambda} = \lambda^*$. □

One may note that the domination in Assumption 5 suffices for several interchanges of integration and differentiation. One consequence is that

$$M = (\partial/\partial\lambda')\ \bar{m}(\gamma^*,\tau^*,\lambda^*)$$

whence, since $\bar{m}(\gamma^*,\tau^*,\lambda^*) = 0$ and $(\partial/\partial m)d(0,\tau) = 0$,

$$\mathcal{J} = -(\partial^2/\partial\lambda\partial\lambda')\bar{s}(\gamma^*,\tau^*,\lambda^*) .$$

<u>Assumption 6</u>. The elements of $m(y,x,\tau,\lambda)\ m'(y,x,\tau,\lambda)$ and $(\partial/\partial\tau')m(y,x,\tau,\lambda)$ are continuous and dominated by $b[q(y,x,\gamma),x]$ on $\mathcal{Y} \times \mathcal{X} \times \bar{T} \times \bar{\Lambda} \times \bar{\Gamma}$; $b(e,x)$ is that of Assumption 3. The elements of $(\partial^2/\partial\tau\partial m')d(m,\tau)$ are continuous on $\bar{\mathcal{M}} \times \bar{T}$ where $\mathcal{M}$ is some open sphere containing the zero vector

$$\int_{\mathcal{E}} m[Y(e,x,\gamma_n^\circ),x,\tau^*,\lambda_n^\circ]\, dP(e) = 0$$

$$\int_{\mathcal{X}}\int_{\mathcal{E}} (\partial/\partial\tau')m[Y(e,x,\gamma^*),x,\tau^*,\lambda^*]\, dP(e)\, d\mu(x) = 0 \quad .$$

The first integral condition is central to our results and is apparently an intrinsic property of reasonable estimation procedures. It was verified for the example as an intermediate step in the verification of Assumption 4.

The second integral condition is sometimes encountered in the theory of maximum likelihood estimation; see Durbin (1970) for a detailed discussion. It validates the application of maximum likelihood theory to a subset of the parameters when the remainder are treated as if known in the derivations but are subsequently estimated. The assumption plays the same role here. It can be avoided in maximum likelihood estimation at a cost of additional complexity in the results; see Gallant and Holly (1980) for details. It can probably be avoided here but there is no reason to further complicate the results in view of the intended applications. For the example, there is no dependence on τ hence nothing to verify. Had an iteritively rescaled estimator been considered, $m(y,x,\tau,\lambda) = \Psi\{[y - f(x,\theta)]/\tau\}(\partial/\partial\theta)f(x,\theta)$ with $\hat{\tau}_n$ supplied by a previous fit, the condition would have been satisfied as the off-diagonal corner of our previously computed M is zero for any σ^* .

<u>Theorem 3</u>. (Asymptotic Normality of the Moments) Under Assumptions 1 through 6

$$\sqrt{n}\, m_n(\lambda_n^\circ) \xrightarrow{\mathcal{L}} N(0,S)$$

$$\sqrt{n}\, m_n(\lambda^*) \xrightarrow{\mathcal{L}} N(-M\,\delta,\, S)$$

S may be singular.

<u>Proof</u>. Given ℓ with $\|\ell\| = 1$ consider the triangular array of random variables

$$Z_{tn} = \ell' m[Y(e_t,x_t,\gamma_n^o),x_t,\tau^*,\lambda_n^o] \quad t = 1, \ldots, n; \; n = 1, 2, \ldots \; .$$

Each Z_{tn} has mean, $\int_{\mathcal{E}} Z_{tn}(e)\, dP(e)$, zero by assumption and variance

$$\sigma_{tn}^2 = \ell' \int_{\mathcal{E}} m[Y(e,x_t,\gamma_n^o),x_t,\tau^*,\lambda_n^o]\, m'[Y(e,x_t,\gamma_n^o),x_t,\tau^*,\lambda_n^o]\, dP(e) \ell \; .$$

By Theorem 1 and the assumption that $\ell im_{n\to\infty}(\gamma_n^o,\lambda_n^o) = (\gamma^*,\lambda^*)$ it follows that $\ell im_{n\to\infty}(1/n)V_n = \ell' S\, \ell$ where

$$V_n = \Sigma_{t=1}^n \sigma_{tn}^2 \quad .$$

Now $(1/n)V_n$ is the variance of $(1/\sqrt{n})\Sigma_{t=1}^n Z_{tn}$ and if $\ell' S \ell = 0$ then $(1/\sqrt{n})\Sigma_{t=1}^n Z_{tn}$ converges in distribution to $N(0,\ell' S \ell)$ by Chebyshev's inequality.

Suppose, then, that $\ell' S \ell > 0$. If it is shown that for every $\varepsilon > 0$ $\ell im_{n\to\infty} B_n = 0$ where

$$B_n = (1/n)\Sigma_{t=1}^n \int_{\mathcal{E}} I_{[|z|>\varepsilon\sqrt{V_n}]}[Z_{tn}(e)] Z_{tn}^2(e)\, dP(e)$$

then $\ell im_{n\to\infty}(n/V_n)B_n = 0$. This is the Lindberg-Feller condition (Chung, 1974); it implies that $(1/\sqrt{n})\Sigma_{t=1}^n Z_{tn}$ converges in distribution to $N(0,\ell' S \ell)$.

Let $\eta > 0$ and $\varepsilon > 0$ be given. Choose $a > 0$ such that $\bar{B}(\gamma^*,\lambda^*) < \eta/2$ where

$$\bar{B}(\gamma^*,\lambda^*) = \int_{\mathcal{X}} \int_{\mathcal{E}} I_{[|z|>\varepsilon a]}\{\ell' m[Y(e,x,\gamma^*),x,\tau^*,\lambda^*]\}$$

$$\times \{\ell' m[Y(e,x,\gamma^*),x,\tau^*,\lambda^*]\}^2\, dP(e)\, d\mu(x) \quad .$$

This is possible because $\bar{B}(\gamma^*,\lambda^*)$ exists when $a = 0$. Choose a continuous function $\varphi(z)$ and an N_1 such that, for all $n > N_1$,

$$I_{[|z|>\varepsilon\sqrt{V_n}]}(z) \le \varphi(z) \le I_{[|z|>\varepsilon a]}(z)$$

and set

$$\tilde{B}_n(\gamma,\lambda) = (1/n)\Sigma_{t=1}^n \int_{\mathcal{E}} \varphi\{\ell' m[Y(e,x,\gamma),x_t,\tau^*,\lambda]\}$$

$$\times \{\ell' m[Y(e,x,\gamma),x_t,\tau^*,\lambda]\}^2 dP(e) .$$

By Theorem 1, $\tilde{B}_n(\gamma,\lambda)$ converges uniformly on $\bar{\Gamma}^* \times \bar{\Lambda}^*$ to, say, $\tilde{B}(\gamma,\lambda)$. By assumption $\lim_{n\to\infty}(\gamma_n^o,\lambda_n^o) = (\gamma^*,\lambda^*)$ whence $\lim_{n\to\infty}\tilde{B}_n(\gamma_n^o,\lambda_n^o) = \tilde{B}(\gamma^*,\lambda^*)$. Then there is an N_2 such that, for all $n > N_2$, $\tilde{B}_n(\gamma_n^o,\lambda_n^o) < \tilde{B}(\gamma^*,\lambda^*) + \eta/2$. But, for all $n > N = \max\{N_1,N_2\}, B_n \le \tilde{B}_n(\gamma_n^o,\lambda_n^o)$ whence

$$B_n \le \tilde{B}_n(\gamma_n^o,\lambda_n^o) < \tilde{B}(\gamma^*,\lambda^*) + \eta/2 \le \bar{B}(\gamma^*,\lambda^*) + \eta/2 \le \eta .$$

Now $\hat{\tau}_n$ is tail equivalent to a sequence contained in T . Thus, without loss of generality $\hat{\tau}_n$ may be taken to be in T and Taylor's theorem applied to obtain

$$(1/\sqrt{n})\Sigma_{t=1}^n Z_{tn} = (1/\sqrt{n})\ell'\Sigma_{t=1}^n m(y_t,x_t,\hat{\tau}_n,\lambda_n^o)$$

$$+ [(1/n)(\partial/\partial\tau')\ell'\Sigma_{t=1}^n m(y_t,x_t,\bar{\tau}_n,\lambda_n^o)]\sqrt{n}(\hat{\tau}_n - \tau^*)$$

where $\|\bar{\tau}_n - \tau^*\| \le \|\hat{\tau}_n - \tau^*\|$. By Theorem 1, the almost sure convergence of $\hat{\tau}_n$, and Assumption 6, the vector multiplying $\sqrt{n}(\hat{\tau}_n - \tau^*)$ converges almost surely to zero. This and the assumed probabil..ty bound on $\sqrt{n}(\hat{\tau}_n - \tau^*)$ imply that the last term converges in probability to zero whence $(1/\sqrt{n})\ell'\Sigma_{t=1}^n m(y_t,x_t,\hat{\tau}_n,\lambda_n^o) \xrightarrow{\mathcal{L}} N(0,\ell'\mathcal{S}\ell)$. This holds for every ℓ with $\|\ell\| = 1$ whence the first result obtains.

The sequence $(\gamma_n^o,\lambda_n^o,\hat{\tau}_n,\hat{\lambda}_n)$ converges almost surely to $(\gamma^*,\lambda^*,\tau^*,\lambda^*)$. It is then tail equivalent to a sequence with values in $\Gamma \times \Lambda \times T \times \Lambda$. Without loss of generality let $(\gamma_n^o,\lambda_n^o,\hat{\tau}_n,\hat{\lambda}_n) \in \Gamma \times \Lambda \times T \times \Lambda$. By Taylor's theorem and Theorem 1,

$$\sqrt{n}\ m_n(\lambda^*) = \sqrt{n}\ m_n(\lambda_n^o) + [M + o_s(1)]\ \sqrt{n}(\lambda^* - \lambda_n^o)$$

which establishes the second result as $\sqrt{n}(\lambda^* - \lambda_n^\circ) \to -\delta$ by assumption. □

Theorem 4. Let Assumptions 1 through 6 hold. Then

$$\sqrt{n}(\partial/\partial\lambda)s_n(\lambda^*) \xrightarrow{\mathcal{L}} N(\mathcal{J}\,\delta,\ \mathcal{I}) ,$$

$$\sqrt{n}(\hat{\lambda}_n - \lambda^*) \xrightarrow{\mathcal{L}} N(\delta,\ \mathcal{J}^{-1}\mathcal{I}\,\mathcal{J}^{-1}) .$$

$\mathcal{I}_n(\hat{\lambda}_n)$ converges almost surely to $\mathcal{I}$ and $\mathcal{J}_n(\hat{\lambda}_n)$ converges almost surely to $\mathcal{J}$.

Proof: By the almost sure convergence of $(\gamma_n^\circ,\lambda_n^\circ,\hat{\tau}_n,\hat{\lambda}_n)$ to $(\gamma^*,\lambda^*,\tau^*,\lambda^*)$, tail equivalence, Taylor's theorem, and Theorem 1

$$\sqrt{n}(\partial/\partial\lambda)s_n(\lambda^*) = \sqrt{n}(\partial/\partial\lambda)m_n'(\lambda^*)(\partial/\partial m)d[m_n(\lambda^*),\hat{\tau}_n]$$

$$= \sqrt{n}[M + o_s(1)]'\{(\partial/\partial m)d(0,\hat{\tau}_n) + [-D + o_s(1)]\ m_n(\lambda^*)\}$$

$$= [M + o_s(1)]'[-D + o_s(1)]\ \sqrt{n}\ m_n(\lambda^*) \quad .$$

The first result follows from Theorem 3.

By the same type of argument

$$\sqrt{n}(\partial/\partial\lambda)s_n(\lambda^*) = \sqrt{n}(\partial/\partial\lambda)s_n(\hat{\lambda}_n) + [\ \mathcal{J} + o_s(1)]\sqrt{n}(\hat{\lambda}_n - \lambda^*) .$$

By Theorem 2

$$= o_s(1) + [\ \mathcal{J} + o_s(1)]\ \sqrt{n}(\hat{\lambda}_n - \lambda^*)$$

and the second result follows from the first.

By Theorem 1 and the almost sure convergence of $(\gamma_n^\circ,\lambda_n^\circ,\hat{\tau}_n,\hat{\lambda}_n)$ to $(\gamma^*,\lambda^*,\tau^*,\lambda^*)$ it follows that $[S_n(\hat{\lambda}_n),\ M_n(\hat{\lambda}_n),\ D_n(\hat{\lambda}_n)] \to (S,\ M,\ D)$ whence $[\mathcal{I}_n(\hat{\lambda}_n),\ \mathcal{J}_n(\hat{\lambda}_n)] \to (\mathcal{I},\ \mathcal{J})$. □

To obtain results for estimation one holds γ_n° fixed at γ^* . Then for the example

$$\sqrt{n}\begin{pmatrix}\hat{\theta}_n - \gamma^* \\ \\ \hat{\sigma}_n - \sigma^*\end{pmatrix} \xrightarrow{\mathcal{L}} N\left(\begin{pmatrix}0 \\ \\ 0\end{pmatrix}, \begin{pmatrix}(\sigma^*)^2 \dfrac{\mathcal{E}\Psi^2(e/\sigma^*)}{[\mathcal{E}\Psi'(e/\sigma^*)]^2}(F'F)^{-1} & 0 \\ 0 & \dfrac{(\sigma^*)^4 \; \mathcal{E}[\Psi^2(e/\sigma^*) - \beta]^2}{4\;[\mathcal{E}\, e\Psi(e/\sigma^*)\Psi'(e/\sigma^*)]}\end{pmatrix}\right)$$

The variance formula

$$\mathcal{J}^{-1}\mathcal{I}\,\mathcal{J}^{-1} = (M'D\,M)^{-1}(M'D\,S\,D\,M)\;(M'D\,M)^{-1}$$

is the same as that which would result if the generalized least squares estimator

$$\hat{\beta} = (M'\,D\,M)^{-1}M'D\,y$$

were employed for the linear model

$$y = M\beta + e,\; e \sim (0,S)\;.$$

Thus, the greatest efficiency for given moment equations results when $D = S^{-1}$.

4. TESTS OF HYPOTHESES

Tests of the hypothesis

$$H: h(\lambda^\circ) = 0 \quad \text{against } A: h(\lambda^\circ) \neq 0$$

are considered here. A full rank assumption is imposed below which is not strictly necessary. However, the less than full rank case appears to be of no practical importance and a full rank assumption eliminates much clutter from the theorems and proofs.

Notation:

$\hat{\lambda}_n$ maximizes $s_n(\lambda)$

$\tilde{\lambda}_n$ maximizes $s_n(\lambda)$ subject to $h(\lambda) = 0$

$\hat{\mathfrak{I}} = \mathfrak{I}_n(\hat{\lambda}_n)$, $\tilde{\mathfrak{I}} = \mathfrak{I}_n(\tilde{\lambda}_n)$

$\hat{\mathcal{J}} = \mathcal{J}_n(\hat{\lambda}_n)$, $\tilde{\mathcal{J}} = \mathcal{J}_n(\tilde{\lambda}_n)$

$V = \mathcal{J}^{-1}\mathfrak{I}\,\mathcal{J}^{-1}$, $\hat{V} = \hat{\mathcal{J}}^{-1}\hat{\mathfrak{I}}\,\hat{\mathcal{J}}^{-1}$, $\tilde{V} = \tilde{\mathcal{J}}^{-1}\,\tilde{\mathfrak{I}}\,\tilde{\mathcal{J}}^{-1}$

$H(\lambda) = (\partial/\partial\lambda')\,h(\lambda)$ (the Jacobian of h of order r x p)

$h = h(\lambda^*)$, $\hat{h} = h(\hat{\lambda}_n)$, $\tilde{h} = h(\tilde{\lambda}_n)$,

$H = H(\lambda^*)$, $\hat{H} = H(\hat{\lambda}_n)$, $\tilde{H} = H(\tilde{\lambda}_n)$.

Assumption 7. The r-vector valued function $h(\lambda)$ defining the hypothesis $H: h(\lambda^\circ) = 0$ is continuously differentiable with Jacobian $H(\lambda) = (\partial/\partial\lambda')h(\lambda)$; $H(\lambda)$ has full rank at $\lambda = \lambda^*$. The matrix $V = \mathcal{J}^{-1}\mathfrak{I}\,\mathcal{J}^{-1}$ has full rank. The statement "the null hypothesis is true" means that $h(\lambda_n^\circ) = 0$ for all n .

Theorem 5. Under Assumptions 1 through 7 the statistics

$$W = n\,h'(\hat{\lambda}_n)(\hat{H}\hat{V}\hat{H}')^{-1}h(\hat{\lambda}_n)$$
$$R = n[(\partial/\partial\lambda)s_n(\tilde{\lambda}_n)]\,\tilde{\mathcal{J}}^{-1}\tilde{H}'(\tilde{H}\,\tilde{V}\,\tilde{H}')^{-1}\tilde{H}\,\tilde{\mathcal{J}}^{-1}[(\partial/\partial\lambda)s_n(\tilde{\lambda}_n)]$$

converge in distribution to the non-central chi square distribution with r degrees of freedom and noncentrality parameter $\alpha = \delta' H'(H V H')^{-1} H\delta/2$. Under the null hypothesis, the limiting distribution is the central chi square with r degrees of freedom.

Proof. (The statistic W) By Theorem 2 there is a sequence which is tail equivalent to $\hat{\lambda}_n$ and takes its values in Λ. The remarks refer to the tail equivalent sequence but a new notation is not introduced. Taylor's theorem applies to this sequence whence

$$\sqrt{n}[h_i(\hat{\lambda}_n) - h_i(\lambda^*)] = (\partial/\partial\lambda')h_i(\bar{\lambda}_{in})\sqrt{n}(\hat{\lambda}_n - \lambda^*) \quad i = 1, 2, \ldots, r$$

where $\|\bar{\lambda}_{in} - \lambda^*\| \le \|\hat{\lambda}_n - \lambda^*\|$. By Theorem 2 $\ell im_{n\to\infty}\|\bar{\lambda}_{in} - \lambda^*\| = 0$ almost surely whence $\ell im_{n\to\infty}(\partial/\partial\lambda)h_i(\bar{\lambda}_{in}) = (\partial/\partial\lambda)h_i(\lambda^*)$ almost surely. Now, in addition, $h(\lambda^*) = 0$ so the Taylor's expansion may be written $\sqrt{n}\, h(\hat{\lambda}_n) = [H + o_s(1)]\sqrt{n}(\hat{\lambda}_n - \lambda^*)$. Then by Theorem 4 $\sqrt{n}h(\hat{\lambda}_n)$ has the same asymptotic distribution as $H\sqrt{n}(\hat{\lambda}_n - \lambda^*)$. Now $(\hat{H}\, \hat{V}\, \hat{H}')^{-\frac{1}{2}}$ exists for n sufficiently large and converges almost surely to $(H\, V\, H')^{-\frac{1}{2}}$ whence $(\hat{H}\, \hat{V}\, \hat{H}')^{-\frac{1}{2}}\sqrt{n}\, h(\hat{\lambda}_n)$ and $(H\, V\, H')^{-\frac{1}{2}}H\sqrt{n}\,(\hat{\lambda}_n - \lambda^*)$ have the same asymptotic distribution. But

$$(H\, V\, H')^{-\frac{1}{2}}H\sqrt{n}\,(\hat{\lambda}_n - \lambda^*) \xrightarrow{\mathcal{L}} N[(H\, V\, H')^{-\frac{1}{2}}H\,\delta,\ I_r]$$

whence W converges in distribution to the non-central chi-square.

When the null hypothesis is true, it follows from Taylor's theorem that

$$0 = \sqrt{n}[h_i(\lambda_n^\circ) - h_i(\lambda^*)] = [(\partial/\partial\lambda')h_i(\overline{\lambda}_{in})]\ \sqrt{n}(\lambda_n^\circ - \lambda^*) .$$

Taking the limit as n tends to infinity this equation becomes

$$0 = (\partial/\partial\lambda')h_i(\lambda^*)\delta \text{ whence } H\delta = 0 \text{ and } \alpha = 0 .$$

(The statistic H) By Theorem 2 there is a sequence which is tail equivalent to $\tilde{\lambda}_n$ and takes its values in Λ. The remarks below refer to the tail equivalent sequence but a new notation is not introduced. By Taylor's theorem

$$(\partial/\partial\lambda_i)s_n(\tilde{\lambda}_n) = (\partial/\partial\lambda_i)s_n(\lambda^*) + [(\partial^2/\partial\lambda\partial\lambda_i)s_n(\tilde{\lambda}_{in})]'(\tilde{\lambda}_n - \lambda^*)$$

$$h_j(\tilde{\lambda}_n) = h_j(\lambda^*) + [(\partial/\partial\lambda')h_j(\bar{\lambda}_{jn})](\tilde{\lambda}_n - \lambda^*)$$

where $\|\bar{\lambda}_{in} - \lambda^*\|$, $\|\bar{\bar{\lambda}}_{jn} - \lambda^*\| \le \|\tilde{\lambda}_n - \lambda^*\|$ for $i = 1, 2, \ldots, p$ $j = 1, 2, \ldots, r$. By Theorem 2 there is for every realization of $\{e_t\}$ an N such that $h(\tilde{\lambda}_n) = 0$ for all $n > N$. Thus $h(\tilde{\lambda}_n) = o_s(1/\sqrt{n})$ and recall that $h(\lambda^*) = 0$. Then the continuity of $H(\lambda)$, the almost sure convergence of $\tilde{\lambda}_n$ to λ^* given by Theorem 2, and Theorem 1 permit these Taylor's expansions to be rewritten as

$$(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) = (\partial/\partial\lambda)s_n(\lambda^*) - [\mathcal{J} + o_s(1)](\tilde{\lambda}_n - \lambda^*)$$

$$[H + o_s(1)](\tilde{\lambda}_n - \lambda^*) = o_s(1/\sqrt{n}) .$$

These equations may be reduced algebraicly to

$$[H + o_s(1)][\mathcal{J} + o_s(1)]^{-1}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) = [H + o_s(1)][\mathcal{J} + o_s(1)]^{-1}(\partial/\partial\lambda)s_n(\lambda^*) + o_s(1/\sqrt{n}) .$$

Then it follows from Theorem 4 that

$$[H + o_s(1)][\mathcal{J} + o_s(1)]^{-1}\sqrt{n}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) \xrightarrow{\mathcal{L}} N(H\,\delta,\ H\,V\,H') .$$

The continuity of $H(\lambda)$, Theorem 2, and Theorem 1 permit the conclusion that

$$(\tilde{H}\,\tilde{V}\,\tilde{H}')^{-\frac{1}{2}}\,\tilde{H}\,\tilde{\mathcal{J}}^{-1}\sqrt{n}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) \xrightarrow{\mathcal{L}} N[(H\,V\,H')^{-\frac{1}{2}}H\,\delta,\ I_r]$$

whence R converges in distribution to the non-central chi-square.

This completes the argument but note for the next proof that

$$H \mathcal{J}^{-1} \sqrt{n}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) \xrightarrow{\mathcal{L}} N(H\,\delta,\ H\,V\,H') \ . \ \square$$

Theorem 6. Under Assumptions 1 through 6 the statistic

$$L = -2n[s_n(\tilde{\lambda}_n) - s_n(\hat{\lambda}_n)]$$

converges in distribution to the law of the quadratic form $Y = Z' \mathcal{J} Z$ where Z is distributed as the multivariate normal

$$Z \sim N[\mathcal{J}^{-1}H'(H\mathcal{J}^{-1}H')^{-1}H\delta\ ,\ \mathcal{J}^{-1}H'(H\mathcal{J}^{-1}H')^{-1}(HVH')(H\mathcal{J}^{-1}H')^{-1}H\mathcal{J}^{-1}] \ .$$

If $\mathcal{I} = \mathcal{J}$ then Y has the non-central chi-square distribution with r degrees of freedom and non-centrality parameter $\alpha = \delta'H'(HVH')^{-1}H\delta/2$. Under the null hypothesis Y is distributed as the central chi-square with r degrees of freedom provided that $\mathcal{I} = \mathcal{J}$. ($HVH' = H\mathcal{J}^{-1}H'$ will also suffice.)

Proof. By Theorem 2 there are sequences which are tail equivalent to $\hat{\lambda}_n$ and $\tilde{\lambda}_n$ and take their values in Λ^*. The remarks below refer to the tail equivalent sequences but a new notation is not introduced. By Taylor's theorem

$$-2n[s_n(\tilde{\lambda}_n) - s_n(\hat{\lambda}_n)]$$
$$= -2n[(\partial/\partial\lambda)s_n(\hat{\lambda}_n)]'(\tilde{\lambda}_n - \hat{\lambda}_n) - n(\tilde{\lambda}_n - \hat{\lambda}_n)'[(\partial^2/\partial\lambda\partial\lambda')s_n(\bar{\lambda}_n)](\tilde{\lambda}_n - \hat{\lambda}_n)$$

where $\|\bar{\lambda}_n - \hat{\lambda}_n\| \le \|\tilde{\lambda}_n - \hat{\lambda}_n\|$. Theorem 1 and the almost sure convergence of $(\tilde{\lambda}_n, \hat{\lambda}_n)$ to (λ^*, λ^*) imply that $(\partial^2/\partial\lambda\partial\lambda')s_n(\bar{\lambda}_n) = -[\mathcal{J} + o_s(1)]$. Now, by tail equivalence, $-2n(\partial/\partial\lambda')s_n(\hat{\lambda}_n) = o_s(1)$ whence

$$-2n[s_n(\tilde{\lambda}_n) - s_n(\hat{\lambda}_n)] = n(\tilde{\lambda}_n - \hat{\lambda}_n)'[\mathcal{J} + o_s(1)](\tilde{\lambda}_n - \hat{\lambda}_n) + o_s(1) \ .$$

By tail equivalence, and Theorem 2 there is for every realization of $\{e_t\}$ an N such that for $n \geq N$ there are Lagrange multipliers θ_n such that

$$\sqrt{n}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) - H'(\tilde{\lambda}_n)\sqrt{n}\ \theta_n = 0\ .$$

Thus,

$$[H + o_s(1)]'\sqrt{n}\ \theta_n = \sqrt{n}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) + o_s(1)$$

by a previous argument $\sqrt{n}(\partial/\partial\lambda)s_n(\hat{\lambda}_n) = o_s(1)$ whence

$$= \sqrt{n}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) - \sqrt{n}(\partial/\partial\lambda)s(\hat{\lambda}_n) + o_s(1)$$

by Taylor's theorem and previous arguments

$$= [\mathcal{J} + o_s(1)]\ \sqrt{n}(\hat{\lambda}_n - \tilde{\lambda}_n) + o_s(1)\quad .$$

From this string of equalities one has

$$H\,\mathcal{J}^{-1}[H + o_s(1)]'\sqrt{n}\ \theta_n = \sqrt{n}\ H\mathcal{J}^{-1}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) + o_s(1)$$

whence by the last line of the previous proof

$$H\,\mathcal{J}^{-1}[H + o_s(1)]'\sqrt{n}\ \theta_n \xrightarrow{\mathcal{L}} N(H\delta,\ HV\ H')\ .$$

Thus

$$\sqrt{n}\ \theta \xrightarrow{\mathcal{L}} N[(H\mathcal{J}^{-1}H')^{-1}H\delta,\ (H\mathcal{J}^{-1}H')^{-1}(HVH')(H\mathcal{J}^{-1}H')^{-1}]\quad .$$

Again from the string of equalities one has

$$\mathcal{J}^{-1}[H + o_s(1)]'\sqrt{n}\ \theta_n = \mathcal{J}^{-1}[\mathcal{J} + o_s(1)]\ \sqrt{n}(\hat{\lambda}_n - \tilde{\lambda}_n) + o_s(1)$$

whence

$$\sqrt{n}(\hat{\lambda}_n - \tilde{\lambda}_n) \xrightarrow{\mathcal{L}} N[\mathcal{J}^{-1}H'(H\mathcal{J}^{-1}H')^{-1}H\delta, \mathcal{J}^{-1}H'(H\mathcal{J}^{-1}H')^{-1}(HVH')(H\mathcal{J}^{-1}H')^{-1}H\,\mathcal{J}^{-1}]\ .$$

Then $\sqrt{n}(\hat{\lambda}_n - \tilde{\lambda}_n)$ converges in distribution to the distribution of the random variable Z and $\sqrt{n}(\hat{\lambda}_n - \tilde{\lambda}_n) = O_P(1)$. From the first paragraph of the proof,

$$-2n[s_n(\tilde{\lambda}_n) - s_n(\hat{\lambda}_n)] = n(\hat{\lambda}_n - \tilde{\lambda}_n)'[\mathcal{J} + o_s(1)](\hat{\lambda}_n - \tilde{\lambda}_n) + o_s(1)$$

$$= n(\hat{\lambda}_n - \tilde{\lambda}_n)'\mathcal{J}(\hat{\lambda}_n - \tilde{\lambda}_n) + O_P(1)\ o_s(1)\ O_P(1) + o_s(1)\ .$$

If $\mathcal{I} = \mathcal{J}$ then $V = \mathcal{J}^{-1}$ and

$$Z \sim N[\mathcal{J}^{-1}H'(H\mathcal{J}^{-1}H')^{-1}H\delta\ ,\ \mathcal{J}^{-1}H'(H\mathcal{J}^{-1}H')^{-1}H\mathcal{J}^{-1}]\ .$$

The conclusion that Y is chi-square follows at once from Theorem 2 of Searle (1971, p. 57). □

In a typical application, λ and τ are subvectors of γ or some easily computed function of γ. If γ_n^o is specified then λ_n^o and τ_n^o become specified. Thus, in a typical application, $(\gamma_n^o,\ \tau_n^o,\ \lambda_n^o)$ is specified and the noncentrality parameter $\alpha = \delta'H'(HVH')^{-1}H\delta/2$ is to be computed. The annoyance of having to specify $(\gamma^*,\ \tau^*,\ \lambda^*)$ in order to make this computation may be eliminated by application of Theorem 7.

NOTATION.

$$S^o = (1/n)\sum_{t=1}^n \int_{\mathcal{E}} m[Y(e,x_t,\gamma_n^o),x_t,\tau_n^o,\lambda_n^o]\quad m'[Y(e,x_t,\gamma_n^o),x_t,\tau_n^o,\lambda_n^o]\ dP(e)$$

$$M^o = (1/n)\sum_{t=1}^n \int_{\mathcal{E}} (\partial/\partial\lambda')m[Y(e,x_t,\gamma_n^o),x_t,\tau_n^o,\lambda_n^o]\ dP(e)$$

$$D^o = (\partial^2/\partial m\partial m')\ d(o,\tau_n^o)$$

$$\mathcal{I}^o = (M^o)'D^oS^oD^oM^o$$

$$\mathcal{J}^o = -(M^o)'D^oM^o$$

$$V^o = (\mathcal{J}^o)^{-1}\mathcal{I}^o(\mathcal{J}^o)^{-1}$$

$$\alpha_n^o = n\ h^{o\prime}[H^oV^oH^{o\prime}]^{-1}h^o/2\quad .$$

<u>Theorem 7</u>. Let Assumptions 1 through 7 hold and let $\{(\gamma_n^o,\tau_n^o,\lambda_n^o)\}$ be any sequence with $\ell im_{n\to\infty}(\gamma_n^o,\tau_n^o,\lambda_n^o) = (\gamma^*,\ \tau^*,\ \lambda^*)$ and $\ell im_{n\to\infty}\sqrt{n}(\lambda_n^o - \lambda^*) = \delta$. Then $\ell im_{n\to\infty}\alpha_n^o = \alpha$.

<u>Proof</u>. By the continuity of $H(\lambda)$, Theorem 1, and the assumption that $\ell im_{n\to\infty}(\gamma_n^o,\ \tau_n^o,\ \lambda_n^o) = (\gamma^*,\ \tau^*,\ \lambda^*)$ it follows that $\ell im_{n\to\infty}[H(\lambda_n^o)V^oH'(\lambda_n^o)]^{-1} = (HVH')^{-1}$.

By Taylor's theorem

$$\sqrt{n}\ h_i(\lambda_n^o) = \sqrt{n}[h_i(\lambda_n^o) - h_i(\lambda^*)] = [(\partial/\partial\lambda)h_i(\bar{\lambda}_{in})]'\sqrt{n}(\lambda_n^o - \lambda^*)$$

where $\|\bar{\lambda}_{in} - \lambda^*\| \le \|\lambda_n^o - \lambda^*\|$ for $i = 1, 2, \ldots, r$. Thus, $\ell im_{n\to\infty}\sqrt{n}\ h(\lambda_n^o) = H\delta$. □

5. EXAMPLES

Scale Invariant M-Estimators

Recent literature:

Ruskin (1978)

Model:

$$y_t = f(x_t, \theta^*) + e_t$$

e is distributed symmetrically about zero

Moment equations:

$$m_n(\lambda) = (1/n)\, \Sigma_{t=1}^{n} \begin{pmatrix} \Psi\{[y_t - f(x_t,\theta)]/\sigma\}\, (\partial/\partial\theta)\, f(x_t,\theta) \\ \Psi^2\{[y_t - f(x_t,\theta)]/\sigma\} - \beta \end{pmatrix}$$

$\lambda = (\theta', \sigma)'$

$\Psi(u)$ is an odd function, $0 < \beta < 1$.

Distance function:

$$d(m) = -\tfrac{1}{2} m'm$$

Asymptotic distribution parameters:

$$S = \begin{pmatrix} \int_{\mathcal{E}} \Psi^2(e/\sigma^*)\, dP(e)\, F'F & 0 \\ 0' & \int_{\mathcal{E}} [\Psi^2(e/\sigma^*) - \beta]^2\, dP(e) \end{pmatrix}$$

$$M = \begin{pmatrix} -(1/\sigma^*)\int_{\mathcal{E}} \Psi'(e/\sigma^*)\, dP(e)\, F'F & 0 \\ 0' & -2(1/\sigma^*)^2 \int_{\mathcal{E}} \Psi(e/\sigma^*)\Psi(e/\sigma^*) e\, dP(e) \end{pmatrix}$$

$D = -I$

$F'F = \int_{\mathcal{X}} (\partial/\partial\theta)\, f(x,\theta^*)\, (\partial/\partial\theta')\, f(x,\theta^*)\, d\mu(x)$

σ^* solves $\int_{\mathcal{E}} \Psi^2(e/\sigma)\, dP(e) = \beta$

$$V = \begin{pmatrix} (\sigma^*)^2 \dfrac{\mathcal{E}\Psi^2(e/\sigma^*)}{[\mathcal{E}\Psi'(e/\sigma^*)]^2}(F'F)^{-1} & 0 \\ 0' & \dfrac{(\sigma^*)^4\, \mathcal{E}[\Psi^2(e/\sigma^*) - \beta]^2}{4[\mathcal{E}\; e\; \Psi(e/\sigma^*)\Psi'(e/\sigma^*)]^2} \end{pmatrix}$$

Single Equation Nonlinear Least Squares

Recent literature:

Jennrich (1969), Malinvaud (1970), Gallant (1973, 1975a, 1975b)

Model:

$$y_t = f(x_t, \theta_n^\circ) + e_t$$

$$\mathcal{E}(e_t) = 0\ , \ \mathcal{E}(e_t^{\,2}) = (\sigma^*)^2$$

Moment equations:

$$m_n(\lambda) = (1/n)\Sigma_{t=1}^{n}(\partial/\partial\theta)f(x_t,\theta)[y_t - f(x_t,\theta)]/\hat{\tau}_n$$

$\sqrt{n}\,(\hat{\tau}_n - (\sigma^*)^2)$ is bounded in probability, $\hat{\tau}_n = (1/n)\ \Sigma_{t=1}^{n}\ \hat{e}_t^2$ where $\hat{e}_t$ are least squares residuals will suffice

$$\lambda = \theta$$

Distance function:

$$d(m) = -\tfrac{1}{2}m'm$$

Asymptotic distribution parameters:

$$S = -M = (\sigma^*)^2 \int_{\mathcal{X}} (\partial/\partial\theta)f(x,\theta^*)(\partial/\partial\theta')\ f(x,\theta^*)\ d\mu(x)$$

$$D = -I$$

$$V = S^{-1}$$

Multivariate Nonlinear Least Squares

Recent literature:

Malinvaud (1970b), Gallant (1975c), Holly (1978)

Model:

$$y_t = f(x_t, \theta_n^\circ) + e_t$$

$$\mathcal{E}(e_t) = 0 \ , \ \mathcal{E}(e_t e_t') = \Sigma^*$$

Moment equations:

$$m_n(\lambda) = (1/n)\Sigma_{t=1}^{n}[(\partial/\partial\theta')\ f(x_t,\theta)]'\ \hat{\Sigma}^{-1}[y_t - f(x_t,\theta)]$$

$\hat{\Sigma} = (1/n)\Sigma_{t=1}^{n}\ \hat{e}_t\hat{e}_t'$; $\hat{e}_t$ are single equation residuals

$$\lambda = \theta$$

Distance function:

$$d(m) = -\tfrac{1}{2}m'm$$

Asymptotic distribution parameters:

$$S = M = \int_{\mathcal{X}} [(\partial/\partial\theta')f(x,\theta^*)]'(\Sigma^*)^{-1}[(\partial/\partial\theta')\ f(x,\theta^*)]\ d\mu(x)$$

$$D = -I$$

$$V = S^{-1}$$

Single Equation Maximum Likelihood

Model:

$$y_t = f(x_t, \theta_n^o) + e_t$$

$$\mathcal{E}(e_t) = 0, \quad \mathcal{E}(e_t^{\,2}) = (\sigma^*)^2$$

Moment equations:

$$m_n(\lambda) = (1/n)\Sigma_{t=1}^{n} \begin{pmatrix} [y_t - f(x_t,\theta)](\partial/\partial\theta)\, f(x_t,\theta) \\ [y_t - f(x_t,\theta)]^2 - \sigma^2 \end{pmatrix}$$

$$\lambda = (\theta', \sigma^2)'$$

Distance function:

$$d(m) = -\tfrac{1}{2}m'm$$

Asymptotic distribution parameters:

$$S = \begin{pmatrix} \sigma^2 F'F & \mathcal{E}(e^3)f \\ \mathcal{E}(e^3)f' & \mathrm{Var}(e^2) \end{pmatrix}$$

$$M = \begin{pmatrix} -F'F & 0 \\ 0' & -1 \end{pmatrix}$$

$$D = -I$$

$$F'F = \int_{\mathcal{X}} (\partial/\partial\theta)\, f(x,\theta^*)(\partial/\partial\theta')\, f(x,\theta^*)\, d\mu(x)$$

$$f = \int_{\mathcal{X}} (\partial/\partial\theta)\, f(x,\theta^*)\, d\mu(x)$$

$$V = \begin{pmatrix} \sigma^2(F'F)^{-1} & \mathcal{E}(e^3)(F'F)^{-1}f \\ \mathcal{E}(e^3)f'(F'F)^{-1} & \mathrm{Var}(e^2) \end{pmatrix}$$

Comment:

Under symmetry $\mathcal{E}(e^3) = 0$.

Multivariate Maximum Likelihood

Recent literature:

Malinvaud (1970b), Barnett (1976), Holly (1978)

Model:

$$y_t = f(x_t, \theta_n^o) + e_t$$

$$\mathcal{E}(e_t) = 0\,, \quad \mathcal{E}(e_t, e_t') = \Sigma^*$$

Moment equations:

$$m_n(\lambda) = (1/n)\Sigma_{t=1}^{n} \begin{pmatrix} [(\partial/\partial\theta')\, f(x_t,\theta)]'\Sigma^{-1}[y_t - f(x_t,\theta)] \\ \mathrm{vec}\{[y_t - f(x_t,\theta)][y_t - f(x_t,\theta)]' - \Sigma\} \end{pmatrix}$$

$$\lambda = (\theta', \sigma')'$$

$\sigma' = (\sigma_{11}, \sigma_{12}, \sigma_{23}, \ldots, \sigma_{1M}, \sigma_{2M}, \ldots, \sigma_{MM})$, upper triangle of Σ

$\mathrm{vec}(\Sigma) = A\sigma$, A an $M^2 \times M(M+1)/2$ matrix of zeroes and ones

Distance function:

$$d(m) = -\tfrac{1}{2}m'm$$

Asymptotic distribution parameters:

$$S = \begin{pmatrix} F'\Sigma^{-1}F & f'\Sigma^{-1}\mathcal{E}[e\ \mathrm{vec}'(ee')] \\ \mathcal{E}[\mathrm{vec}(ee')e']\Sigma^{-1}f & \mathrm{Var}[\mathrm{vec}(ee')] \end{pmatrix}$$

$$M = \begin{pmatrix} -F'\Sigma^{-1}F & 0 \\ 0' & -A \end{pmatrix}$$

$$D = -I$$

$$F'\Sigma^{-1}F = \int_{\mathcal{X}} [(\partial/\partial\theta')\, f(x,\theta^*)]'(\Sigma^*)^{-1}[(\partial/\partial\theta')\, f(x,\theta^*)]\, d\mu(x)$$

$$\Sigma^{-1}f = (\Sigma^{*})^{-1}\int_{\mathcal{X}} (\partial/\partial\theta') \, f(x,\theta^{*}) \, d\mu(x)$$

$$V = \begin{pmatrix} (F'\Sigma^{-1}F)^{-1} & (F'\Sigma^{-1}F)^{-1}f'\Sigma^{-1}\mathcal{E}[e \, \mathrm{vec}'(ee')]A(A'A)^{-1} \\ (A'A)^{-1}A'\mathcal{E}[\mathrm{vec}(ee')e']\Sigma^{-1}f(F'\Sigma^{-1}F)^{-1} & (A'A)^{-1}A'\mathrm{Var}[\mathrm{vec}(ee')]A(A'A)^{-1} \end{pmatrix}$$

Comment:

Under normality $\mathcal{E}[\mathrm{vec}(ee')e'] = 0$, $A'\mathrm{Var}[\mathrm{vec}(ee')]A = 2A'(\Sigma^{*}\otimes \Sigma^{*})A$.

Two-Stage Nonlinear Least-Squares

Recent literature:

Amemiya (1974), Gallant and Jorgenson (1979)

System:

$$q(y_t, x_t, \theta_n^*) = e_t$$

$$\mathcal{E}(e_t) = 0\,, \quad \mathcal{C}(e_t e_t') = \Sigma^*$$

Equation of interest:

$$q_\alpha(y_t, x_t, \theta_\alpha^\circ) = e_{\alpha t}$$

$$\mathcal{E}(e_t) = 0\,, \quad \mathrm{Var}(e_{\alpha t}) = \sigma_{\alpha\alpha}^*$$

Moment equations:

$$m_n(\lambda) = (1/n)\Sigma_{t=1}^n z_t q_\alpha(y_t, x_t, \theta_\alpha)$$

$$\lambda = \theta_\alpha$$

$$z_t = z(x)\,, \quad z(x) \text{ continuous}$$

Distance function:

$$d(m, \tau) = -\tfrac{1}{2} m' \tau^{-1} m$$

$$\hat{\tau}_n = \hat{\sigma}_{\alpha\alpha}\,(1/n)\,\Sigma_{t=1}^n z_t z'_t$$

$\sqrt{n}(\hat{\sigma}_{\alpha\alpha} - \sigma_{\alpha\alpha}^*)$ is bounded in probability, $\hat{\sigma}_{\alpha\alpha} = (1/n)\Sigma_{t=1}^n \hat{e}_{\alpha t}^2$ where $\hat{e}_{\alpha t}$ are two-stage least-squares residuals will suffice

Asymptotic distribution parameters:

$$S = (\sigma_{\alpha\alpha}^*)^2 Z'Z$$

$$M = Z'Q_\alpha$$

$$D = -S^{-1}$$

$$Z'Z = \int_{\mathcal{X}} z(x)\, z'(x)\, d\mu(x)$$

$$Z'Q_{\alpha} = \int_{\mathcal{X}}\int_{\mathcal{E}} z(x)(\partial/\partial\theta_{\alpha}')\, q_{\alpha}[Y(e,x,\theta_{\alpha}^{*}),x,\theta_{\alpha}]\, dP(e)\, d\mu(x)\Big|_{\theta_{\alpha}=\theta_{\alpha}^{*}}$$

$$V = \sigma_{\alpha\alpha}^{*} Q_{\alpha}' Z(Z'Z)^{-1} Z' Q_{\alpha}$$

Comment:

$$\mathcal{I} = \mathcal{J}$$

Three-Stage Nonlinear Least-Squares

Recent literature:

Jorgenson and Laffont (1979), Gallant (1977), Amemiya (1977), Gallant and Jorgenson (1979)

Model:

$$q(y_t, x_t, \theta_n^\circ) = e_t$$

$$\mathcal{E}(e_t) = 0 \;, \quad C(e_t e_t') = \Sigma^*$$

Moment equations:

$$m_n(\lambda) = (1/n) \Sigma_{t=1}^n q(y_t, x_t, \theta) \otimes z_t$$

$$\lambda = \theta$$

$$z_t = z(x_t) \quad , \; z(x) \text{ continuous}$$

Distance function:

$$d(m,\tau) = -\tfrac{1}{2} m' \tau^{-1} m$$

$$\tau_n = [\hat{\Sigma} \otimes (1/n)\Sigma_{t=1}^n z_t z_t']$$

$\hat{\Sigma} = (1/n)\Sigma_{t=1}^n \hat{e}_t \hat{e}_t'$; $\hat{e}_t$ are two-stage least-squares residuals

Asymptotic distribution parameters:

$$S = \Sigma^* \otimes (Z'Z) = \Sigma^* \otimes \int_{\mathcal{X}} z(x)\, z'(x)\, d\mu(x)$$

$$M = Q \otimes Z = \int_{\mathcal{X}} (\partial/\partial\theta')\, q[Y(e,x,\theta^*), x, \theta] \otimes z(x)\, d\mu(x) \,\Big|_{\theta=\theta^*}$$

$$D = -S^{-1}$$

$$V = [(Q \otimes Z)'[(\Sigma^*)^{-1} \otimes (Z'Z)^{-1}](Q \otimes Z)]^{-1}$$

6. M-ESTIMATORS

An M-estimator $\hat{\lambda}_n$ is defined as the solution of the optimization problem

$$\text{Maximize:} \quad s_n(\lambda) = (1/n)\Sigma_{t=1}^{n} s(y_t, x_t, \hat{\tau}_n, \lambda)$$

where $\hat{\tau}_n$ is a random variable which corresponds conceptually to estimators of nuisance parameters. A constrained M-estimator $\tilde{\lambda}_n$ is the solution of the optimization problem

$$\text{Maximize:} \quad s_n(\lambda) \quad \text{subject to } h(\lambda) = 0$$

where $h(\lambda)$ maps R^p into R^r. The objective of this section is to establish conditions such that these estimators are asymptotically normally distributed. These results are due to Souza and Gallant (1979) where proofs may be found.

Two examples are carried throughout the discussion to provide the reader with the flavor of the details in an application. In both cases the data generating model is

$$y_t = f(x_t, \gamma_n^{\circ}) + e_t \quad .$$

The first estimator is a robust estimator with distance function

$$s_1(y,x,\lambda) = -\rho[y - f(x,\lambda)]$$

where $\rho(u) = \ell n \cosh(u/2)$. The second is an iteratively rescaled robust estimator. The distance function is

$$s_2(y,x,\tau,\lambda) = \rho\{[y - f(x,\lambda)]/\tau\} \ .$$

The scale estimator $\hat{\tau}$ is obtained by computing $\hat{\theta}$ to maximize $(1/n)\Sigma_{t=1}^{n} s_1(y_t, x_t, \theta)$ and then solving

$$(1/n)\Sigma_{t=1}^{n} \Psi^2\{[y_t - f(x_t,\theta)]/\tau\} - \int \Psi^2(e)\, d\Phi(e)$$

for τ where $\Psi(u) = \rho'(u) = \frac{1}{2}\tanh(u/2)$.

NOTATION.

$$s_n(\lambda) = (1/n)\Sigma_{t=1}^{n} s(y_t, x_t, \hat{\tau}_n, \lambda)$$

$$\bar{s}(\gamma,\tau,\lambda) = \int_{\mathcal{X}}\int_{\mathcal{E}} s[Y(e,x,\gamma),x,\tau,\lambda]\, dP(e)\, d\mu(x) \; .$$

The identification is

Assumption 4′. The parameter γ° is indexed by n and $\ell im_{n\to\infty}\gamma_n^{\circ} = \gamma^*$ for some point $\gamma^* \in \Gamma$. The sequence $\hat{\tau}_n$ converges almost surely to a point τ^* and $\sqrt{n}(\hat{\tau}_n - \tau^*)$ is bounded in probability. There is a compact subset Λ' of Λ and there are unique points λ^*, λ_1°, λ_2°, ... corresponding to $\gamma = \gamma^*$, γ_1°, γ_2°, ... which maximize $\bar{s}(\gamma^*, \tau^*, \lambda)$, $\bar{s}(\gamma_1^{\circ},\tau^*,\lambda)$, $\bar{s}(\gamma_2^{\circ},\tau^*,\lambda)$, ... over Λ. The function $h(\lambda)$ of the hypothesis H: $h(\lambda_n^{\circ}) = 0$ is a continuous vector valued function on Λ; the point λ^* satisfies $h(\lambda^*) = 0$ and $\ell im_{n\to\infty}\sqrt{n}(\lambda_n^{\circ} - \lambda^*) = \delta$.

A verification that $\hat{\tau}_n$ has the requisite properties is straightforward in typical applications. A verification for the example is deferred until end of the section.

A verification of unique maxima of $\bar{s}(\gamma,\tau^*,\lambda)$ usually commences by proposing an obviously minimal identification condition. Then known results for the location problem are exploited to verify a unique association of λ to γ. To illustrate with the example, it is clearly impossible to identify λ by observing $\{y_t, x_t\}$ if $f(x,\lambda) = f(x,\gamma)$ a. e. μ for some $\lambda \neq \gamma$. Then a minimal identification condition is

$$\lambda \neq \gamma \Rightarrow \mu\{x\colon f(x,\lambda) \neq f(x,\gamma)\} > 0 \; .$$

Now $\varphi(\delta) = \int\rho(e + \delta)\, dP(e)$ is known to have a unique minimum at $\delta = 0$ when P(e) is symmetric about zero and assigns positive probability to every nonempty, open interval. If $\lambda \neq \gamma$ and $\delta(x) = f(x,\gamma) - f(x,\lambda)$ then $\varphi[\delta(x)] \geq \varphi(0)$ for every x and by the identification condition $\varphi[\delta(x)] > \varphi(0)$ on some set of positive μ measure whenc

$$\bar{s}_1(\gamma,\lambda) = -\int_{\mathcal{X}} \varphi[\delta(x)]\, d\mu(x) < -\int_{\mathcal{X}} \varphi(0)\, d\mu(x) = \bar{s}_1(\gamma,\gamma)$$

as required.

The almost sure convergence imposed in Assumption 4′ implies that there is a sequence which takes its values in a neighborhood of τ^* and is tail equivalent to $\hat{\tau}_n$. Thus, without loss of generality, it may be assumed that $\hat{\tau}_n$ takes its values in a compact sphere T for which τ^* is an interior point. Similarly, Γ may be taken as a compact sphere with interior point γ^* . Thus, the effective conditions of the next assumption are domination of the objective function and a requirement that eventually it suffices to minimize $s_n(\lambda)$ over Λ' .

ASSUMPTION 5′. The sets Γ and T are compact spheres containing γ^* and τ^*, respectively. To almost every realization of $\{e_t\}$ there corresponds an N for which $n > N$ implies $\sup_{\Lambda'} s_n(\lambda) = \sup_{\Lambda} s_n(\lambda)$. The function $s(y,x,\tau,\lambda)$ is continuous on $\mathcal{Y} \times \mathcal{X} \times T \times \Lambda'$ and $|s(y,x,\tau,\lambda)| \le b[q(y,x,\gamma),x]$ on $\mathcal{Y} \times \mathcal{X} \times T \times \Lambda' \times \Gamma$. (The function $b(e,x)$ is given by Assumption 3.)

The exhibition of the requisite dominating function is an ad hoc process and one exploits the special characteristics of an application. For the example

$$\begin{aligned} |s_1(y,x,\lambda)| &= \rho[e + f(x,\gamma) - f(x,\lambda)] \\ &\le \tfrac{1}{2}|e + f(x,\gamma) - f(x,\lambda)| \\ &\le |e| + \sup_{\Gamma}|f(x,\gamma)| + \sup_{\Lambda'}|f(x,\lambda)| . \end{aligned}$$

The domination condition obtains if $P(e)$ has a finite mean and if $\mu(x)$ can accomodate the tail behavior of $\sup_{\Gamma}|f(x,\gamma)|$ and $\sup_{\Lambda'}|f(x,\lambda)|$. Or, take $\mathcal{X}$ compact so that these functions are bounded in view of Assumption 2. Some carefully worked examples for more complex situations may be found in Gallant and Holly (1980) and Gallant (1977).

The construction of Λ' is also ad hoc. Most authors have adopted the simple expedient of taking Λ compact and putting $\Lambda' = \Lambda$. There has been some reluctance to impose bounds on scale parameters and location parameters which enter the model linearly. General treatments for unbounded location parameters with least squares methods may be found in Malinvaud (1970) and Gallant (1973). For redescending ρ functions, the use of an initial consistent estimator as a start value for an algorithm which is guaranteed to converge to a local minimum of $s_n(\lambda)$ suffices to confine $\hat{\lambda}_n$ to Λ' eventually. A construction of Λ' for scale parameters with maximum likelihood estimation of the parameters of the general multivariate model is in Gallant and Holly (1980).

However, one does not wander haphazardly into nonlinear estimation. Typically one has need of a considerable knowledge of the situation in order to construct the model. Presuming knowledge of Λ' as well is probably not an unreasonably assumption. Most authors apparently take this position.

THEOREM 2'. (Strong consistency) Let Assumptions 1 through 5'hold. Then $\hat{\lambda}_n$ and $\tilde{\lambda}_n$ converge almost surely to λ^*.

NOTATION.

$$\mathcal{I} = \int_{\mathcal{X}}\int_{\mathcal{E}} \{(\partial/\partial\lambda)s[Y(e,x,\gamma^*),x,\tau^*,\lambda^*]\}\{(\partial/\partial\lambda)s[Y(e,x,\gamma^*),x,\tau^*,\lambda^*]\}' dP(e)\ d\mu(x$$

$$\mathcal{J} = -\int_{\mathcal{X}}\int_{\mathcal{E}} (\partial^2/\partial\lambda\partial\lambda')s[Y(e,x,\gamma^*),x,\tau^*,\lambda^*]\ dP(e)\ d\mu(x)$$

$$\mathcal{I}_n(\lambda) = (1/n)\Sigma_{t=1}^n[(\partial/\partial\lambda)s(y_t,x_t,\hat{\tau}_n,\Lambda)][(\partial/\partial\lambda)s(y_t,x_t,\hat{\tau}_n,\lambda)]'$$

$$\mathcal{J}_n(\lambda) = -(1/n)\Sigma_{t=1}^n(\partial^2/\partial\lambda\partial\lambda')s(y_t,x_t,\hat{\tau}_n,\lambda)\ .$$

ASSUMPTION 6′. There are open spheres Γ^*, T^*, and Λ^* with $\gamma^* \in \Gamma^* \subset \Gamma$, $\tau^* \in T^* \subset T$, and $\lambda^* \in \Lambda^* \subset \Lambda'$. The elements of $(\partial/\partial\lambda)s(y,x,\tau,\lambda)$, $(\partial^2/\partial\lambda\partial\lambda')s(y,x,\tau,\lambda)$, $(\partial^2/\partial\tau\partial\lambda')s(y,x,\tau,\lambda)$, and $[(\partial/\partial\lambda)s(y,x,\tau,\lambda)][(\partial/\partial\lambda)s(y,x,\tau,\lambda)]'$ are continuous and dominated by $b[q(y,x,\gamma),x]$ on $\mathcal{Y} \times \mathcal{X} \times \bar{\Gamma}^* \times \bar{T}^* \times \bar{\Lambda}^*$ where the overbar indicates the closure of a set. Moreover, $\mathcal{J}$ is nonsingular and

$$\int_{\mathcal{E}} (\partial/\partial\lambda)s[Y(e,x,\gamma_n^\circ),x,\tau^*,\lambda_n^\circ]\, dP(e) = 0 ,$$

$$\int_{\mathcal{X}}\int_{\mathcal{E}} (\partial^2/\partial\tau\partial\lambda')s[Y(e,x,\gamma^*),x,\tau^*,\lambda^*]\, dP(e)\, d\mu(x) = 0 .$$

The first integral condition is that the expectation of the "score" is zero. It is central to our results and is apparently an intrinsic property of reasonable estimation procedures. The second integral condition is sometimes encountered in the theory of maximum likelihood estimation; see Durbin (1970) for a detailed discussion. It validates the application of maximum likelihood theory to a subset of the parameters when the remainder are treated as if known in the derivations but are subsequently estimated. The assumption plays the same role here. It can be avoided in maximum likelihood estimation at a cost of additional complexity in the results; see Gallant and Holly (1980) for details. It can probably be avoided here but there is no reason to further complicate the results in view of the intended applications.

Consider the verification of the integral conditions for the example with

$$s_2(y,x,\tau,\lambda) = -\rho\{[y - f(x,\lambda)]/\tau\} .$$

Now

$$(\partial/\partial\lambda)s_2(y,x,\tau,\lambda^*) = (1/\tau)\Psi(e/\tau)(\partial/\partial\lambda)\ f(x,\lambda)\ ,$$

$$(\partial^2/\partial\tau\partial\lambda')s_2(y,x,\tau,\lambda^*) = (-1/\tau^2)[\Psi(e/\tau) + \Psi'(e/\tau)(e/\tau)](\partial/\partial\lambda)\ f(x,\lambda)\ .$$

Both $\Psi(e/\tau)$ and $\Psi'(e/\tau)(e/\tau)$ are odd functions and will integrate to zero for symmetric $P(e)$. Thus, both integral conditions are satisfied.

THEOREM 3′. (Asymptotic Normality of the Scores) Under Assumptions 1 through 6′

$$(1/\sqrt{n})\Sigma_{t=1}^{n}(\partial/\partial\lambda)s(y_t,x_t,\hat{\tau}_n,\lambda_n^{\circ}) \xrightarrow{\mathcal{L}} N(0,\mathcal{J})\ ,$$

$\mathcal{J}$ may be singular.

THEOREM 4′. Let Assumptions 1 through 6′ hold. Then

$$(1/\sqrt{n})\Sigma_{t=1}^{n}(\partial/\partial\lambda)s(y_t,x_t,\hat{\tau}_n,\lambda^*) \xrightarrow{\mathcal{L}} N(\mathcal{Y}\delta,\mathcal{J})\ ,$$

$$\sqrt{n}(\hat{\lambda}_n - \lambda^*) \xrightarrow{\mathcal{L}} N(\delta,\mathcal{Y}^{-1}\mathcal{J}\mathcal{Y}^{-1})\ ,$$

$\mathcal{J}_n(\hat{\lambda}_n)$ converges almost surely to $\mathcal{J}$, and $\mathcal{Y}_n(\hat{\lambda}_n)$ converges almost surely to $\mathcal{Y}$.

For the example with $s_1(y,x,\lambda)$ defining the estimator and with $P(e)$ symmetric

$$\mathcal{J} = \int_{\mathcal{E}}\Psi^2(e)\ dP(e)\ (F'F)\ ,\ \mathcal{Y} = -\int_{\mathcal{E}}\Psi^2(e)\ dP(e)\ (F'F)$$

where $(F'F) = \int_{\mathcal{X}}[(\partial/\partial\lambda)f(x,\lambda^*)][(\partial/\partial\lambda)\ f(x,\lambda^*)]'\ d\mu(x)$

whence $\sqrt{n}(\hat{\lambda}_n - \lambda^*) \xrightarrow{\mathcal{L}} N[\delta,\ \frac{\mathcal{E}\Psi^2}{(\mathcal{E}\Psi')^2}(F'F)^{-1}]$;

with $s_2(y,x,\tau,\lambda)$ defining the estimator

$$\sqrt{n}(\hat{\lambda}_n - \lambda^*) \xrightarrow{\mathcal{L}} N[\delta,(\tau^*)^2\frac{\mathcal{E}\Psi^2(e/\tau^*)}{[\mathcal{E}\Psi'(e/\tau^*)]^2}(F'F)^{-1}]\ .$$

The choice of $\rho(u) = \ell n \cosh (u/2)$ to obtain these results was for specificity in discussing the tail behavior of P and μ and to suggest constructions of the requisite dominating function $b(e,x)$. Other than this, these results apply to any $\rho(u)$ which is a twice continuously differentiable even function provided that $P(e)$ is symmetric and that $\int_{\mathcal{E}} \rho(e+\delta)\, dP(e)$ has a unique minimum at $\delta = 0$.

A verification that $\hat{\tau}_n$ of the example satisfies Assumption 4' was promised. Assume that f, μ, and P are such that Assumptions 1 through 6' are satisfied for the preliminary estimator $\hat{\theta}_n$ which maximizes $(1/n)\Sigma_{t=1}^{n} s_1(y_t, x_t, \theta)$ and take P symmetric so that $(\hat{\theta}_n, \gamma_n^o)$ converges almost surely to (γ^*, γ^*) and $\sqrt{n}[(\hat{\theta}_n \gamma_n^o) - (\gamma^* \gamma^*)]$ is bounded in probability. Define

$$\delta(x,\gamma,\theta) = f(x,\gamma) - f(x,\theta)$$

$$\varphi_n(\tau,\gamma,\theta) = \frac{1}{n}\Sigma_{t=1}^{n} \Psi^2[e_t/\tau + \delta(x_t,\gamma,\theta)/\tau] - \beta$$

$$\bar{\varphi}(\tau,\gamma,\theta) = \iint \Psi^2[e/\tau + \delta(x,\gamma,\theta)/\tau] - \beta \; dP(e)\, d\mu(x)$$

and recall that $\hat{\tau}_n$ solves $\varphi_n(\tau, \gamma_n^o, \hat{\theta}_n) = 0$ for $\beta = \int \Psi^2(e)\, d\Phi(e)$. Both φ_n and $\bar{\varphi}$ are strictly decreasing functions of τ and there is a τ^* with $\bar{\varphi}(\tau^*, \gamma^*, \gamma^*) = 0$.

By Theorem 1 and the almost sure convergence of $(\hat{\theta}_n, \gamma_n^o)$, given $\epsilon > 0$ there is an N such that $n > N$ implies $\varphi_n(\tau^* + \epsilon, \gamma_n^o, \hat{\theta}_n) < \varphi_n(\tau^* - \epsilon, \gamma_n^o, \hat{\theta}_n)$ for almost every realization. Thus $\tau^* - \epsilon < \hat{\tau}_n < \tau^* + \epsilon$ whence $\hat{\tau}_n$ converges almost surely to τ^*. Applications of Taylor's theorem and Theorem 1 yield

$$\sqrt{n}(\hat{\tau}_n - \tau^*) = [(\partial/\partial\tau)\varphi_n(\bar{\tau}, \gamma_n^o, \hat{\theta}_n)]^{-1} \sqrt{n}\,\varphi_n(\tau^*, \gamma_n^o, \hat{\theta}_n)$$

$$= [(\partial/\partial\tau)\bar{\varphi}(\tau^*, \gamma^*, \gamma^*) + o_s(1)]^{-1} \sqrt{n}\,\varphi_n(\tau^*, \gamma_n^o, \hat{\theta}_n)$$

and

$$\sqrt{n}\;\varphi_n(\tau^*, \gamma_n^o, \hat{\theta}_n)$$

$$= \frac{1}{\sqrt{n}}\Sigma_{t=1}^n \psi^2(e_t/\tau^*) + [\;(\partial/\partial\theta)\;\varphi_n(\tau^*, \gamma_n^o, \bar{\theta})]\,\sqrt{n}(\hat{\theta}_n - \gamma_n^o)$$

$$= \frac{1}{\sqrt{n}}\Sigma_{t=1}^n \psi^2(e_t/\tau^*) + [(\partial/\partial\theta)\;\bar{\varphi}(\tau^*, \gamma^*, \gamma^*) + o_s(1)]\,\sqrt{n}(\hat{\theta}_n - \gamma_n^o)\;.$$

By the central limit $(1/\sqrt{n})\Sigma_{t=1}^n \psi^2(e_t/\tau^*)$ is bounded in probability and $\sqrt{n}(\hat{\theta}_n - \gamma_n^o)$ $= \sqrt{n}(\hat{\theta}_n - \gamma^*) + \sqrt{n}(\gamma^* - \gamma_n^o)$ which is bounded in probability.

The results on inference in Section 4 may be used in conjunction with the results of this section by substituting

$$s_n(\lambda),\ \mathscr{I}_n(\lambda),\ \mathscr{J}_n(\lambda),\ \mathscr{I},\ \mathscr{J}$$

as defined in this section for the definitions of Section 3. For the notation in connection with Theorem 7, substitute

NOTATION

$$\mathscr{I}^\circ = (1/n)\Sigma_{t=1}^{n}\int_{\mathcal{E}}\{(\partial/\partial\lambda)s[Y(e,x_t,\gamma_n^\circ),x_t,\tau_n^\circ,\lambda_n^\circ]\}$$

$$\times\ \{(\partial/\partial\lambda)s[Y(e,x_t,\gamma_n^\circ),x_t,\tau_n^\circ,\lambda_n^\circ]\}'dP(e)$$

$$\mathscr{J}^\circ = -(1/n)\Sigma_{t=1}^{n}\int_{\mathcal{E}}(\partial^2/\partial\lambda\partial\lambda')s[Y(e,x_t,\gamma_n^\circ),x_t,\tau_n^\circ,\lambda_n^\circ]\ dP(e)$$

$$V^\circ = (\mathscr{J}^\circ)^{-1}\mathscr{I}^\circ(\mathscr{J}^\circ)^{-1}$$

$$\alpha_n^\circ = n\ h^{\circ\prime}[H^\circ V^\circ H^{\circ\prime}]^{-1}h^\circ/2\quad .$$

7. EXAMPLES

Robust Nonlinear Regression

Recent literature:

Balet-Lawrence (1975), Grossman (1976), Ruskin (1978)

Model:

$$y_t = f(x_t, \theta_n^o) + e_t$$

e_t symmetrically distributed

Objective function:

$$s_n(\lambda) = (1/n)\Sigma_{t=1}^{n} \rho[y_t - f(x_t, \theta)]$$

$$\theta = \lambda$$

$\rho(u)$ an even function with $\rho(0) \le \rho(u)$

Asymptotic distribution parameters:

$$\mathcal{I} = \mathcal{E}\Psi^2(e)(F'F)$$

$$\mathcal{J} = \mathcal{E}\Psi'(e)(F'F)$$

$$F'F = \int_{\mathcal{X}} (\partial/\partial\theta) f(x, \theta^*)(\partial/\partial\theta') \; f(x, \theta^*) \; d\mu(x)$$

$$\Psi(u) = (d/du) \; \rho(u)$$

$$\Psi'(u) = (d/du) \; \Psi(u)$$

Iteratively Rescaled Robust Nonlinear Regression

Model:

$$y_t = f(x_t, \theta_n^o) + e_t$$

e_t symmetrically distributed

Objective function:

$$s_n(\lambda) = (1/n)\Sigma_{t=1}^{n} - \rho\{[y_t - f(x_t, \theta)]/\hat{\tau}_n\}$$

$$\lambda = \theta$$

$\rho(u)$ an even function with $\rho(0) \leq \rho(u)$

$\sqrt{n}(\hat{\tau}_n - \tau^*)$ is bounded in probability

Asymptotic distribution parameters:

$$\mathcal{I} = (1/\tau^*)^2 \, \mathcal{E}\Psi^2(e/\tau^*)(F'F)$$

$$\mathcal{J} = (1/\tau^*)^2 \, \mathcal{E}\Psi'(e/\tau^*)(F'F)$$

$$F'F = \int_{\mathcal{X}} (\partial/\partial\theta) f(x, \theta^*)(\partial/\partial\theta') \, f(x, \theta^*) \, d\mu(x)$$

$$\Psi'(u) = (d/du) \, \rho(u)$$

$$\Psi(u) = (d/du) \, \Psi(u)$$

Single Equation Nonlinear Least Squares

Recent literature:

Jennrich (1969), Malinvaud (1970), Gallant (1973, 1975a, 1975b)

Model:

$$y_t = f(x_t, \theta_n^\circ) + e_t$$

$$\mathcal{E}(e_t) = 0\ ,\ \mathcal{E}(e_t^2) = \sigma^2$$

Objective function:

$$s_n(\lambda) = (1/n)\Sigma_{t=1}^n - [y_t - f(x_t, \theta)]^2/(2\hat{\tau}_n)$$

$$\lambda = \theta$$

$\sqrt{n}(\hat{\tau}_n - \sigma^2)$ is bounded in probability

Asymptotic distribution parameters:

$$\mathcal{I} = \mathcal{J} = \sigma^{-2}\int_{\mathcal{X}} (\partial/\partial\theta)\ f(x, \theta^*)(\partial/\partial\theta')\ f(x, \theta^*)\ d\mu(x)$$

$$V = \mathcal{I}^{-1}$$

Comment:

$$\mathcal{I} = \mathcal{J}$$

<u>Multivariate Nonlinear Least Squares</u>

Recent literature:

Malinvaud (1970b), Gallant (1975c), Holly (1978)

Model:

$$y_t = f(x_t, \theta_n^o) + e_t$$

$$\mathcal{E}(e_t) = 0\ , \ \mathcal{E}(e_t e_t') = \Sigma$$

Objective function:

$$s_n(\lambda) = (1/n)\Sigma_{t=1}^{n}(-\tfrac{1}{2})[y_t - f(x_t,\theta)]'\hat{\tau}_n^{-1}[y_t - f(x_t,\theta)]$$

$$\lambda = \theta$$

$\sqrt{n}(\hat{\tau}_n - \Sigma)$ is bounded in probability

Asymptotic distribution parameters:

$$\mathcal{I} = \mathcal{J} = F'\Sigma^{-1}F = \int_{\mathcal{X}}[(\partial/\partial\theta')\ f(x,\theta^*)]'\Sigma^{-1}[(\partial/\partial\theta')f(x,\theta^*)]\ d\mu(x)$$

$$V = \mathcal{J}^{-1}$$

Comment:

$$\mathcal{I} = \mathcal{J}$$

Single Equation Maximum Likelihood

Model:

$$y_t = f(x_t, \theta_n^o) + e_t$$

$$\mathcal{E}(e_t) = 0 , \quad \mathcal{E}(e_t^2) = (\sigma^*)^2$$

Objective function:

$$s_n(\lambda) = (1/n)\Sigma_{t=1}^n (-\tfrac{1}{2})\{\ell n \ \sigma^2 + [y_t - f(x_t, \theta^*)]^2/\sigma^2\}$$

$$\lambda = (\theta', \sigma^2)'$$

Asymptotic distribution parameters:

$$\mathcal{I} = \begin{pmatrix} (\sigma^*)^{-2} F'F & \tfrac{1}{2}(\sigma^*)^{-6}\mathcal{E}(e^3) f \\ \tfrac{1}{2}(\sigma^*)^{-6}\mathcal{E}(e^3) f' & \tfrac{1}{4}(\sigma^*)^{-8}\mathrm{Var}(e^4) \end{pmatrix}$$

$$\mathcal{J} = \begin{pmatrix} (\sigma^*)^{-2} F'F & 0 \\ 0' & \tfrac{1}{2}(\sigma^*)^{-4} \end{pmatrix}$$

$$F'F = \int_{\mathcal{X}} (\partial/\partial\theta) f(x, \theta^*) (\partial/\partial\theta') \ f(x, \theta^*) \ d\mu(x)$$

$$f = \int_{\mathcal{X}} (\partial/\partial\theta) \ f(x, \theta^*) \ d\mu(x)$$

Comment:

For any hypothesis of the form $h(\theta) = 0$, $HVH' = H\mathcal{J}H'$. Under normality, $\mathcal{I} = \mathcal{J}$.

Multivariate Maximum Likelihood

Recent literature:

Malinvaud (1970b), Barnett (1976), Holly (1978)

Model:

$$y_t = f(x_t,\theta_n^o) + e_t$$

$$\mathcal{E}(e_t) = 0\ ,\quad \mathcal{E}(e_t e_t') = \Sigma^*$$

Objective function:

$$s_n(\lambda) = (1/n)\Sigma_{t=1}^n(-\tfrac{1}{2})\{\ell n \det \Sigma + [y_t - f(x_t,\theta)]'\Sigma^{-1}[y_t - f(x_t,\theta)]\}$$

$$\lambda = (\theta',\sigma')'$$

$\sigma' = (\sigma_{11}, \sigma_{12}, \ldots, \sigma_{1M}, \sigma_{2M}, \ldots, \sigma_{MM})$, upper triangle of Σ

$\text{vec}(\Sigma) = A\sigma$, A an $M^2 \times M(M+1)/2$ matrix of zeroes and ones

Asymptotic distribution parameters:

$$\mathcal{I} = \begin{pmatrix} F'\Sigma^{-1}F & \tfrac{1}{2}f'\Sigma^{-1}\mathcal{E}[e\ \text{vec}'(ee')](\Sigma\otimes\Sigma)^{-1}A \\ \text{sym} & \tfrac{1}{4}A'(\Sigma\otimes\Sigma)^{-1}\text{Var}[\text{vec}(ee')](\Sigma\otimes\Sigma)^{-1}A \end{pmatrix}$$

$$\mathcal{J} = \begin{pmatrix} F'\Sigma^{-1}F & 0 \\ 0 & \tfrac{1}{2}A'(\Sigma\otimes\Sigma)^{-1}A \end{pmatrix}$$

$$F'\Sigma^{-1}F = \int_{\mathcal{X}}[(\partial/\partial\theta')f(x,\theta^*)]'(\Sigma^*)^{-1}[(\partial/\partial\theta')f(x,\theta^*)]\ d\mu(x)$$

$$\Sigma^{-1}f = (\Sigma^*)^{-1}\int_{\mathcal{X}}(\partial/\partial\theta')\ f(x,\theta^*)\ d\mu(x)$$

$$(\Sigma\otimes\Sigma) = (\Sigma^*\otimes\Sigma^*)$$

Comment:

For any hypothesis of the form $h(\theta) = 0$, $HVH' = H\mathcal{J}^{-1}H'$. Under normality, $\mathcal{I} = \mathcal{J}$.

REFERENCES

Amemiya, T. (1974), "The nonlinear two-stage least squares estimator," Journal of Econometrics 2, 105-110.

_________ (1977), "The maximum likelihood estimator and the nonlinear three-stage least squares estimator in the general nonlinear simultaneous equation model," Econometrica 45, 955-968.

Balet-Lawrence, Sonia (1975), Estimation of the parameters in an implicit model by minimizing the sum of absolute value of order p. Unpublished Ph.D. Dissertation, North Carolina State University, Raleigh, NC.

Barnett, William A. (1976), "Maximum likelihood and iterated Aitken estimation of nonlinear systems of equations," Journal of the American Statistical Association 71, 354-360.

Chung, K. L. (1974), A Course in Probability, 2nd ed., Academic Press.

Durbin, J. (1970), "Testing for serial correlation in least-squares regression when some of the regressions are lagged dependent variables," Econometrica 38, 410-429.

Gallant, A. Ronald (1973), "Inference for nonlinear models," Institute of Statistics Mimeograph Series No. 875, North Carolina State University, Raleigh, NC.

_________ (1975a), "The power of the likelihood ratio test of location in nonlinear regression models," Journal of the American Statistical Association 70, 199-203.

_________ (1975b), "Testing a subset of the parameters of a nonlinear regression model," Journal of the American Statistical Association 70, 927-932.

_________ (1975c), "Seemingly unrelated nonlinear regressions," Journal of Econometrics 3, 35-50.

________________ (1977), "Three-stage least squares estimation for a system of simultaneous nonlinear implicit equations," Journal of Econometrics 5, 71-88.

________________ and Alberto Holly (1980), "Statistical inference in an implicit, nonlinear, simultaneous equation model in the context of maximum likelihood estimation," Econometrica 48, 697-720.

________________ and Dale W. Jorgenson (1979), "Statistical inference for a system of simultaneous, nonlinear implicit equations in the context of instrumental variable estimation," Journal of Econometrics, 11.

Grossman, W. (1976), "Robust nonlinear regression," in Johannes Gordesch and Peter Naeve, eds. Compstat 1976, Proceedings in Computational Statistics, Wien, Austria: Physica-Verlag.

Hartley, H. O. and A. Booker (1965), "Nonlinear least squares estimation," Annals of Mathematical Statistics 36, 638-650.

Holly, A. (1978), "Tests of nonlinear statistical hypotheses in multiple equation nonlinear models," Cashiers du Laboratoire d'Econometrie, Ecole Polytechnique, Paris.

Huber, Peter J. (1964), "Robust estimation of a location parameter," Annals Mathematical Statistics 35, 73-101.

Jennrich, Robert I. (1969), "Asymptotic properties of nonlinear least squares estimators," The Annals of Mathematical Statistics 40, 633-643.

Jorgenson, D. W. and J. Laffont (1974), "Efficient estimation of nonlinear simultaneous equations with additive disturbances," Annals of Economic and Social Measurement 3, 615-640.

Malinvaud, E. (1970a), "The consistency of nonlinear regressions," The Annals of Mathematical Statistics 41, 956-969.

_____________ (1970b), Statistical Methods of Econometrics. Amsterdam: North Holland. Chapter 9.

Ruskin, David M. (1978), M-Estimates of Nonlinear Regression Parameters and Their Jackknife Constructed Confidence Intervals. Ph.D. Dissertation, UCLA.

Souza, Geraldo and A. Ronald Gallant (1979), "Statistical inference based on M-estimators for the multivariate nonlinear regression model in implicit form," Institute of Statistics Mimeograph Series, No. 1229, North Carolina State University, Raleigh, NC.