THE INSTITUTE OF STATISTICS

UNIVERSITY OF NORTH CAROLINA SYSTEM

Nonlinear Statistical Models

by
A. Ronald Gallant

Chapter 4. Univariate Nonlinear Regression:
Asymptotic Theory

Institute of Statistics Mimeograph Series No. 1674
October 1985

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

NONLINEAR STATISTICAL MODELS

by

A. Ronald Gallant

CHAPTER 4. Univariate Nonlinear Regression: Asymptotic Theory

Copyright © 1983 by A. Ronald Gallant. All rights reserved.

NOT FOR GENERAL DISTRIBUTION

This copy is reproduced by special permission of the author and publisher from:

A. Ronald Gallant. Nonlinear Statistical Models. New York: John Wiley and Sons, Inc, forthcoming.

and is being circulated privately for discussion purposes only. Reproduction from this copy is expressly prohibited.

Please send comments and report errors to the following address:

A. Ronald Gallant
Institute of Statistics
North Carolina State University
Post Office Box 8203
Raleigh, NC 27695-8203

Phone 1-919-737-2531

NONLINEAR STATISTICAL MODELS

Table of Contents

Table of Contents (continued)

CHAPTER 4. Univariate Nonlinear Regression: Asymptotic Theory

In this chapter, the results of the previous chapter are specialized to the case of a correctly specified univariate nonlinear regression model estimated by least squares. Specialization is simply a matter of restating Assumptions 1 through 7 of Chapter 3 in context. This done, the asymptotic theory follows immediately. The characterizations used in Chapter 1 are established using probability bounds that follow from the asymptotic theory.

1. INTRODUCTION

Let us review some notation. The univariate nonlinear model is written as

$$y_t = f(x_t, \theta^\circ) + e_t \qquad t = 1, 2, \ldots, n$$

with θ° known to lie in some compact set Θ^*. The functional form of $f(x,\theta)$ is known, x is k-dimensional, θ is p-dimensional, and the model is assumed to be correctly specified. Following the conventions of Chapter 1, the model can be written in a vector notation as

$$y = f(\theta^\circ) + e$$

with the Jacobian of $f(\theta)$ written as $F(\theta) = (\partial/\partial\theta')f(\theta)$. The parameter θ is estimated by $\hat{\theta}$ that minimizes

$$s_n(\theta) = (1/n)\ \|y - f(\theta)\|^2 = (1/n)\Sigma_{t=1}^n[y_t - f(x_t,\theta)]^2 .$$

We are interested in testing the hypothesis

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

which we assume can be given the equivalent representation

$$H\colon \theta^\circ = g(\rho^\circ) \text{ for some } \rho^\circ \text{ against } A\colon \theta^\circ \neq g(\rho) \text{ for any } \rho$$

where $h\colon R^p \to R^q$, $g\colon R^r \to R^p$, and $p = r + q$. The correspondence with the notation of Chapter 3 is as follows.

NOTATION 1

General (Chapter 3)	Specific (Chapter 4)
$e_t = q(y_t, x_t, \gamma_n^o)$	$e_t = y_t - f(x_t, \theta_n^o)$
$\gamma \in \Gamma$	$\theta \in \Theta^*$
$y = Y(e, x, \gamma)$	$y = f(x, \theta) + e$
$s(y_t, x_t, \hat{\tau}_n, \lambda)$	$[y_t - f(x_t, \theta)]^2$
$\lambda \in \Lambda^*$	$\theta \in \Theta^*$
$s_n(\lambda) = (1/n)\Sigma_{t=1}^n s(y_t, x_t, \hat{\tau}_n, \lambda)$	$s_n(\theta) = (1/n)\Sigma_{t=1}^n [y_t - f(x_t, \theta)]^2$
$s_n^o(\lambda) = (1/n)\Sigma_{t=1}^n \int_{\mathcal{E}} s[Y(e, x_t, \gamma_n^o), x_t, \tau_n^o, \lambda] dP(e)$	$s_n^o(\theta) = \sigma^2 + (1/n)\Sigma_{t=1}^n [f(x_t, \theta_n^o) - f(x_t, \theta$
$s^*(\lambda) = \int_{\mathcal{X}} \int_{\mathcal{E}} s[Y(e, x, \gamma^*), x, \tau^*, \lambda] dP(e)\ d\mu(x)$	$s^*(\theta) = \sigma^2 + \int_{\mathcal{X}} [f(x, \theta^*) - f(x, \theta)]^2 d\mu(x)$
$\hat{\lambda}_n$ minimizes $s_n(\lambda)$	$\hat{\theta}_n$ minimizes $s_n(\theta)$
$\tilde{\lambda}_n$ minimizes $s_n(\lambda)$ subject to $h(\lambda) = 0$	$\tilde{\theta}_n = g(\hat{\rho}_n)$ minimizes $s_n(\theta)$ subject to $h(\theta) = 0$
λ_n^o minimizes $s_n^o(\lambda)$	θ_n^o minimizes $s_n^o(\theta)$
λ_n^* minimizes $s_n^o(\lambda)$ subject to $h(\lambda) = 0$	$\theta_n^* = g(\rho_n^o)$ minimizes $s_n^o(\theta)$ subject to $h(\theta) = 0$
λ^* minimizes $s^*(\lambda)$	θ^* minimizes $s^*(\theta)$

2. REGULARITY CONDITIONS

Application of the general theory to a correctly specified univariate nonlinear regression is just a matter of restating Assumptions 1 through 7 of Chapter 3 in terms of the notation above. As the data is presumed to be generated according to

$$y_t = f(x_t, \theta_n^\circ) + e_t \qquad t = 1, 2, \ldots, n$$

Assumptions 1 through 5 of Chapter 3 read as follows.

ASSUMPTION 1'. The errors are independently and identically distributed with common distribution $P(e)$. □

ASSUMPTION 2'. $f(x,\theta)$ is continuous on $\mathcal{X} \times \Theta^*$ and Θ^* is compact. □

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 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 $\int_{\mathcal{E}} \sup_{N_x} b(e,x)\, dP(e) < \infty$. □

ASSUMPTION 4'. (Identification) The parameter θ° is indexed by n and the sequence $\{\theta_n^\circ\}$ converges to θ^* .

$$s^*(\theta) = \sigma^2 + \int_{\mathcal{X}} [f(x,\theta^*) - f(x,\theta)]^2\, d\mu(x)$$

has a unique minimum over Θ^* at θ^* . □

ASSUMPTION 5'. Θ^* is compact, $[e + f(x,\theta^\circ) - f(x,\theta)]^2$ is dominated by $b(e,x)$; $b(e,x)$ is that of Assumption 3. □

The sample objective function is

$$s_n(\theta) = (1/n)\,\|y - f(\theta)\|^2$$

with expectation

$$s_n^\circ(\theta) = \sigma^2 + (1/n)\,\|f(\theta_n^\circ) - f(\theta)\|^2 .$$

By Lemma 1 of Chapter 3, both $s_n(\theta)$ and $s_n^\circ(\theta)$ have uniform, almost sure limit

$$s^*(\theta) = \sigma^2 + \int_{\mathcal{X}} [f(x,\theta^*) - f(x,\theta)]^2 \, d\mu(x) .$$

Note that the true value θ_n° of the unknown parameter is also a minimizer of $s_n^\circ(\theta)$ so that our use of θ_n° to denote them both is not ambiguous. We may apply Theorem 3 of Chapter 3 and conclude that

$$\lim_{n\to\infty} \theta_n^\circ = \theta^* ,$$

$$\lim_{n\to\infty} \hat{\theta}_n = \theta^* \quad \text{almost surely.}$$

Assumption 6 of Chapter 3 may be restated as follows.

ASSUMPTION 6'. Θ^* contains a closed ball Θ centered at θ^* with finite, nonzero radius such that

$$(\partial/\partial\theta_i)s[Y(e,x,\theta^\circ),x,\theta] = -2[e + f(x,\theta^\circ) - f(x,\theta)](\partial/\partial\theta_i)\, f(x,\theta)$$

$$(\partial^2/\partial\theta_i\partial\theta_j)s[Y(e,x,\theta^\circ),x,\theta] = 2[(\partial/\partial\theta_i)f(x,\theta)][(\partial/\partial\theta_j)f(x,\theta)]$$
$$- 2[e + f(x,\theta^\circ) - f(x,\theta)](\partial^2/\partial\theta_i\partial\theta_j)f(x,\theta)$$

$$\{(\partial/\partial\theta_i)s[Y(e,x,\theta^\circ),x,\theta]\}\{(\partial/\partial\theta_j)s[Y(e,x,\theta^\circ),x,\theta]\}$$
$$= 4[e + f(x,\theta^\circ) - f(x,\theta)]^2\,[(\partial/\partial\theta_i)f(x,\theta)][(\partial/\partial\theta_j)f(x,\theta)]$$

are continuous and dominated by $b(e,x)$ on $\mathcal{E} \times \mathcal{X} \times \Theta^* \times \Theta$ for $i, j = 1, 2, \ldots, p$. Moreover,

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

is nonsingular. □

Define

NOTATION 2

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

$$Q_n^\circ = (1/n)\ F'(\theta_n^\circ)\ F(\theta_n^\circ) \,,$$

$$Q_n^* = (1/n)\ F'(\theta_n^*)\ F(\theta_n^*) \,. \quad \square$$

Direct computation according to Notations 2 and 3 of Chapter 3 yields (Problem 1).

$$\mathcal{I}^* = 4\sigma^2\ Q$$

$$\mathcal{J}^* = 2\ Q$$

$$\mathcal{U}^* = 0$$

$$\mathcal{I}_n^\circ = 4\sigma^2\ Q_n^\circ$$

$$\mathcal{J}_n^\circ = 2\ Q_n^\circ$$

$$\mathcal{U}_n^\circ = 0$$

$$\mathcal{I}_n^* = 4\sigma^2\ Q_n^*$$

$$\mathcal{J}_n^* = 2\ Q_n^* - (2/n)\Sigma_{t=1}^n[f(x_t,\theta_n^\circ) - f(x_t,\theta_n^*)](\partial^2/\partial\theta\partial\theta')f(x_t,\theta_n^*)$$

$$\mathcal{U}_n^* = (4/n)\Sigma_{t=1}^n[f(x_t,\theta_n^\circ) - f(x_t,\theta_n^*)]^2[(\partial/\partial\theta)f(x_t,\theta_n^*)][(\partial/\partial\theta)\ f(x_t,\theta_n^*)]'$$

Noting that

$$(\partial/\partial\theta)s_n(\theta) = (-2/n)F'(\theta)[e + f(\theta_n^o) - f(\theta)]$$

we have from Theorem 4 of Chapter 3 that

$$(1/\sqrt{n})F'(\theta_n^o)e \xrightarrow{\mathcal{L}} N(0,\sigma^2 Q)$$

and from Theorem 5 that

$$\sqrt{n}\ (\hat\theta_n - \theta_n^o) \xrightarrow{\mathcal{L}} N(0,\sigma^2 Q^{-1})$$

$$\lim_{n\to\infty} Q_n^o = Q\ .$$

The Pitman drift assumption is restated as follows.

ASSUMPTION 7'. (Pitman drift) The sequence θ_n^o is chosen such that $\lim_{n\to\infty} \sqrt{n}(\theta_n^o - \theta_n^*) = \Delta$. Moreover, $h(\theta^*) = 0$.

Noting that

$$(\partial/\partial\theta)s_n^o(\theta) = (-2/n)F'(\theta)\ [f(\theta_n^o) - f(\theta)]$$

we have from Theorem 6 that

$$\lim_{n\to\infty} \tilde\theta_n = \theta^* \text{ almost surely}$$

$$\lim_{n\to\infty} \theta_n^* = \theta^*$$

$$\lim_{n\to\infty} Q_n^* = Q$$

$$(1/\sqrt{n})\ F'(\theta_n^*)\ e \xrightarrow{\mathcal{L}} N(0,\sigma^2 Q)$$

$$\lim_{n\to\infty} (1/\sqrt{n})\ F'(\theta_n^*)[f(\theta_n^o) - f(\theta_n^*)] = Q\Delta\ .$$

Assumption 13 of Chapter 3 is restated as follows.

ASSUMPTION 13'. The function $h(\theta)$ is a once continuously differentiable mapping of Θ into R^q. Its Jacobian $H(\theta) = (\partial/\partial\theta')h(\theta)$ has full rank $(=q)$ at $\theta = \theta^*$. □

PROBLEMS

1. Use the derivatives given in Assumption 6 to compute $\bar{\bar{\mathfrak{I}}}(\theta)$, $\bar{\bar{\mathcal{J}}}(\theta)$, $\bar{\bar{\mathfrak{u}}}(\theta)$ and $\bar{\mathfrak{I}}(\theta)$, $\bar{\mathcal{J}}(\theta)$, $\bar{\mathfrak{u}}(\theta)$ as defined in Notations 2 and 3 of Chapter 3.

3. CHARACTERIZATIONS OF LEAST SQUARES ESTIMATORS AND TEST STATISTICS

The first of the characterizations appearing in Chapter 1 is

$$\hat{\theta}_n = \theta_n^\circ + [F'(\theta_n^\circ)\ F(\theta_n^\circ)]^{-1}F'(\theta_n^\circ)e + o_p(1/\sqrt{n}) \ .$$

It is derived using the same sort of arguments as used in the proof of Theorem 5 of Chapter 3 so we shall be brief here; one can look at Theorem 5 for details. By Lemma 2 of Chapter 3 we may assume without loss of generality that $\hat{\theta}_n$ and θ_n° are in Θ and that $(\partial/\partial\theta)s_n(\hat{\theta}_n) = o_p(1/\sqrt{n})$. Recall that $Q_n^\circ = Q + o(1)$ whence $\mathcal{J}_n^\circ = \mathcal{J}^* + o(1)$. By Taylor's theorem

$$\sqrt{n}\ (\partial/\partial\theta)s_n(\theta_n^\circ) = \sqrt{n}\ (\partial/\partial\theta)s_n(\hat{\theta}_n) + \bar{\mathcal{J}}\ \sqrt{n}\ (\theta_n^\circ - \hat{\theta}_n)$$

where $\bar{\mathcal{J}} = \mathcal{J}^* + o_s(1)$. Then

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

which can be rewritten as

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

Now $[\mathcal{J}^* - \mathcal{J}_n^\circ + o_s(1)] = o_s(1)$ and $\sqrt{n}\ (\hat{\theta}_n - \theta_n^\circ) \xrightarrow{\mathcal{L}} N(0,\sigma^2 Q)$ which implies that $\sqrt{n}\ (\hat{\theta}_n - \theta_n^\circ) = O_p(1)$ whence $[\mathcal{J}^* - \mathcal{J}_n^\circ + o_s(1)]\ \sqrt{n}\ (\hat{\theta}_n - \theta_n^\circ) = o_p(1)$. Thus we have that

$$\mathcal{J}_n^\circ\ \sqrt{n}\ (\hat{\theta}_n - \theta_n^\circ) = \sqrt{n}\ (\partial/\partial\theta)s_n(\theta_n^\circ) + o_p(1) \ .$$

There is an N such that for $n > N$ the inverse of $\mathcal{J}_n^\circ$ exists whence

$$\sqrt{n}\ (\hat{\theta}_n - \theta_n^\circ) = -\sqrt{n}\ (\mathcal{J}_n^\circ)^{-1}(\partial/\partial\theta)s_n(\theta_n^\circ) + o_p(1)$$

or

$$\hat{\theta}_n = \theta_n^\circ - (\mathscr{J}_n^\circ)^{-1}(\partial/\partial\theta)s_n(\theta_n^\circ) + o_p(1/\sqrt{n}) \ .$$

Finally, $-(\mathscr{J}_n^\circ)^{-1}(\partial/\partial\theta)s_n(\theta_n^\circ) = [F'(\theta_n^\circ)F(\theta_n^\circ)]^{-1}F'(\theta_n^\circ)e$ which completes the argument.

The next characterization that needs justification is

$$s^2 = e'\{I - F(\theta_n^\circ)[F'(\theta_n^\circ)F(\theta_n^\circ)]^{-1}F'(\theta_n^\circ)\}e/(n-p) + o_p(1/n) \quad .$$

The derivation is similar to the arguments used in the proof of Theorem 15 of Chapter 3; again we shall be brief and one can look at the proof of Theorem 15 for details. By Taylor's theorem

$$\begin{aligned} n[s_n(\theta_n^\circ) - s_n(\hat{\theta}_n)] &\\ &= n[(\partial/\partial\theta)s_n(\hat{\theta}_n)]'(\theta_n - \theta_n^\circ) \\ &\quad + (n/2)(\hat{\theta}_n - \theta_n^\circ)'[(\partial^2/\partial\theta\partial\theta')s_n(\bar{\theta}_n)](\hat{\theta}_n - \theta_n^\circ) \\ &= n\ o_s(1/\sqrt{n})(\hat{\theta}_n - \theta_n^\circ) + (n/2)(\hat{\theta}_n - \theta_n^\circ)'[\mathscr{J}_n^\circ + o_s(1)](\hat{\theta}_n - \theta_n^\circ) \\ &= (n/2)(\hat{\theta}_n - \theta_n^\circ)'\mathscr{J}_n^\circ(\hat{\theta}_n - \theta_n^\circ) + o_p(1) \quad . \end{aligned}$$

From the proceeding result we have

$$(\hat{\theta}_n - \theta_n^\circ) = [F'(\theta_n^\circ)F(\theta_n^\circ)]^{-1}F'(\theta_n^\circ)e + o_p(1/\sqrt{n})$$

whence

$$n[s_n(\theta_n^\circ) - s_n(\hat{\theta}_n)] = n\ e'F(\theta_n^\circ)[F'(\theta_n^\circ)F(\theta_n^\circ)]^{-1}F'(\theta_n^\circ)e + o_p(1) \ .$$

This equation reduces to

$$\|y - f(\hat{\theta})\|^2 = e'\{I - F(\theta_n^\circ)[F'(\theta_n^\circ)F(\theta_n^\circ)]^{-1}F'(\theta_n^\circ)\}e + o_p(1/n)$$

which completes the argument.

Next we show that

$$h(\hat{\theta}_n) = h(\theta_n^\circ) + H(\theta_n^\circ)[F'(\theta_n^\circ)F(\theta_n^\circ)]^{-1}F'(\theta_n^\circ)e + o_p(1/\sqrt{n}) \; .$$

A straightforward argument using Taylor's theorem yields

$$\sqrt{n}\; h(\hat{\theta}_n) = \sqrt{n}\; h(\theta_n^\circ) + \bar{H}\sqrt{n}\;(\hat{\theta}_n - \theta_n^\circ)$$

where $\bar{H}$ has rows $(\partial/\partial\theta')\; h(\bar{\theta}_i)$ with $\bar{\theta}_i = \lambda_i\hat{\theta}_n + (1-\lambda_i)\theta_n^\circ$ for some λ_i with $0 \leq \lambda_i \leq 1$ whence

$$\sqrt{n}\; h(\hat{\theta}_n) = \sqrt{n}\; h(\theta_n^\circ) + [H(\theta_n^\circ) + o_s(1)]\sqrt{n}\;(\hat{\theta}_n - \theta_n^\circ) \quad .$$

Since $\sqrt{n}\;(\hat{\theta}_n - \theta_n^\circ)$ is bounded in probability we have

$$\begin{aligned}
\sqrt{n}\; h(\hat{\theta}_n) &= \sqrt{n}\; h(\theta_n^\circ) + \sqrt{n}\; H(\theta_n^\circ)(\hat{\theta}_n - \theta_n^\circ) + o_s(1) \\
&= \sqrt{n}\; h(\theta_n^\circ) + H(\theta_n^\circ)\sqrt{n}\;\{[F'(\theta_n^\circ)F(\theta_n^\circ)]^{-1}F'(\theta_n^\circ)e + o_p(1/\sqrt{n})\} + o_s(1) \\
&= \sqrt{n}\; h(\theta_n^\circ) + \sqrt{n}\; H(\theta_n^\circ)[F'(\theta_n^\circ)F(\theta_n^\circ)]^{-1}F'(\theta_n^\circ)e + o_p(1) \; .
\end{aligned}$$

We next show that

$$1/s^2 = (n-p)/e'(I-P_F)e + o_p(1/n)$$

where

$$P_F = F(\theta_n^\circ)[F'(\theta_n^\circ)F(\theta_n^\circ)]^{-1}F'(\theta_n^\circ) \; .$$

Fix a realization of the errors $\{e_t\}$ for which $\ell im_{n\to\infty} s^2 = \sigma^2$ and $\ell im_{n\to\infty} e'(I-P_F)e/(n-p) = \sigma^2$; almost every realization is such (Problem 2). Choose N so that if $n > N$ then $s^2 > 0$ and $e'(I-P_F)e > 0$. Using

$$s^2 = e'(I-P_F)e/(n-p) + o_p(1/n)$$

and Taylor's theorem we have

$$1/s^2 = (n-p)/e'(I-P_F)e - [(n-p)/e'(I-P_F)e]^2 o_p(1/n) \; .$$

The term $[(n-p)/e'(I-P_F)e]^2$ is bounded for $n > N$ because $\ell im_{n\to\infty}[(n-p)/e'(I-P_F)e]^2 = 1/\sigma^4$. One concludes that $1/s^2 = (n-p)/e'(I-P_F)e + o_p(1/n)$ which completes the argument.

The next task is to show that if the errors are normally distributed then

$$W = Y + o_p(1)$$

where

$$Y \sim F'(q,\ n-p,\ \lambda)$$

$$\lambda = h'(\theta_n^\circ)\{H(\theta_n^\circ)[F'(\theta_n^\circ)F(\theta_n^\circ)]^{-1}H'(\theta_n^\circ)\}^{-1}h(\theta_n^\circ)/(2\sigma^2) .$$

Now

$$W = n\ h'(\hat\theta_n)\{\hat H[(1/n)F'(\hat\theta_n)F(\hat\theta_n)]^{-1}\hat H'\}^{-1}h(\hat\theta_n)/(qs^2)$$

and as notation write

$$\sqrt{n}\ h(\hat\theta_n) = \sqrt{n}\ h(\theta_n^\circ) + \sqrt{n}\ H(\theta_n^\circ)[F'(\theta_n^\circ)F(\theta_n^\circ)]^{-1}F'(\theta_n^\circ)e + o_p(1)$$

$$= \mu + U + o_p(1)$$

$$\{\hat H[(1/n)F'(\hat\theta_n)F(\hat\theta_n)]^{-1}\hat H'\}^{-1} = \{H(\theta_n^\circ)[(1/n)F'(\theta_n^\circ)F(\theta_n^\circ)]^{-1}H'(\theta_n^\circ)\}^{-1} + o_p(1)$$

$$= A^{-1} + o_p(1)$$

whence

$$W = [\mu + U + o_p(1)]'A^{-1}[\mu + U + o_p(1)][(n-p)/e'(I-P_F)e + o_p(1)]/q$$

$$= \frac{(\mu+U)'A^{-1}(\mu+U)/(q\sigma^2)}{e'(I-P_F)e/[\sigma^2(n-p)]} + o_p(1)$$

$$= Y + o_p(1) .$$

Assuming normal errors then

$$U \sim N_q(0,\ \sigma^2 A)$$

which implies that (Appendix 1)

$$(\mu+U)'A^{-1}(\mu+U)/\sigma^2 \sim \chi^{2'}(q,\lambda)$$

with

$$\lambda = \mu'A^{-1}\mu/(2\sigma^2)$$

$$= n\ h'(\theta_n^\circ)\{H(\theta_n^\circ)[(1/n)F'(\theta_n^\circ)F(\theta_n^\circ)]^{-1}H'(\theta_n^\circ)\}^{-1}h(\theta_n^\circ)/(2\sigma^2)\ .$$

Since $A(I-P_F) = 0$, U and $(I-P_F)e$ are independently distributed whence $(\mu+U)'A^{-1}(\mu+U)$ and $e'(I-P_F)e = e'(I-P_F)'(I-P_F)e$ are independently distributed. This implies that $Y \sim F'(q,n-p,\lambda)$ which completes the argument.

Simply by rescaling s^2 in the foregoing we have that

$$(SSE_{full})/n = e'P_F^\perp e/n + o_p(1/n)$$

$$n/(SSE_{full}) = n/e'P_F^\perp e + o_p(1/n)$$

where

$$P_F^\perp = I - P_F = I - F(\theta_n^\circ)[F'(\theta_n^\circ)F(\theta_n^\circ)]^{-1}F'(\theta_n^\circ)\ ;$$

recall that

$$SSE_{full} = \|y - f(\hat\theta_n)\|^2$$

$$SSE_{reduced} = \|y - f(\tilde\theta_n)\|^2 = \|y - f[g(\hat\rho_n)]\|^2\ .$$

The claim that

$$(SSE_{reduced})/n = (e+\delta)'P_{FG}^\perp(e+\delta)/n + o_p(1/n)$$

with

$$\delta = f(\theta_n^\circ) - f(\theta_n^*) = f(\theta_n^\circ) - f[g(\rho_n^\circ)]$$

$$P_{FG}^\perp = I - P_{FG} = I - F(\theta_n^\circ)G(\rho_n^\circ)[G'(\rho_n^\circ)F'(\theta_n^\circ)F(\theta_n^\circ)G(\rho_n^\circ)]^{-1}G'(\rho_n^\circ)F'(\theta_n^\circ)$$

comes fairly close to being a restatement of a few lines of the proof of

Theorem 13 of Chapter 3. In that proof we find the equations

$$\bar{H}\sqrt{n}\,(\tilde{\theta}_n - \theta_n^*) = o_s(1)$$

$$\sqrt{n}\,(\tilde{\theta}_n - \theta_n^*) = \bar{\mathcal{J}}^{-1}\sqrt{n}\,(\partial/\partial\theta)s_n(\tilde{\theta}_n) - \bar{\mathcal{J}}^{-1}\sqrt{n}\,(\partial/\partial\theta)s_n(\theta_n^*) + o_s(1)$$

which, using arguments that have become repetitive at this point, can be rewritten as

$$H\sqrt{n}\,(\tilde{\theta}_n - \theta_n^*) = o_s(1)$$

$$\sqrt{n}\,(\tilde{\theta}_n - \theta_n^*) = \mathcal{J}^{-1}[\sqrt{n}\,(\partial/\partial\theta)s_n(\tilde{\theta}_n) - \sqrt{n}\,(\partial/\partial\theta)s_n(\theta_n^*)] + o_p(1)$$

with $\mathcal{J} = \mathcal{J}_n^{\circ}$ and $H = H(\theta_n^*)$. Using the conclusion of Theorem 13 of Chapter 3 one can substitute for $\sqrt{n}\,(\partial/\partial\theta)s_n(\tilde{\theta}_n)$ to obtain

$$\sqrt{n}\,[(\partial/\partial\theta)s_n(\tilde{\theta}_n)]'\sqrt{n}\,(\tilde{\theta}_n - \theta_n^*) = o_p(1)$$

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

Then using Taylor's theorem

$$n[s_n(\tilde{\theta}_n) - s_n(\theta_n^*)]$$

$$= -n[(\partial/\partial\theta)s_n(\tilde{\theta}_n)]'(\tilde{\theta}_n - \theta_n^*) - (n/2)(\tilde{\theta}_n - \theta_n^*)'[\mathcal{J} + o_s(1)](\tilde{\theta}_n - \theta_n^*)$$

$$= (-n/2)(\tilde{\theta}_n - \theta_n^*)'\mathcal{J}(\tilde{\theta}_n - \theta_n^*) + o_p(1)$$

$$= (-n/2)[(\partial/\partial\theta)s_n(\theta_n^*)]'[\mathcal{J}^{-1} - \mathcal{J}^{-1}H'(H\mathcal{J}^{-1}H')^{-1}H\mathcal{J}^{-1}][(\partial/\partial\theta)s_n(\theta_n^*)]\ .$$

Using the identify obtained in Section 6 of Chapter 3 we have

$$\mathcal{J}^{-1} - \mathcal{J}^{-1}H'(H\mathcal{J}^{-1}H')^{-1}H\mathcal{J}^{-1} = G(G'\mathcal{J}G)^{-1}G'$$

whence

$$n\, s_n(\tilde{\theta}_n) = n s_n(\theta_n^*) - (n/2)[(\partial/\partial\theta)s_n(\theta_n^*)]'G(G'\mathcal{J}\, G)^{-1}G'[(\partial/\partial\theta)s_n(\theta_n^*)] + o_p(1)$$

Using Taylor's theorem, the Uniform Strong Law, and the Pitman drift assumption we have

$$\begin{aligned}
(\partial/\partial\theta)s_n(\theta_n^*) &= (-2/n)\Sigma_{t=1}^n[e_t + f(x_t,\theta_n^\circ) - f(x_t,\theta_n^*)](\partial/\partial\theta)f(x_t,\theta_n^*) \\
&= (-2/n)\Sigma_{t=1}^n[e_t + f(x_t,\theta_n^\circ) - f(x_t,\theta_n^*)](\partial/\partial\theta)f(x_t,\theta_n^\circ) \\
&\quad + (1/\sqrt{n})(-2/n)\Sigma_{t=1}^n[e_t + f(x_t,\theta_n^\circ) - f(x_t,\theta_n^*)] \\
&\quad \times \begin{pmatrix} (\partial/\partial\theta')(\partial/\partial\theta_1)f(x_t,\bar{\theta}_{1n}) \\ \vdots \\ (\partial/\partial\theta')(\partial/\partial\theta_p)f(x_t,\bar{\theta}_{pn}) \end{pmatrix} \sqrt{n}\,(\theta_n^\circ - \theta_n^*) \\
&= (-2/n)\, F'(\theta_n^\circ)(e + \delta) + o_p(1/\sqrt{n}) \ .
\end{aligned}$$

Substitution and algebraic reduction yields (Problem 3)

$$n\, s_n(\tilde{\theta}_n) = (e + \delta)'(e + \delta) - (e + \delta)'P_{FG}(e + \delta) + o_p(1)$$

which proves the claim.

The following are the characterizations used in Chapter 1 that have not yet been verified

$$(SSE_{reduced})/(SSE_{full}) = (e+\delta)' P^{\perp}_{FG}(e+\delta)/e' P^{\perp}_{F} e = o_p(1/n)$$

$$\tilde{D}'(\tilde{F}'\tilde{F})^{-1}\tilde{D}/n = (e+\delta)'(P_F - P_{FG})(e+\delta)/n + o_p(1/n)$$

$$\frac{\tilde{D}'(\tilde{F}'\tilde{F})\tilde{D}/q}{SSE(\hat{\theta})/(n-p)} = \frac{(e+\delta)'(P_F - P_{FG})(e+\delta)/q}{e'(I - P_F)e/(n-p)} + o_p(1)$$

$$\frac{n\ \tilde{D}'(\tilde{F}'\tilde{F})\ \tilde{D}}{SSE(\tilde{\theta})} = \frac{n(e+\delta)'(P_F - P_{FG})(e+\delta)}{(e+\delta)'(I - P_{FG})(e+\delta)} + o_p(1) \quad .$$

Except for the second, these are obvious at sight. Let us sketch the verification of the second characterization

$$\begin{aligned}
\tilde{D}'\tilde{F}'\tilde{F}\ \tilde{D} &= [y - f(\tilde{\theta}_n)]\tilde{F}(\tilde{F}'\tilde{F})^{-1}\tilde{F}'[y - f(\tilde{\theta}_n)] \\
&= (n/4)[(\partial/\partial\theta)s_n(\tilde{\theta}_n)]'[(1/n)\tilde{F}'\tilde{F}]^{-1}[(\partial/\partial\theta)s_n(\tilde{\theta}_n)] \\
&= (n/2)[(\partial/\partial\theta)s_n(\tilde{\theta}_n)]'[\mathcal{J} + o_s(1)]^{-1}[(\partial/\partial\theta)s_n(\tilde{\theta}_n)] \\
&= (n/2)[(\partial/\partial\theta)s_n(\theta^*_n)]'\mathcal{J}^{-1}H'(H\mathcal{J}^{-1}H')^{-1}H\mathcal{J}^{-1}[(\partial/\partial\theta)s_n(\theta^*_n)] + o_p(1) \\
&= (n/2)[(\partial/\partial\theta)s_n(\theta^*_n)]'[\mathcal{J}^{-1} - G(G'\mathcal{J}G)^{-1}G'][(\partial/\partial\theta)s_n(\theta^*_n)] + o_p(1) \\
&= (1/n)(e+\delta)'F(\theta^\circ_n)[(Q^\circ_n)^{-1} - G(G'Q^\circ_n G)^{-1}G']F'(\theta^\circ_n)(e+\delta) + o_p(1) \\
&= (1/n)(e+\delta)'F(\theta^\circ_n)[(Q^\circ_n)^{-1} - G(G'Q^\circ_n G)^{-1}G']F'(\theta^\circ_n)(e+\delta) + o_p(1) \\
&= (e+\delta)'(P_F - P_{FG})(e+\delta) + o_p(1) \ .
\end{aligned}$$

PROBLEMS

1. Give a detailed derivation of the four characterizations listed in the preceding paragraph.

2. Cite the theorem which permits one to claim that $\ell im_{n\to\infty} s^2 = \sigma^2$ almost surely and prove directly that $\ell im_{n\to\infty} e'(I - P_F)e/(n-p) = \sigma^2$ almost surely.

3. Show in detail that $(\partial/\partial\theta)s_n(\theta_n^*) \doteq (-2/n)F'(\theta_n^\circ)(e+\delta) + o_p(1/\sqrt{n})$ suffices to reduce $(n/2)[(\partial/\partial\theta)s_n(\theta_n^*)]'G(G'\mathcal{J}\,G)^{-1}G'[(\partial/\partial\theta)s_n(\theta_n^*)]$ to $(e+\delta)'P_{FG}(e+\delta)$.

4. REFERENCES

Gallant, A. Ronald and Alberto Holly (1980), "Statistical Inference in an Implicit, Nonlinear, Simultaneous Equation Model in the Context of Maximum Likelihood Estimation," Econometrica 48, 697-720.

5. INDEX TO CHAPTER 4