ON THE ASYMPTOTIC POWER OF THE LACK OF FIT TEST IN NONLINEAR REGRESSION

by

A. Ronald Gallant

Institute of Statistics
Mimeo Series No. 1083
Raleigh - August 1976

Some notation which is used throughout is set forth here for convenient reference.

Notation:

$y = (y_1, y_2, \ldots, y_n)'$ $(n \times 1)$,

$g(\Psi) = [g(x_1,\Psi), g(x_2,\Psi), \ldots, g(x_n, \Psi)]'$ $(n \times 1)$,

$h(\omega) = [h(x_1,\omega), \ldots, h(x_n,\omega)]'$ $(n \times 1)$,

$e = (e_1, e_2, \ldots, e_n)$ $(n \times 1)$,

$G(\Psi)$ = the n by u matrix with typical element $(\partial/\partial\Psi_j)\, g(x_t, \Psi)$ where t is the row index,

Z = the n by w matrix with rows z_t' ,

$g = g(\Psi^*)$,

$h = h(\omega^*)$,

$G = G(\Psi^*)$,

$P_G = G(G'G)^{-1}G'$ $(n \times n)$,

$P_{GZ} = [G \vdots Z]\{[G \vdots Z]'[G \vdots Z]\}^{-1}[G \vdots Z]$ $(n \times n)$,

$Q_G = I - P_G$,

$Q_{GZ} = I - P_{GZ}$.

The following regularity conditions govern throughout.

Assumptions: The sequence of inputs $\{(x_t,z_t)\}_{t=1}^{n}$ are chosen from $\mathcal{X} \times \mathcal{Z}$, where $\mathcal{X}$ and $\mathcal{Z}$ are compact, such that the measure μ_n defined on the Borel subsets of $\mathcal{X} \times \mathcal{Z}$ by

$$\mu_n(A) = \text{the proportion of } (x_t,z_t) \text{ in } A \text{ for } t \leq n$$

converges weakly to a measure μ defined on the Borel subsets of $\mathcal{X} \times \mathcal{Z}$; see [2]. The sets Ψ and Δ are compact. The functions $g(x,\Psi)$, $(\partial/\partial\Psi_i)\, g(x,\Psi)$, and

This technical note serves as an appendix to [4]. Ambiquities as to objectives, definitions, notation, etc. may be resolved by reference to [4].

Consider testing

$$H: \quad y_t = g(x_t, \Psi) + e_t$$

against

$$A^{\#}: \quad y_t = g(x_t, \Psi) + z_t' \delta + e_t$$

using the likelihood ratio test, assuming normal errors with variance unknown. The true model is, however,

$$A: \quad y_t = g(x_t, \Psi) + \tau\, h(x_t, \omega) + e_t \quad .$$

An asymptotic approximation of the power of the test against A, not $A^{\#}$, is required.

In the above expressions, y_t is univariate, x_t is k-dimensional, z_t is w-dimensional, and the index $t = 1,2,\ldots, n$. The parameter Ψ is u-dimensional, τ is univariate, ω is v-dimensional, and δ is ω-dimensional. The asterisk is used in connection with parameters - Ψ^*, τ^*, and ω^* - to emphasize that it is the true but unknown value which is meant. The omission of the asterisk does not necessarily denote the contrary. The parameter Ψ is contained in a compact set Ψ. For convenience, δ is constrained to lie in a compact set Δ but this assumption may be eliminated if desired; see [1].

The test statistic, itself, is

$$T^{\#} = \tilde{\sigma}^2/(\sigma^2)^{\#}$$

where: $\tilde{\Psi}$ minimizes $\Sigma_{t=1}^{n}[y_t - g(x_t, \Psi)]^2$, and $\tilde{\sigma}^2 = (1/n)\, \Sigma_{t=1}^{n}[y_t - g(x_t, \tilde{\Psi})]^2$;
$(\Psi^{\#}, \delta^{\#})$ minimizes $\Sigma_{t=1}^{n}[y_t - g(x_t, \Psi) - z_t'\delta]^2$, and $(\sigma^2)^{\#} = (1/n)\, \Sigma_{t=1}^{n}[y_t - g(x_t, \Psi^{\#}) - z_t'\delta^{\#}]^2$

One rejects when $T^{\#}$ is larger than

$$c^* = 1 + w\, F_\alpha/(n-u-w)$$

where F_α denotes the upper $\alpha\cdot 100$ percentage point of an F random variable with w numerator degrees freedom and n-u-w denominator degrees freedom.

$(\partial^2/\partial\psi_i\partial\psi_j)\, g(x, \psi)$ are continuous in (x, ψ) on $\mathfrak{X} \times \Psi$. The function $h(x, \omega^*)$ is continuous on $\mathfrak{X}$. The true value ψ^* is contained in an open set which is, in turn, contained in Ψ ; Δ contains an open neighborhood of the zero vector. If $g(x, \psi) = g(x, \psi^*)$ except on a set of μ measure zero, it is assumed that $\psi = \psi^*$; likewise, $g(x, \psi) + z'\delta = g(x, \psi^*)$ a.e. implies $\psi = \psi^*$ and $\delta = 0$. The matrix $\lim_{n\to\infty}[G\vdots Z]'[G\vdots Z]$ is non-singular. (The limit exists by Lemma 1 of [3].). As n increases, $\sqrt{n}\,\tau^*$ tends to a finite limit. The errors $\{e_t\}$ are independently and normally distributed each with mean zero and unknown variance σ^2.

<u>Lemma 1</u>: The random variables $\tilde{\psi}$ and $\psi^{\#}$ converge almost surely to ψ^*. The random variable $\delta^{\#}$ converges almost surely to the zero vector. The random variables $\tilde{\sigma}^2$ and $(\sigma^2)^{\#}$ converge almost surely to σ^2.

<u>Proof</u>: Denote τ^* by τ^*_n to emphasize the assumed variation with n.

Consider the sequence of random variables

$$Q_n(\psi, \tau) = (1/n)\, \| e + g + \tau h - g(\psi) \|^2$$

$$= e'e/n + 2e'[g + \tau h - g(\psi)]/n + [g + \tau h - g(\psi)]'[g + \tau h - g(\psi)]/n .$$

Note that $\tilde{\psi}$ minimizes $Q_n(\psi, \tau^*_n)$ for each realization of e. $Q_n(\psi, \tau)$ converges almost surely to

$$\bar{Q}(\psi, \tau) = \sigma^2 + \int [g(x, \psi^*) + \tau h(x, \omega^*) - g(x, \psi)]^2\, d\mu(x, z)$$

uniformly in (ψ, τ) over $\Psi \times [-\frac{1}{2}, \frac{1}{2}]$ by the Strong Law of Large Numbers and Lemma 1 of [3], Parts 1 and 2. Consider a realization of the errors $\{e_t\}_{t=1}^{\infty}$ which does not belong to the exceptional set. Let $\{\tilde{\psi}_n\}$ be the sequence of points minimizing $Q_n(\psi, \tau^*_n)$ corresponding to this realization. Since Ψ is compact, there is at least one limit point $\bar{\psi}$ and at least one subsequence $\{\tilde{\psi}_{n_m}\}$ such that $\lim_{m\to\infty} \tilde{\psi}_{n_m} = \bar{\psi}$. As a direct consequence of the uniform convergence of the continuous functions $Q_n(\psi, \tau)$ to $\bar{Q}(\psi, \tau)$, we have

$$\sigma^2 \le \bar{Q}(\bar{\Psi}, 0)$$

$$= \lim_{m\to\infty} Q_{n_m}(\tilde{\Psi}_{n_m}, \tau^*_{n_m})$$

$$\le \lim_{m\to\infty} Q_{n_m}(\Psi^*, \tau^*_{n_m})$$

$$= \bar{Q}(\Psi^*, 0)$$

$$= \sigma^2 .$$

This implies

$$\int[g(x, \Psi^*) + \tau h(x, \omega^*) - g(x, \Psi)]^2 \, d\mu(x, z) = 0$$

which implies $\bar{\Psi} = \Psi^*$ by assumption. Thus, the sequence $\{\tilde{\Psi}_n\}$ has only one limit point Ψ^* ; moreover, this implies

$$\lim_{n\to\infty} \tilde{\sigma}^2_n = \lim_{n\to\infty} Q(\tilde{\Psi}_n, \tau^*_n) = \bar{Q}(\Psi^*, 0) = \sigma^2 .$$

Next, consider the sequence of random variables

$$Q_n(\Psi,\delta,\tau) = (1/n)\|e + g + \tau h - g(\Psi) - Z\delta\|^2 .$$

Note that $(\Psi^{\#},\delta^{\#})$ minimizes $Q_n(\Psi,\delta,\tau^*_n)$ for each realization of e. $Q_n(\Psi,\delta,\tau)$ converges almost surely to

$$\bar{Q}(\Psi,\delta,\tau) = \sigma^2 + \int[g(x,\Psi^*) + \tau h(x,\omega^*) - g(x,\Psi) - z'\delta]^2 \, d\mu(x,z)$$

uniformly in (Ψ,δ,τ) over $\Psi \times \Delta \times [-\frac{1}{2},\frac{1}{2}]$ as above. The remainder of the proof is entirely analagous to the above with $(\Psi^{\#},\delta^{\#})$ replacing $\tilde{\Psi}$ throughout. $\square$

<u>Lemma 2</u>: The random vector $(1/\sqrt{n})\,[G \vdots Z]'(e + \tau^* h)$ converges in distribution to a u + w - variate normal.

<u>Proof</u>: By Lemma 3.5 of [1], $(1/\sqrt{n})\,[G \vdots Z]'e$ converges in distribution to a u + w - variate normal. By Part 1 of Lemma 1 of [3],

$$\lim_{n\to\infty}(1/\sqrt{n})\,[G \vdots Z]'(\tau^* h)$$

$$= (\lim_{n\to\infty}\sqrt{n}\,\tau^*)\,[\lim_{n\to\infty}(1/n)[G \vdots Z]'h]. \ \square$$

<u>Theorem 1</u>. The random variable $(\sigma^2)^{\#}$ may be characterized as

$$(\sigma^2)^{\#} = (e + \tau^* h)' Q_{GZ}(e + \tau^* h)/n + a_n$$

where $n\, a_n$ converges in probability to zero.

The random variable $\tilde{\sigma}^2$ may be characterized as

$$\tilde{\sigma}^2 = (e + \tau^* h)' Q_Z (e + \tau^* h)/n + b_n$$

where $n\, b_n$ converges in probability to zero.

<u>Proof</u>: Recall that τ^* varies with n. By Lemma 1, $(\Psi^{\#},\delta^{\#})$ will almost surely be contained in an open subset of $\Psi \times \Delta$ containing $(\Psi^*,0)$, allowing the use of Taylor's expansions in the proof and causing $(\Psi^{\#},\delta^{\#})$ to eventually become a stationary point of

$$Q_n(\Psi,\delta,\tau^*) = (1/n)\| y-g(\Psi) - Z\delta\|^2 \quad .$$

We will now obtain intermediate results based on Taylor's expansions which will be used later in the proof. By Taylor's theorem,

$$g(\Psi^{\#}) + Z\delta^{\#} = g + G(\Psi^{\#}-\Psi^*) + Z\delta^{\#} + D(\Psi^{\#}-\Psi^*)$$

where D is the n by u matrix with typical row

$$\tfrac{1}{2}(\Psi^{\#}-\Psi^*)' \nabla^2_{\Psi}\, g(x_t,\, \bar{\Psi})$$

and $\bar{\Psi}$ is on the line segment joining $\Psi^{\#}$ to Ψ^* .

Using Lemma 1 and Lemma 1 of [3], one can show that $(1/n)\, G'(\Psi^{\#})D$, $(1/n)G'D$, $(1/n)Z'D$, and $(1/n)\, D'D$ converge almost surely to the zero matrix. Again by Taylor's theorem,

$$[G'(\Psi^{\#}) - G'](e + \tau^* h) = E(\Psi^{\#}-\Psi^*)$$

where E is the u by u matrix with typical element

$$e_{ij} = \Sigma_{t=1}^{n} (\partial^2/\partial\Psi_j \partial\Psi_i) g(x_t,\bar{\Psi})\, [e_t + \tau^* h(x_t,\omega^*)]$$

Using Lemma 1 of [3] and the assumed convergence of τ^* to zero, one can show that $(1/n)\, E$ converges almost surely to the zero matrix.

We will obtain the probability order of $\psi^{\#}$ and $\delta^{\#}$. As mentioned earlier, for almost every realization of the errors $\{e_t\}$, $(\psi^{\#}, \delta^{\#})$ is eventually a stationary point of $(-\sqrt{n}/2)\ Q_n(\psi, \delta, \tau^*)$ so that the random vector

$$\begin{aligned}(-\sqrt{n}/2)\ &\nabla_{\psi\delta} Q_n(\psi^{\#}, \delta^{\#}, \tau^*)\\ &= (1/\sqrt{n})[G(\psi^{\#}) \vdots Z]'(y - g(\psi^{\#}) - Z\ \delta^{\#})\end{aligned}$$

converges almost surely to the zero vector. Substituting the expansions of the previous paragraph, we have that

$$\begin{aligned}(1/\sqrt{n})\ &[G \vdots Z]'(e + \tau^* h)\\ &- \{(1/n)\ [G(\psi^{\#}) \vdots Z]'[G + D \vdots Z] + (1/n) \begin{bmatrix} E & 0 \\ 0 & 0 \end{bmatrix}\} \sqrt{n} \begin{bmatrix} \psi^{\#} - \psi^* \\ \delta^{\#} \end{bmatrix}\end{aligned}$$

converges almost surely to the zero vector. Lemma 1 of [3] and our previous results imply that the matrix in braces converges almost surely to a non-singular matrix. Lemma 2 implies that $(1/\sqrt{n})[G \vdots Z]'(e + \tau^* h)$ converges in distribution to a $u + w$ - variate normal. These facts allow the conclusion that

$$u_n = \sqrt{n}\ \{\begin{bmatrix} \psi^{\#} - \psi^* \\ \delta^{\#} \end{bmatrix} - ([G \vdots Z]'[G \vdots Z])^{-1}[G \vdots Z]'(e + \tau^* h)\}$$

converges in probability to zero and that

$$v_n = \sqrt{n} \begin{bmatrix} \psi^{\#} - \psi^* \\ \delta^{\#} \end{bmatrix}$$

$$w_n = \sqrt{n}\ (\psi^{\#} - \psi^*)$$

are bounded in probability.

The sum of squares

$$\begin{aligned}\|y - g(\psi^{\#}) - Z\ \delta^{\#}\|^2 &= \|\ e + g + \tau^* h - g\ (\psi^{\#}) - Z\ \delta^{\#}\ \|^2\\ &= \|Q_{GZ}(e + \tau^* h) + P_{GZ}(e + \tau^* h) - G(\psi^{\#} - \psi^*) - Z\ \delta^{\#} - D(\psi^{\#} - \psi^*)\|^2\\ &= \|Q_{GZ}(e + \tau^* h)\|^2 - 2(e + \tau^* h)' Q_{GZ}\ D(\psi^{\#} - \psi^*)\\ &\quad + \|P_{GZ}(e + \tau^* h) - G(\psi^{\#} - \psi^*) - Z\ \delta^{\#} - D(\psi^{\#} - \psi^*)\|^2\ .\end{aligned}$$

The cross product term may be written as

$$(e + \tau^* h)\, Q_{GZ} D(\Psi^\# - \Psi^*)$$

$$= \tfrac{1}{2} w_n \{(1/n) \sum_{t=1}^{n} [e_t + \tau^* h(x_t, \omega^*)]\, \nabla^2 g(x_t, \bar{\Psi})\}\, w_n$$

$$- \{(1/\sqrt{n})(e + \tau^* h)'[G\vdots Z]\}\, \{(1/n)[G\vdots Z]'[G\vdots Z]\}^{-1} \{(1/n)[G\vdots Z]'D\}\, w_n$$

Both terms converge almost surely to zero by Lemmas 1 and 2, Lemma 1 of [3], and our previous results. (Some care must be taken with the argument concerning the first term; see [1, p. 14a-14b] for details.) By the triangle inequality

$$\| P_{GZ}(e + \tau^* h) - G(\Psi^\# - \Psi^*) - Z\,\delta^\# - D(\Psi^\# - \Psi^*) \|$$

$$\leq \| [G\vdots Z]\{([G\vdots Z]'[G\vdots Z])^{-1}[G\vdots Z]'(e + \tau^* h) - \begin{bmatrix} \Psi^\# - \Psi^* \\ \delta^\# \end{bmatrix}\} \|$$

$$+ \| D(\Psi^\# - \Psi^*) \|$$

$$= (u_n'\{(1/n)[G\vdots Z]'[G\vdots Z]\}u_n)^{\frac{1}{2}} + (w_n' \{(1/n)\, D'D\}\, w_n)^{\frac{1}{2}} .$$

The two terms on the right converge almost surely to zero by our previous results.

The proof for $\tilde{\sigma}^2$ is analogous and, therefore, omitted. $\square$

<u>Theorem 2</u>. The statistic $T^\#$ may be characterized as $T^\# = X + c_n$ where

$$X = (e + \tau^* h)' Q_G (e + \tau^* h)/(e + \tau^* h)'\, Q_{GZ}(e + \tau^* h)$$

and $n\, c_n$ converges in probability to zero.

The probability $P(X > c^*)$ is given by the doubly non-central F distribution as defined in [5; p.75] with: numerator degrees freedom w, and non-centrality parameter

$$\lambda_1 = (\tau^*)^2\, h'(P_{GZ} - P_G)h/(2\sigma^2)$$

and denominator degrees freedom n - u - w, and non-centrality parameter

$$\lambda_2 = (\tau^*)^2\, h'\, Q_{GZ} h/(2\sigma^2).$$

<u>Proof</u>: The proof of Lemma 1 of [2] may be used almost word for word to prove that

$$1/(\sigma^2)^\# = n/(e + \tau^* h)'\, Q_{GZ}(e + \tau^* h) + d_n$$

where $n\, d_n$ converges in probability to zero. Thus, $T^{\#} = X + c_n$ where

$$n\, c_n = n\, b_n[n/(e + \tau^{*}h)'\, Q_{GZ}(e + \tau^{\#}h)] + nd_n[(e + \tau^{*}h)' P_{GZ}(e + \tau^{*}h)/n] + nb_n d_n ;$$

the term b_n is as defined in Theorem 1. By Lemma 1 and Theorem 1, each term of nc_n converges in probability to zero.

Set $z = (1/\sigma)e$, $\gamma = (1/\sigma)\, \tau^{*}h$, and $R = P_{GZ} - P_G$.

The random variables $z_1, z_2, \ldots, z_n$ are independent normal random variables each with mean zero and variance one. Thus, the random variable $(z + \gamma)'R(z + \gamma)$ is a noncentral chi-squared with w degrees freedom and noncentrality parameter

$$\lambda_1 = \gamma' R\, \gamma/2 = (\tau^{*})^2 h'(P_{GZ} - P_G)h/(2\sigma^2)$$

Similarly, $(z + \gamma)'Q_{GZ}(z + \gamma)$ is a noncentral chi-squared random variable with $n - u - w$ degrees freedom and noncentrality parameter

$$\lambda_2 = \gamma' Q_{GZ}\gamma/2 = (\tau^{*})^2\, h'Q_{GZ}h/(2\sigma^2).$$

These two random variables are independent because $R\, Q_{GZ} = 0$ (see [Graybill 5, p.79ff]).

Now,

$$P(X \geqslant c^{*})$$

$$= P[(e + \tau^{*}h)'Q_G(e + \tau^{*}h)/(e + \tau^{*}h)'Q_{GZ}(e + \tau^{*}h) > 1 + wF_\alpha/(n-u-w)]$$

$$= P[(z + \gamma)'(Q_G - Q_{GZ})(z + \gamma)/(z + \gamma)'Q_{GZ}(z + \gamma) \geqslant w\, F_\alpha/(n-u-w)]$$

$$= P\{[(z + \gamma)'R(z + \gamma)/w]/[(z + \gamma)'Q_{GZ}(z + \gamma)/(n-u-w)] > F_\alpha\} \ . \ \square$$

Alternative expressions for the noncentrality parameters are useful in the context of an attempt to maximize λ_1 and minimize λ_2 via choice of Z.

$$\lambda_1 = (\tau^{*})^2 h'Q_G Z(Z'Q_G Z)^{-1}Z'Q_G\, h/(2\sigma^2)$$

$$\lambda_2 = (\tau^{*})^2 h'Q_G h/(2\sigma^2) - \lambda_1 \quad .$$

These expressions may be verified as follows. The vectors

$$b = (Z'Q_G Z)^{-1}Z'Q_G h$$

$$a = (G'G)^{-1}G'(h - Zb)$$

solve the equations

$$\begin{bmatrix} G'G & G'Z \\ Z'G & Z'Z \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} G'h \\ Z'h \end{bmatrix} .$$

as may be verified by substitution. Thus,

$$h'P_{GZ}h = [h'G \quad h'Z] \begin{bmatrix} G'G & G'Z \\ Z'G & Z'Z \end{bmatrix}^{-1} \begin{bmatrix} G'h \\ Z'h \end{bmatrix}$$

$$= [h'G \quad h'Z] \begin{bmatrix} (G'G)^{-1}G'(h - Zb) \\ (Z'Q_G Z)^{-1}Z'Q_G h \end{bmatrix}$$

$$= h'P_G(h - Zb) + h'Z(Z'Q_G Z)^{-1}Z'Q_G h$$

$$= h'P_G h - h'P_G Z(Z'Q_G Z)^{-1}Z'Q_G h + h'Z(Z'Q_G Z)^{-1}Z'Q_G h$$

$$= h'P_G h + h'Q_G Z(Z'Q_G Z)^{-1}Z'Q_G h .$$

The formulas for λ_1 and λ_2 are obtained by substitution of this expression for $h'P_{GZ}h$ in the formulas given in Theorem 2.

As mentioned in [4], one may use the statistic $S^{\#}$ instead of $T^{\#}$ to test H against A. For convenience, the definition of $S^{\#}$ is repeated here. Let

$$(s^2)^{\#} = \sum_{t=1}^{n} [y_t - g(x_t, \psi^{\#}) - z_t' \delta^{\#}]^2/(n-u-w) .$$

Evaluate the matrix $G(\psi)$ at $\psi = \psi^{\#}$ and put

$$C^{\#} = \{[G(\psi^{\#}) \vdots Z]'[G(\psi^{\#}) \vdots Z]\}^{-1} .$$

Let $C_{22}^{\#}$ be the matrix formed by deleting the first u rows and columns of C; then

$$S^{\#} = \frac{(\delta^{\#})'(C_{22}^{\#})^{-1} \delta^{\#}/w}{(s^2)^{\#}} .$$

H is rejected when $S^{\#}$ exceeds the upper $\alpha \cdot 100$ percentage point F_α of an F random variable with w numerator degrees freedom and n-u-w denominator degrees freedom.

The tests based on $S^{\#}$ and $T^{\#}$ are asymptotically equivalent in the sense of the following theorem.

<u>Theorem 3</u>. The statistic $S^{\#}$ may be characterized as $S^{\#} = Y + d_n$ where

$$Y = \frac{(e + \tau^* h)'(P_{GZ} - P_G)(e + \tau^* h)/w}{(e + \tau^* h)'Q_{GZ}(e + \tau^* h)/(n-u-w)}$$

and d_n converges in probability to zero. The probabilities $P(Y > F_\alpha)$ and $P(X > c^*)$ are equal where X is as in Theorem 2.

<u>Proof</u>. On page 6, as an intermediate step in the proof of Theorem 1, a characterization of $\sqrt{n}\begin{pmatrix}\Psi^{\#} - \Psi^* \\ \delta^{\#}\end{pmatrix}$ was obtained. Rearranging terms, we have

$$\delta^{\#} = (C_{21}G' + C_{22}Z')(e + \tau^* h) + z_n$$

$$= W_2 + z_n$$

where

$$C = ([G \vdots Z]'[G \vdots Z])^{-1}$$

has been partitioned as

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} .$$

and $\sqrt{n}\, z_n$ converges in probability to zero. Also, recall that $\sqrt{n}\,\delta^{\#}$ is bounded in probability as is $\sqrt{n}\, W_2$.

On page 7, as an intermediate step in the proof of Theorem 2, a characterization of $(\sigma^2)^{\#}$ was obtained. From this, it follows that

$$1/(s^2)^{\#} = (n-u-w)/(e + \tau^* h)'Q_{GZ}(e + \tau^* h) + w_n$$

where $n\, w_n$ converges in probability to zero.

The random variable $S^{\#}$ may be written

$$S^{\#} = (\delta^{\#})' C_{22}^{-1} \delta^{\#}/[w(s^2)^{\#}]$$

$$+ (\sqrt{n}\, \delta^{\#})'[(1/n)(C_{22}^{\#})^{-1} - (1/n)C_{22}^{-1}](\sqrt{n}\, \delta^{\#})/[w(s^2)^{\#}] .$$

The second term converges in probability to zero because: $\sqrt{n}\, \delta^{\#}$ is bounded in probability, $(s^2)^{\#}$ converges almost surely to σ^2, and $[(1/n)(C_{22}^{\#})^{-1} - (1/n)C_{22}^{-1}]$ converges almost surely to the zero matrix by Part 3 of Lemma 1 of [3].
Denote this second term by u_n . Now

$$(\delta^{\#})' C_{22}^{-1} \delta^{\#} = W_2' C_{22}^{-1} W_2 + 2(\sqrt{n}\, W_2)'[(1/n)C_{22}^{-1}](\sqrt{n}\, z_n)$$

$$+ (\sqrt{n}\, z_n)'[(1/n)\, C_{22}^{-1}](\sqrt{n}\, z_n)$$

where the latter two terms converge in probability to zero. Denote these latter terms by v_n .

We have, now, that

$$S^{\#} = W_2' C_{22}^{-1} W_2/[w(s^2)^{\#}] + u_n + v_n/[w(s^2)^{\#}]$$

$$= W_2' C_{22}^{-1} W_2/[w(e + \tau^* h)' Q_{GZ}(e + \tau^* h)/(n-u-w)]$$

$$+ (n\, w_n)\, W_2'[(1/n)\, C_{22}^{-1}]\, W_2/w + u_n + v_n/[w(s^2)^{\#}].$$

The remainder term d_n of the theorem to be proved equals the last three terms of this expression; and, converges in probability to zero. It remains to show that Y is the first term of this expression.

Now,

$$(G\ C_{12} + Z\ C_{22})\ C_{22}^{-1}\ (C_{21}G' + C_{22}Z')$$

$$= G\ C_{12}C_{22}^{-1}\ C_{21}G' + G\ C_{12}Z' + Z\ C_{21}G' + Z\ C_{22}Z'$$

$$= [G \vdots Z] \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} G' \\ Z' \end{bmatrix} - G'[C_{11} - C_{12}\ C_{22}^{-1}\ C_{21}]\ G$$

$$= P_{GZ} - P_G$$

using the fact that $(G'G)^{-1} = C_{11} - C_{12}\ C_{22}^{-1}\ C_{21}$. It follows that

$$W_2'\ C_{22}^{-1}\ W_2 = (e + \tau^* h)'(P_{GZ} - P_G)(e + \tau^* h)$$

as required. ▯

References

[1] Gallant, A. Ronald., "Inference for Nonlinear Models," Institute of Statistics Mimeograph Series, No. 875 (revised), Raleigh: North Carolina State University, 1975.

[2] ______________, "The Power of the Likelihood Ratio Test of Location in Nonlinear Models," Journal of the American Statistical Association, 70 (March 1975), 198-203.

[3] ______________, "Testing a Subset of the Parameters of a Nonlinear Regression Model," Journal of the American Statistical Association, 70 (December 1975), 927-932.

[4] ______________, "Testing a Nonlinear Regression Specification; A Nonregular Case," Institute of Statistics Mimeograph Series, No. 1017, Raleigh: North Carolina State University, 1975.

[5] Graybill, Franklin A. An Introduction to Linear Statistical Models, Vol. I, New York: McGraw-Hill Book Co., 1961.