THE INSTITUTE OF STATISTICS

UNIVERSITY OF NORTH CAROLINA SYSTEM

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

INFERENCE FOR NONLINEAR MODELS

by

A. RONALD GALLANT

Institute of Statistics
Mimeograph Series No. 875
Raleigh - June 1973
REVISED - June 1975

ABSTRACT

Inference for Nonlinear Models

The study considers estimation and hypothesis testing problems for a regression model whose response function is nonlinear in the unknown parameters.

The results of Malinvaud are extended to obtain asymptotic normality for the least squares estimates of the nonlinear parameters. The conditions set forth do not require the existence of second order partial derivatives of the response function in the parameters and do not require that the parameter space be bounded.

Hypothesis testing problems in the situation where the errors are normally and independently distributed are considered. For the hypothesis of location with variance known, a necessary and sufficient condition for the existence of a uniformly most powerful test is obtained and the asymptotic null and non-null distribution of the likelihood ratio test is derived. A special case investigated is a test of location when some parameters enter the model in a linear fashion.

INFERENCE FOR NONLINEAR MODELS [1]

A. R. GALLANT

1. Introduction. The regression model considered has the structure

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

The unknown parameter θ° is contained in the parameteter space Ω a subset of R^p. The input variables x_t are contained in X a subset of R^k; the rule of formation of the sequence $\{x_t\}_{t=1}^{\infty}$ is assumed known. The sequence of random variables $\{e_t\}_{t=1}^{\infty}$ are assumed independent each with distribution function $F(e)$. This determines the (marginal) probability measures P_n^e on $(R^n, \mathcal{B}_n)$ where $\mathcal{B}_n$ denotes the n dimensional Borel sets and also determines the measure P_∞^e over $(R^\infty, \mathcal{B}_\infty)$. A sample $(y_1, y_2, \ldots, y_n)$ is generated according to the model in order to estimate θ and/or test hypotheses as to the location of θ.

In Section 3 we extend the results of Malinvaud (1970) to obtain measurability, strong consistency, and aymptotic normality for the least squares estimator. In addition technical results on characterization and rates of convergence are obtained for use in later sections. The assumptions employed allow Ω to be an unbounded set and do not require that the second order partial derivatives of f in θ exist. These relaxations of the usual assumptions were

[1] This research was supported in part by the Bureau of the Census Contract No. 1-35044 and by a N.D.E.A. Title IV Fellowship.

motivated by a desire to accommodate Problem B below and to develop an estimation theory sufficiently general to include the grafted polynomial response functions which have seen use in applications. A grafted quadratic-quadratic model is used as the example in Section 9. It should be noted that our assumptions do not rule out the case when the response function is linear in the parameters.

The remainder of the paper considers these two hypothesis testing problems:

Problem A. The form of the response function $f(x,\theta)$ is known as well as the sequence $\{x_t\}$. For given $\alpha \in (0,1)$ and given n we wish to test $H: \theta^\circ \in \Omega_H$ against $K: \theta^\circ \in \Omega_K$ when $\Omega_H = \{\theta_0\}$ and $\Omega_K = \Omega \sim \Omega_H$.

Problem B. The form of the response function is known as well as the sequence $\{x_t\}$. Some of the parameters enter the model in a linear fashion so that the parameter space has the form

$$\Omega = \{\theta = (\theta_{(1)}, \theta_{(2)}): \theta_{(1)} \in R^{p_1}, \theta_{(2)} \in \Omega_{(2)} \subset R^{p_2}\}$$

and the response function has the form

$$f(x,\theta) = a_0(x,\theta_{(2)}) + \Sigma_{i=1}^{p_1} \theta_i a_i(x,\theta_{(2)}) .$$

For given $\alpha \in (0,1)$ and given n we wish to test $H: \theta^\circ \in \Omega_H$ against $K: \theta^\circ \in \Omega_K$ when

$$\Omega_H = \{\theta: \theta_{(1)} \in R^{p_1}; \theta_{(2)} = {}_0\theta_{(2)}\}$$

and $\Omega_K = \Omega \sim \Omega_H$.

The theoretical setting we adopt to approach Problems A and B follows Lehmann (1959). Briefly, the probability $P^y_\theta(B)$ of obtaining a sample falling in the set B from $\mathcal{B}_n$ is given by $P^e_n(B_\theta)$ where

$$B_\theta = \{(e_1, e_2, \ldots, e_n): (e_1 + f(x_1,\theta), \ldots, e_n + f(x_n,\theta)) \in B\} .$$

A test of H: $\theta^\circ \in \Omega_H$ against K: $\theta^\circ \in \Omega_K$ is a $(\mathcal{B}_n)$ measurable function $\varphi(y)$ mapping the sample $(y_1, \ldots, y_n)$ into $[0,1]$. The power function of φ is given by $E_\theta\varphi = \int \varphi(y)dP^y_\theta(y)$. The usage of the terms level α, uniformly most powerful (UMP), and most powerful (MP) are standard (Lehmann, 1959, p. 60 ff).

2. Notation, definitions, assumptions.

Definition 2.1. For the regression model described in Section 1 and given n define

W' = the transpose of a matrix W.

W^+ = the Moore-Penrose inverse of a matrix W (Rao, 1965, p. 25).

$\| \|$ = the Euclidean norm.

e = $(e_1, e_2, \ldots, e_n)'$.

y = $(y_1, y_2, \ldots, y_n)'$.

$f_n(\theta)$ = $(f(x_1,\theta), f(x_2,\theta), \ldots, f(x_n,\theta))'$.

$\nabla f(x,\theta)$ = the $p \times 1$ vector whose j^{th} element is $\frac{\partial}{\partial\theta_j} f(x,\theta)$.

$\nabla^2 f(x,\theta)$ = the $p \times p$ matrix whose i,j^{th} element is $\frac{\partial^2}{\partial\theta_i\partial\theta_j} f(x,\theta)$.

$F_n(\theta)$ = the $n \times p$ matrix whose t^{th} row is $\nabla' f(x_t, \theta)$.

$Q_n(\theta)$ = $\|y - f_n(\theta)\|^2$.

$P_n(\theta)$ = $F_n(\theta)[F_n'(\theta)F_n(\theta)]^+ F_n'(\theta) = F_n(\theta)F_n^+(\theta)$.

$P_n^\perp(\theta)$ = $I - P_n(\theta)$.

$I_A(x)$ = 1 if $x \in A$; = 0 if $x \notin A$.

$\mathcal{Q}$ = the Borel subsets of X.

$\hat{\theta}_n(y)$ = a $(\mathcal{B}_n)$ measurable function such that $Q_n(\hat{\theta}_n) = \inf_\Omega Q_n(\theta)$.

$\hat{\sigma}_n^2(y)$ = $n^{-1} Q_n(\hat{\theta}_n)$.

When considering Problem B, the following additional notation is convenient:

Definition 2.2. Given the conditions of Problem B, define:

$a(\theta_{(2)})$ = the $n \times 1$ vector whose t^{th} element is $a_0(x_t, \theta_{(2)})$.

$A(\theta_{(2)})$ = the $n \times p_1$ matrix whose t,j^{th} element is $a_j(x_t, \theta_{(2)})$.

a_H = $a({}_0\theta_{(2)})$.

A_H = $A({}_0\theta_{(2)})$.

$\gamma(\theta)$ = $([A_H^+(f_n(\theta) - a_H)]', {}_0\theta_{(2)})$.

$P_H = A_H[A_H'A_H]^+A_H' = A_H A_H^+$.

$P_H^{\perp} \quad = I - P_H$.

$\hat{\hat{\theta}}_n(y) \quad =$ a $(\mathcal{B}_n)$ measurable function such that $Q_n(\hat{\hat{\theta}}_n) = \inf_{\Omega_H} Q_n(\theta)$.

Definition 2.3. (Malinvaud, 1970) Given the sequence $\{x_t\}_{t=1}^{\infty}$ from X and an n define μ_n to be the measure on $(X,\mathcal{Q})$ which satisfies

$$\mu_n(A) = n^{-1} \Sigma_{t=1}^n I_A(x_t)$$

for each $A \in \mathcal{Q}$.

Definition 2.4. (Billingsly and Topsoe, 1967) A sequence of measures $\{\nu_n\}$ on $(X,\mathcal{Q})$ is said to converge weakly to a measure ν on $(X,\mathcal{Q})$ if for every real valued, bounded, continuous function g over X

$$\int g(x)\, d\nu_n(x) \to \int g(x)\, d\nu(x)$$

as $n \to \infty$.

The following definition is an extension to a sequence of vector valued random variables of the definition given by Pratt (1958).

Definition 2.5. Let $\{Z_n\}$ be a sequence of vector valued random variables, and $\{a_n\}$ a sequence of (strictly) positive real numbers.

We say that Z_n is order in probability a_n and write $Z_n = O_p(a_n)$ if for every $\epsilon > 0$ there an M and an N such that $P[a_n^{-1}\|Z_n\| \geq M] < \epsilon$ for $n > N$.

We say that Z_n is of smaller order in probability than a_n and write $Z_n = o_p(a_n)$ if $a_n^{-1}Z_n \xrightarrow{P} 0$ as $n \to \infty$.

For each assumption below it is implicitly assumed that any lower numbered assumption necessary for existence of terms is satisfied. For example, the statement of Assumption 7 requires Assumption 6 for definition of μ and Assumption 2 for the $\mathcal{G}$ measurability of $\{x: f(x,\theta) \neq f(x,\theta^\circ)\}$.

Assumption 1. Ω is a closed subset of R^p.

Assumption 2. $f(x,\theta)$ is continuous on $X \times \Omega$.

Assumption 3. For given n and almost every y $(P^y_{\theta^\circ})$ there is a θ^* in Ω minimizing $\|y - f_n(\theta)\|^2$.

Assumption 4. The errors $\{e_t\}$ are independent and identically distributed with mean zero and finite variance $\sigma^2 > 0$.

Assumption 5. X is a compact subset of R^k.

Assumption 6. The sequence of measures $\{\mu_n\}$ determined by $\{x_t\}$ converges weakly to a measure μ on $(X,\mathcal{G})$.

Assumption 7. If $\theta \neq \theta^\circ$ and $\theta \in \Omega$ then $\mu\{x: f(x,\theta) \neq f(x,\theta^\circ)\} > 0$.

Assumption 8. Given $M > 0$ there is an N and a K such that for all $n > N$ and all $\theta \in \Omega$ if $n^{-1} \Sigma_{t=1}^n f^2(x_t,\theta) < M$ then $\|\theta\| < K$.

Assumption 9. There is a bounded open sphere Ω° containing θ° whose closure $\bar{\Omega}^\circ$ is a subset of Ω.

Assumption 10. $\nabla f(x,\theta)$ exists and is continuous on $X \times \Omega$.

Assumption 11. The matrix

$$F'F = [\int \frac{\partial}{\partial\theta_i} f(x,\theta^\circ) \frac{\partial}{\partial\theta_j} f(x,\theta^\circ) d\mu(x)]_{p \times p}$$

is non-singular.

Assumption 12. There is a function $\varphi(x,\theta)$ which is uniformly bounded for $(x,\theta) \in X \times \bar{\Omega}^{\circ}$ such that

$$f(x,\theta) = f(x,\theta^{\circ}) + \nabla' f(x,\theta^{\circ})(\theta - \theta^{\circ})' + \varphi(x,\theta)\ \|\theta - \theta^{\circ}\|^2 .$$

Assumption 13. The response function f, inputs $\{x_t\}$ and errors $\{e_t\}$ are such that given a sequence of random variables $\{\tilde{\theta}_n\}$ with $\tilde{\theta}_n \xrightarrow{a.s.} \theta^{\circ}$ and $P_n^e[Q_n(\tilde{\theta}_n) = \inf_{\Omega} Q_n(\theta)] \to 1$ as $n \to \infty$ it follows that

$$n^{-\frac{1}{2}}\ \Sigma_{t=1}^{n}\{\frac{\partial}{\partial\theta_i} f(x_t,\tilde{\theta}_n) - \frac{\partial}{\partial\theta_i} f(x_t,\theta^{\circ})\}\ e_t \xrightarrow{P} 0$$

as $n \to \infty$ for $i = 1, 2, \ldots, p$.

As shown later the verification of Assumption 8 is unnecessary when those which preceed it are satisfied and

Assumption 14. Ω is bounded.

Also, the verification of Assumptions 12, 13 is unnecessary when those which preceed them and

Assumption 15. The partial derivatives $\frac{\partial^2}{\partial\theta_i\partial\theta_j} f(x,\theta)$ exist and are continuous over $X \times \Omega$.

3. Large sample estimation. As stated earlier, our objective is to obtain a large sample theory sufficiently general to accommodate the linear model, the partially linear model of Problem B, and the class of nonlinear models typified by the example.

Theorem 1. If a regression model satisfies Assumptions 1, 2, 3 for given n then there is a Borel measurable function $\hat{\theta}_n$ mapping R^n into

Ω such that $Q_n(\hat{\theta}_n) = \inf_\Omega Q_n(\theta)$ for all y in a set with $(P^y_{\theta^\circ})$ probability one.

Proof. Construction. Let $\Omega_i = \{\theta \in \Omega : ||\theta|| \leq i\}$ which is compact by Assumption 1. Apply Lemma 2 of Jennrich (1969) with $\Theta = \Omega_i$, $Y = R^n$ and $Q(\theta,y) = Q_n(\theta)$ obtaining the measurable function $\tilde{\theta}_i : R^n \to \Omega_i$ which satisfies $Q_n(\tilde{\theta}_i) = \inf_{\Omega_i}(\theta)$. Since $\inf_\Omega Q_n(\theta)$ is measurable by Assumption 2 the sets

$$Y_i^* = \{y \in R^n : Q_n(\tilde{\theta}_i) = \inf_\Omega Q_n(\theta)\}$$

are in $\mathcal{B}_n$. Let $Y_i = Y_i^* \sim \cup_{j=1}^{i-1} Y_j^*$; note that the Y_i are disjoint. Let $Y = \cup_{i=1}^\infty Y_i$ and set

$$\hat{\theta}_n(y) = \theta^\circ I_{\tilde{Y}}(y) + \lim_{M \to \infty} \Sigma_{i=1}^M \tilde{\theta}_i(y) I_{Y_i}(y) .$$

Verification of properties. By Assumption 3 $P^y_{\theta^\circ}(\cup_{i=1}^\infty Y_i) = 1$ and $\hat{\theta}_n$ is the limit of measurable functions hence measurable. For given $y \in \cup_{i=1}^\infty Y_i$ y is in $Y_i \subset Y_i^*$ for some i hence $Q_n(\hat{\theta}_n) = Q_n(\tilde{\theta}_i) = \inf_\Omega Q_n(\theta)$. ▯

Lemma 3.1. Let a regression model satisfy Assumption 1 through 8 for all $n > N$. Then there is a bounded open subset S of R^p containing θ° such that for almost every realization (P^e_∞) a sequence of least squares estimators is in S for n sufficiently large.

Proof.

$$n^{-\frac{1}{2}} ||f_n(\hat{\theta}_n)|| \leq n^{-\frac{1}{2}} ||y - f_n(\hat{\theta}_n)|| + n^{-\frac{1}{2}} ||y||$$

$$\leq n^{-\frac{1}{2}} \|y - f_n(\theta^\circ)\| + n^{-\frac{1}{2}} \|e\| + n^{-\frac{1}{2}} \|f_n(\theta^\circ)\|$$

$$= 2n^{-\frac{1}{2}} \|e\| + n^{-\frac{1}{2}} \|f_n(\theta^\circ)\| .$$

Since $n^{-1} \|f_n(\theta^\circ)\|^2 \leq \sup_X f^2(x,\theta^\circ)$ and $n^{-1}\|e\|^2 \xrightarrow{a.s.} \sigma^2$ by the Strong Law of Large numbers, there is an M for almost every realization of $\{e_t\}$ such that $n^{-\frac{1}{2}} \|f_n(\hat{\theta}_n)\| < M$. By Assumption 8 $\|\hat{\theta}_n\| < K$ for $n > N$. Choose $B > K$ large enough that $\theta^\circ \in S = \{\theta: \|\theta\| < B\}$. ‖

Lemma 3.2. Let $\Theta \subset R^p$ and $X \subset R^k$ be compact sets. Let the real valued function $g(x,\theta)$ be continuous on $X \times \Theta$. If a sequence of measures $\{\nu_n\}$ converges weakly to a measure ν on $(X,\mathcal{G})$ then

$$\int g(x,\theta)\, d\nu_n(x) \to \int g(x,\theta)\, d\nu(x)$$

uniformly in θ over Θ as $n \to \infty$.

Proof. Malinvaud (1970, p. 967). ‖

Theorem 2. Let a regression model satisfy Assumptions 1 through 8 for all $n > N$ and let $\hat{\theta}_n$ be a least squares estimator. Then $\hat{\theta}_n \xrightarrow{a.s.} \theta^\circ$ and $\hat{\sigma}_n^2 \xrightarrow{a.s.} \sigma^2$ as $n \to \infty$.

Proof. Let S be as in Lemma 3.1. If we take $\Theta = \Omega \cap \bar{S}$ as the parameter space of the regression model it will satisfy Assumption (a) of Jennrich (1969). We next show the model satisfies Assumption (b) of Jennrich when Θ is the parameter space.

$$n^{-1} \Sigma_{t=1}^n f^2(x_t,\theta) \to \int f^2(x,\theta)\, d\mu(x)$$

uniformly on Θ by Assumptions 1, 2, 5, 6 and Lemma 3.2; if $\theta \in \Theta$ and $\theta \neq \theta^\circ$ then

$$\int [f(x,\theta) - f(x,\theta^\circ)]^2 \, d\mu(x) > 0$$

by Assumption 7.

Let θ_n^* be a least squares estimator for the model with parameter space Θ and set

$$\tilde{\theta}_n = \hat{\theta}_n + (\theta_n^* - \hat{\theta}_n)(1 - I_S(\hat{\theta}_n)) .$$

Now $\tilde{\theta}_n$ is a least squares estimator for the model with parameter space Θ and by Theorem 6 of Jennrich $\tilde{\theta}_n \xrightarrow{a.s.} \theta^\circ$ and $n^{-1}Q_n(\tilde{\theta}_n) \xrightarrow{a.s.} \sigma^2$. By Lemma 3.1, except for realizations in E with $P_\infty^e(E) = 0$, we have $I_S(\hat{\theta}_n) = 1$, $\tilde{\theta}_n = \hat{\theta}_n$, and $n^{-1}Q_n(\tilde{\theta}_n) = \hat{\sigma}_n^2$ for n sufficiently large. ▯

Lemma 3.3. Let a regression model satisfy Assumptions 5, 6, 9, 10, 11. Let $\{\tilde{\alpha}_n\}$ and $\{\tilde{\beta}_n\}$ be sequences of random variables with range in Ω such that α_n, $\beta_n \xrightarrow{a.s.} \theta^\circ$. Then

$$\frac{1}{n} F_n'(\tilde{\alpha}_n) \, F_n(\tilde{\beta}_n) \xrightarrow{a.s.} F'F .$$

Proof. We may write the elements of $n^{-1}F_n'(\alpha)F_n(\beta)$ where $(\alpha, \beta) \in \Omega \times \Omega$ as $f_{nij}(\alpha,\beta) = \int \frac{\partial}{\partial\theta_i} f(x,\alpha) \frac{\partial}{\partial\theta_j} f(x,\beta) \, d\mu_n(x)$. By Lemma 3.2 $f_{nij}(\alpha,\beta) \to f_{ij}(\alpha,\beta) = \int \frac{\partial}{\partial\theta_i} f(x,\alpha) \frac{\partial}{\partial\theta_j} f(x,\beta) \, d\mu(x)$ as $n \to \infty$ uniformly on $\bar{\Omega}^\circ \times \bar{\Omega}^\circ$; further, the uniform limit $f_{ij}(\alpha,\beta)$ of continuous functions is continuous on $\bar{\Omega}^\circ \times \bar{\Omega}^\circ$. As a consequence $f_{nij}(\tilde{\alpha}_n, \tilde{\beta}_n) \xrightarrow{a.s.} f_{ij}(\theta^\circ, \theta^\circ)$. ▯

Lemma 3.4. Let $\{e_t\}$, X , $\{x_t\}$ satisfy Assumptions 4, 5, 6 respectively and let $\Theta \subset R^p$ be compact. If $g(x,\theta)$ is continuous on $X \times \Theta$ then for almost all realizations of $\{e_t\}$

$$\frac{1}{n} \Sigma_{t=1}^{n} g(x_t,\theta)\ e_t \to 0$$

uniformly for all $\theta \in \Theta$.

Proof. By Lemma 3.2, $n^{-1} \Sigma_{t=1}^{n} g^2(x_t,\theta) \to \int g^2(x,\theta)\ d\mu(x)$ uniformly over θ . Apply Theorem 4 of Jennrich (1969). ‖

Lemma 3.5. Let $\{e_t\}$, X , $\{x_t\}$ satisfy Assumptions 4, 5, 6 respectively. Let

$$g(x) = (g_1(x), \ldots, g_p(x))'$$

where each g_i is continuous over X . Then

$$n^{-\frac{1}{2}} \Sigma_{t=1}^{n} e_t g(x_t) \xrightarrow{\mathcal{L}} N_p\{0,\sigma^2 V\}$$

where $V = [\int g_i(x)\ g_j(x)\ d\mu(x)]_{p \times p}$ and $N_p\{\mu,V\}$ denotes the p dimensional multivariate normal distribution with mean μ and dispersion matrix V .

Proof. Apply Corollary 1 of Jennrich (1969). The existence of the requisite tail products follows from Assumption 6. ‖

Conclusions (a), (c) of the theorem which follows characterize $\hat{\theta}_n$, $\hat{\sigma}^2_n$, as linear and quadratic functions of the errors $\{e_t\}$ similar to those occurring in linear regression models plus remainders of specified probability order. Conclusion (b) specifies the rate at which $\hat{\theta}_n$ converges to θ° .

Theorem 3. Let a regression model satisfy Assumptions 1 through 13 for all $n > N$ and let $\hat{\theta}_n$ be a least squares estimator. Then there is an M such that for all $n > M$ $\det[F_n'(\theta^\circ)F_n(\theta^\circ)] > 0$ and:

(a) $(\hat{\theta}_n - \theta^\circ)' = [F_n'(\theta^\circ)F_n(\theta^\circ)]^{-1}F_n'(\theta^\circ)e + o_p(\frac{1}{\sqrt{n}})$

(b) $(\hat{\theta}_n - \theta^\circ)' = O_p(\frac{1}{\sqrt{n}})$

(c) $\hat{\sigma}_n^2 = n^{-1}\|P_n^\perp(\theta^\circ)e\|^2 + O_p(\frac{1}{n})$.

If, in addition, Assumption 15 is satisfied then:

(d) $\hat{\sigma}_n^2 = n^{-1}\|P_n^\perp(\theta^\circ)e\|^2 + o_p(\frac{1}{n})$.

Proof. Assumptions 6, 10, 11 imply $\det[n^{-1}F_n'(\theta^\circ)F_n(\theta^\circ)] \to \det(F'F) > 0$ as $n \to \infty$ hence the existence of M. All statements below are for $n > \max[M,N]$.

Conclusion (a). Let $\varphi_n(\theta) = (\varphi(x_1,\theta), \ldots, \varphi(x_n,\theta))'$ where $\varphi(x,\theta)$ is given by Assumption 12 and let $\tilde{\theta}_n = \hat{\theta}_n I_{\Omega^\circ}(\hat{\theta}_n) + \theta^\circ(1 - I_{\Omega^\circ}(\hat{\theta}_n))$. Note that $\varphi_n(\tilde{\theta}_n)$ is measurable by Assumptions 2, 10 and the measurability of $\hat{\theta}_n$. If $\hat{\theta}_n \in \Omega^\circ$ then $\sqrt{n}\,\hat{\theta}_n = \sqrt{n}\,\tilde{\theta}_n$ and by Assumption 10 $\sqrt{n}\,\nabla Q_n(\hat{\theta}_n) = 0$ hence by Theorem 2 $\sqrt{n}\,\hat{\theta}_n = \sqrt{n}\,\tilde{\theta}_n + o_p(1)$ and $\sqrt{n}\,\nabla Q_n(\tilde{\theta}_n) = o_p(1)$. Using the fact that $\nabla Q_n(\tilde{\theta}_n) = 2F_n'(\tilde{\theta}_n)\{y - f_n(\tilde{\theta}_n)\}$ and substituting the expression for $f_n(\tilde{\theta}_n)$ given by Assumption 12 we have

$$o_p(1) = n^{-\frac{1}{2}}F_n'(\tilde{\theta}_n)e - G_n\sqrt{n}\,(\tilde{\theta}_n - \theta^\circ)' ,$$

where

$$G_n = n^{-1}\{F_n'(\tilde{\theta}_n)F_n(\theta^\circ) + F_n'(\tilde{\theta}_n)\varphi_n(\tilde{\theta}_n)(\tilde{\theta}_n - \theta^\circ)\} .$$

By Assumption 13

$$o_p(1) = n^{-\frac{1}{2}} F_n'(\theta^\circ)e - G_n \sqrt{n}\, (\tilde{\theta}_n - \theta^\circ)' .$$

Now $G_n \xrightarrow{a.s.} F'F$ as we will show below and $n^{-\frac{1}{2}}F_n'(\theta^\circ)e \xrightarrow{\mathcal{L}} N_p$ by Lemma 3.5 so that $\sqrt{n}\,(\tilde{\theta}_n - \theta^\circ)' \xrightarrow{\mathcal{L}} N_p$. Thus

$$(I - [\tfrac{1}{n} F_n'(\theta^\circ)F_n(\theta^\circ)]^{-1}G_n) \sqrt{n}\, (\tilde{\theta}_n - \theta^\circ) \xrightarrow{P} 0$$

hence

$$\sqrt{n}\, (\hat{\theta}_n - \theta^\circ)' = \sqrt{n}\, (\tilde{\theta}_n - \theta^\circ)' + o_p(1) = \sqrt{n}\, [F_n'(\theta^\circ)F_n(\theta^\circ)]^{-1}F_n'(\theta^\circ)e + o_p(1) .$$

To see that $G_n \xrightarrow{a.s.} F'F$ note that $n^{-1}F_n'(\tilde{\theta}_n)F_n(\theta^\circ) \xrightarrow{a.s.} F'F$ by Lemma 3.3 and the elements of $n^{-1}F_n'(\tilde{\theta}_n)\varphi_n(\tilde{\theta}_n)$ are uniformly bounded by Assumption 12 hence

$$n^{-1}F_n'(\tilde{\theta}_n)\varphi_n(\tilde{\theta}_n)(\tilde{\theta}_n - \theta^\circ) \xrightarrow{a.s.} 0 \quad \text{by Theorem 2.}$$

Conclusion (b). By Chebishev's inequality the elements of $\sqrt{n}\, [F_n'(\theta^\circ)F_n(\theta^\circ)]^{-1}F_n'(\theta^\circ)e$ are $O_p(1)$.

Conclusion (c). Omitting the subscript n we have

$$\begin{aligned} n\hat{\sigma}_n^2 &= \|y - f(\tilde{\theta})\|^2 + (\|y - f(\hat{\theta})\|^2 - \|y - f(\tilde{\theta})\|^2)(1 - I_{\Omega^\circ}(\hat{\theta})) \\ &= \|y - f(\theta^\circ) - F(\theta^\circ)(\tilde{\theta} - \theta^\circ)' - \|\tilde{\theta} - \theta^\circ\|^2\varphi(\tilde{\theta})\|^2 + o_p(1) \\ &= \|P^\perp(\theta^\circ)e + P(\theta^\circ)e - F(\theta^\circ)(\tilde{\theta} - \theta^\circ) - \|\tilde{\theta} - \theta^\circ\|^2\varphi(\tilde{\theta})\|^2 + o_p(1) \end{aligned}$$

using conclusion (a) and letting j represent a random $p \times 1$ vector of probability order $O_p(1)$ we have

$$= \|P^{\perp}(\theta^{\circ})e - o_p(n^{-\frac{1}{2}})F(\theta^{\circ})j - \|\tilde{\theta}-\theta^{\circ}\|^2\ \varphi(\tilde{\theta})\|^2 + o_p(1)$$

$$= \|P^{\perp}(\theta^{\circ})e\|^2 + \|o_p(n^{-\frac{1}{2}})F(\theta^{\circ})j + \|\tilde{\theta} - \theta^{\circ}\|^2\varphi(\tilde{\theta})\|^2$$

$$- 2e'P^{\perp}(\theta^{\circ})[F(\theta^{\circ})jo_p(n^{-\frac{1}{2}}) + \|\tilde{\theta} - \theta^{\circ}\|^2\varphi(\tilde{\theta})] + o_p(1)$$

$$= \|P^{\perp}(\theta^{\circ})e\|^2 + \|o_p(n^{-\frac{1}{2}})F(\theta^{\circ})j + O_p(n^{-1})\varphi(\tilde{\theta})\|^2$$

$$- 2\|\tilde{\theta} - \theta^{\circ}\|^2 e'P^{\perp}(\theta^{\circ})\varphi(\tilde{\theta}) + o_p(1) .$$

Now the second term of this expression is bounded by the square of the sum

$$[o_p(n^{-1})j'F'(\theta^{\circ})F(\theta^{\circ})j]^{\frac{1}{2}} + [O_p(n^{-2})\varphi'(\tilde{\theta})\varphi(\tilde{\theta})]^{\frac{1}{2}}$$

which converges in probability to zero since $(1/n)F'(\theta^{\circ})F(\theta^{\circ}) \to F'F$ by Lemma 3.3 and $(1/n)\varphi'(\tilde{\theta})\varphi(\tilde{\theta})$ is uniformly bounded for all n. The absolute value of the term

$$2\|\tilde{\theta} - \theta^{\circ}\|^2 e'P^{\perp}(\theta^{\circ})\varphi(\tilde{\theta})$$

is bounded by

$$2O_p(1)\|n^{-\frac{1}{2}}e'P^{\perp}(\theta^{\circ})\|\|n^{-\frac{1}{2}}\varphi(\tilde{\theta})\|$$

$$\le 2O_p(1)[e'e/n]^{\frac{1}{2}}[\varphi'(\tilde{\theta})\varphi(\tilde{\theta})/n]^{\frac{1}{2}}$$

and the latter term is $O_p(1)$ by the Strong Law of Large Numbers and the uniform bound on $\varphi(x,\theta)$ given by Assumption 12. Thus, we have shown that

$$n\hat{\sigma}_n^2 = \|P^{\perp}(\theta^{\circ})e\|^2 + o_p(1) + O_p(1) + o_p(1)$$

which establishes conclusion (c).

Conclusion (d). What is required is to use Assumption 15 to show that the term $2\|\tilde{\theta} - \theta^\circ\|^2 e' P^\perp(\theta^\circ)\varphi(\tilde{\theta})$ obtained above is $o_p(1)$ rather than $O_p(1)$. By Taylor's theorem, there is a function $\bar{\theta}$: $X \times \bar{\Omega}^\circ \to \bar{\Omega}^\circ$ such that

$$\|\theta - \theta^\circ\|^2 \varphi(x,\theta) = \tfrac{1}{2}(\theta - \theta^\circ)\nabla^2 f(x,\bar{\theta})(\theta - \theta^\circ)'$$

and $\|\bar{\theta}(x,\theta) - \theta^\circ\| \le \|\theta - \theta^\circ\|$ for all $(x,\theta) \in X \times \bar{\Omega}^\circ$. Let $u = P^\perp(\theta^\circ)e$; then

$$2\|\tilde{\theta} - \theta^\circ\|^2 e' P^\perp(\theta^\circ)\varphi(\tilde{\theta}) = (\tilde{\theta} - \theta^\circ)\Sigma_{t=1}^n u_t \nabla^2 f(x_1,\bar{\theta})(\tilde{\theta} - \theta^\circ)' .$$

If we show that $(1/n)\,\Sigma_{t=1}^n u_t d(x_t,\bar{\theta})$ converges in probability to zero where $d(x,\theta)$ denotes a typical element of $\nabla^2 f(x,\theta)$ the desired result will follow from conclusion (b). The sum

$$\begin{aligned} &(1/n)\,\Sigma_{t=1}^n u_t d(x_t,\bar{\theta}) \\ &= (1/n)\Sigma_{t=1}^n u_t[d(x_t,\bar{\theta}) - d(x_t,\theta^\circ)] + (1/n)\Sigma_{t=1}^n u_t d(x_t,\theta^\circ) \\ &= Z_2 + Z_1 . \end{aligned}$$

Now $E(Z_1) = 0$ and

$$\begin{aligned} \mathrm{Var}(Z_1) &= (1/n^2) d'(\theta^\circ) C(u,u') d(\theta^\circ) \\ &= (1/n^2)\sigma^2 d'(\theta^\circ) P^\perp(\theta^\circ) d(\theta^\circ) \\ &\le (1/n^2)\sigma^2 d'(\theta^\circ) d(\theta^\circ) \\ &= (1/n)\sigma^2 \int d^2(x,\theta^\circ)\, d\mu_n(x) \\ &\to 0 \end{aligned}$$

as $n \to \infty$. Thus, by Chebishev's inequality $Z_1 = o_p(1)$. Since $X \times \bar{\Omega}^\circ$ is compact $d(x,\theta)$ is uniformly continuous over $X \times \bar{\Omega}^\circ$. Given $\epsilon > 0$ there is a δ such that $\|\bar{\theta} - \theta^\circ\| < \delta$ and $(x,\bar{\theta}) \in X \times \bar{\Omega}^\circ$ imply $|d(x,\bar{\theta}) - d(x,\theta^\circ)| < (\sigma + \epsilon)^{-1}\epsilon$. Then for almost every realization of $\{e_t\}$ there is an N such that

$$\|n^{-\frac{1}{2}}[d(\bar{\theta}) - d(\theta^\circ)]\| < (\sigma + \epsilon)^{-1}\epsilon$$

$$\|n^{-\frac{1}{2}}P^{\perp}(\theta^\circ)e\| \le \|n^{-\frac{1}{2}}e\| < \sigma + \epsilon$$

so that

$$|Z_2| \le \|n^{-\frac{1}{2}}[d(\bar{\theta}) - d(\theta^\circ)]\|\|n^{-\frac{1}{2}}P^{\perp}(\theta^\circ)e\| < \epsilon .$$

Thus $Z_2 \xrightarrow{a.s.} 0$ which implies $Z_2 = o_p(1)$. □

Theorem 4. Let a regression model satisfy Assumptions 1 through 13 for all $n > N$ and let $\hat{\theta}_n$ be a least squares estimator. Then

$$\sqrt{n}\,(\hat{\theta}_n - \theta^\circ)' \xrightarrow{\mathcal{L}} N_p\{0,\sigma^2(F'F)^{-1}\}$$

$$n^{-1}F_n'(\hat{\theta}_n)F_n(\hat{\theta}_n) \xrightarrow{a.s.} F'F .$$

Proof. The first conclusion follows from (a) of Theorem 3, Lemma 3.3 with $\tilde{\alpha}_n = \tilde{\beta}_n = \theta^\circ$, and Lemma 3.5. The second conclusion follows from Lemma 3.3 with $\tilde{\alpha}_n = \tilde{\beta}_n = \hat{\theta}_n$. □

Theorem 5. If a regression model satisfies Assumptions 1, 2, and 14 it satisfies Assumptions 3 and 8 for all n .

If a regression model satisfies Assumptions 1 through 11 and 15 for all $n > N$ it satisfies Assumptions 12 and 13.

<u>Proof</u>. First statement. Assumptions 1, 2, 14 imply Assumption 3. Set $K = \sup_\Omega \|\theta\| + 1$ and Assumption 8 follows.

Second statement. By Assumption 15 and Taylor's theorem there is a function $\bar{\theta}: X \times \bar{\Omega}^\circ \longrightarrow \bar{\Omega}^\circ$ such that Assumption 12 holds with

$$\varphi(x,\theta) = \begin{cases} \frac{1}{2}\|\theta - \theta^\circ\|^{-2}(\theta - \theta^\circ)\nabla^2 f(x,\bar{\theta})(\theta - \theta^\circ)' & \theta \neq \theta^\circ \\ 0 & \theta = \theta^\circ \end{cases}$$

Let $\{\tilde{\theta}_n\}$ satisfying $\tilde{\theta}_n \in \bar{\Omega}^\circ$, $\tilde{\theta}_n \xrightarrow{a.s.} \theta^\circ$, and $P_n^e[Q_n(\tilde{\theta}_n) = \inf_\Omega Q_n(\theta)] \to 1$ be given. By Jennrich (1969) Lemma 3 there are measurable functions $\bar{\theta}_{itn}(e)$ with range in $\bar{\Omega}^\circ$ such that

$$n^{-\frac{1}{2}}\{F_n'(\tilde{\theta}_n) - F_n'(\theta^\circ)\}e = \sqrt{n}\, D_n(\tilde{\theta}_n - \theta^\circ)'$$

where D_n is the $p \times p$ matrix with row index i and column index j defined by

$$D_n = [n^{-1} \Sigma_{t=1}^n \{\frac{\partial^2}{\partial\theta_j\partial\theta_i} f(x_t, \bar{\theta}_{itn}(e))\}e_t] .$$

Assumptions 1 through 11, 15 imply that Assumptions (a) through (d) of Jennrich are satisfied for the regression model with parameter space replaced by $\bar{\Omega}^\circ$. By Theorems 6, 7 of Jennrich $\tilde{\theta}_n \xrightarrow{a.s.} \theta^\circ$ and $\sqrt{n}\,(\tilde{\theta}_n - \theta^\circ)' \xrightarrow{\mathcal{L}} N_p$. By Lemma 3.4 the elements of D_n converge almost surely to zero. Thus $D_n \sqrt{n}\,(\tilde{\theta}_n - \theta^\circ)' \xrightarrow{P} 0$ and Assumption 13 is satisfied. ▮

4. Uniformly most powerful tests for Problem A. In this section, we will consider Problem A under the assumption that the errors e_t are independently and normally distributed with known variance σ^2, $0 < \sigma^2 < \infty$. Letting m represent Lebesgue measure, P_θ^y will have the density function

$$p_\theta(y) = (2\pi\sigma^2)^{-n/2} \exp\{-\tfrac{1}{2} \sigma^{-2}\|y - f_n(\theta)\|^2\}$$

with respect to m. The results of this section are stated in terms of

Condition 4.1. There is an $n \times 1$ vector η and a function $c: \Omega \longrightarrow [0, \infty)$ such that $f_n(\theta_o) - f_n(\theta) = c(\theta)\eta$.

Lemma 4.1. Let a regression model have normally distributed errors, σ^2 known. For the hypothesis $H: \theta^\circ = \theta_o$ vs. $K_1: \theta^\circ = \theta_1$ at level α where $f_n(\theta_o) \neq f_n(\theta_1)$ and $0 < \alpha < 1$ the unique a.e. m most powerful test is

$$\varphi(y) = \begin{cases} 1 & a'y \leq c \\ 0 & a'y > c \end{cases}$$

where

$$a = f_n(\theta_o) - f_n(\theta_1)$$

$$c = \|a\| \sigma z_\alpha + a'f_n(\theta_o)$$

and z_α satisfies $\Phi(z_\alpha) = \alpha$. (Φ is the distribution function of a standard normal random variable.)

Proof. The proof follows from the Neyman-Pearson fundamental lemma (Lehman, 1959, p. 65). ∎

Theorem 6. Let a regression model have normally distributed errors, σ^2 known.

a) If Condition 4.1 is satisfied there is a uniformly most powerful test for Problem A given by

$$\varphi(y) = \begin{cases} 1 & \eta' y \leq c \\ 0 & \eta' y > c \end{cases}$$

where

$$c = \|\eta\| \sigma z_\alpha + \eta' f_n(\theta_o) .$$

(η is given by Condition 4.1; if $\eta = 0$ set $\varphi(y) \equiv \alpha$.)

b) If Condition 4.1 is not satisfied there does not exist a uniformly most powerful test for Problem A.

Proof. Conclusion (a). Let $\theta_1 \in \Omega_K$ and consider testing $H: \theta^\circ \in \Omega_H$ vs. $K: \theta^\circ = \theta_1$ at level α. If η or $c(\theta_1) = 0$ then $f_n(\theta_o) = f_n(\theta_1)$ and for any test ψ which is level α for H we have $E_{\theta_1} \psi = E_{\theta_o} \psi \leq \alpha = E_{\theta_o} \varphi = E_{\theta_1} \varphi$ so φ is MP for (H, K_1). If $c(\theta_1) > 0$ and $\eta \neq 0$ the MP test φ_1 of (H, K_1) at level α is given by Lemma 4.1 with $a = c(\theta_1)\eta$ and $c = c(\theta_1) [\|\eta\| \sigma z_\alpha + \eta' f_n(\theta_o)]$. Hence $\varphi(y) = \varphi_1(y)$. Thus for every $\theta \in \Omega_K$ φ is MP for $H: \theta^\circ \in \Omega_H$ vs. $K: \theta^\circ = \theta$.

Conclusion (b). Given choices of $\theta_1, \theta_2 \in \Omega_K$ let $a_i = [f_n(\theta_o) - f_n(\theta_i)]$ $(i = 1, 2)$. There is a choice of θ_1 such that $a_1 \neq 0$ or else Condition 4.1 will hold with $\eta = 0$. There is a choice of θ_2 such that $a_2 \neq ca_1$ for any $c \geq 0$ or else Condition 4.1 will hold with $\eta = a_1$. Consider

$H:\ \theta^{\circ} \in \Omega_H$ vs. $K_i:\ \theta^{\circ} = \theta_i$ at level α with associated unique a.e. m MP tests

$$\varphi_i(y) = I_{\{y:\ a_i'y \le c_i\}}(y) \qquad (i = 1, 2)$$

given by Lemma 4.1. If there exists a UMP test φ for Problem A it is MP for (H, K_i) $(i = 1, 2)$ and we have $\varphi_1(y) = \varphi(y) = \varphi_2(y)$ a.e. m.

We will show that $\varphi_1 \neq \varphi_2$ a.e. m so that a UMP test for Problem A cannot exist. Set

$$A = \{y:\ a_1'y < c_1,\ a_2'y > c_2\} \subset \{y:\ \varphi_1(y) \neq \varphi_2(y)\} .$$

Now A is the intersection of two open half spaces of R^n so that $A = \emptyset$ implies $a_1 = ca_2$ where $c > 0$. This is false by the choice of $\theta_1, \theta_2 \in \Omega_K$. Then A is a non-empty open set and $m(A) > 0$. ‖

5. Uniformly most powerful tests for Problem B. In this section we consider Problem B under the assumption of normally distributed errors, σ^2 known. The results of this section are stated in terms of

Condition 5.1. There is an $n \times 1$ vector η and a function $c:\ \Omega_K \to [0, \infty)$ such that $P_H^{\perp}[a_H - f_n(\theta)] = c(\theta)\eta$.

Lemma 5.1. Let the regression model of Problem B have normally distributed errors, σ^2 known, and let $\theta_1 \in \Omega_K$. For the hypothesis $H:\ \theta^{\circ} \in \Omega_H$ vs. $K_1:\ \theta^{\circ} = \theta_1$ at level α where $f_n(\gamma(\theta_1)) \neq f_n(\theta_1)$ the unique a.e. m most powerful test is

$$\varphi(y) = \begin{cases} 1 & a'y \le c \\ 0 & a'y > c \end{cases}$$

where

$$a = P_H^{\perp} [a_H - f_n(\theta_1)]$$

$$c = \|a\| \sigma z_\alpha + a' a_H \ .$$

Proof. By Lemma 4.1 φ is the unique a.e. MP test of H_γ: $\theta^\circ = \gamma(\theta_1)$ vs. K_1: $\theta^\circ = \theta_1$. Since any test which is level α for (H, K_1) must be level α for (H_γ, K_1) we will have φ MP for (H, K_1) if φ is level α for (H, K_1). Let $\theta \in \Omega_H$ then

$$E_\theta \varphi = P_\theta^y \{y: \ a'y \leq c\}$$

$$= \Phi(z_\alpha + (\|a\| \sigma)^{-1} a'[a_H - f_n(\theta)])$$

$$= \Phi(z_\alpha + (\|a\| \sigma)^{-1} a' A_H \theta'_{(1)})$$

$$= \Phi(z_\alpha)$$

because $P_H^{\perp} A_H = 0$ implies $a'A_H = 0$. ∎

Theorem 7. Let a regression model have normally distributed errors, σ^2 known.

a) If Condition 5.1 is satisfied there is a uniformly most powerful test for Problem B given by

$$\varphi(y) = \begin{cases} 1 & \eta' y \leq c \\ 0 & \eta' y > c \end{cases}$$

where

$$c = \|\eta\| \sigma z_\alpha + \eta' a_H .$$

(η is given by Condition 5.1; if $\eta = 0$ set $\varphi(y) \equiv \alpha$.)

b) If Condition 5.1 is not satisfied there does not exist a uniformly most powerful test for Problem B.

Proof. The proof of Theorem 6 may be used word for word with the substitution of $\gamma(\theta_i)$ for θ_o $(i = 1, 2)$, Lemma 5.1 for Lemma 4.1, Condition 5.1 for Condition 4.1, and Problem B for Problem A. ‖

Proposition 5.1. For Problem B if $a(\theta_{(2)}) = 0$ for all $\theta_{(2)} \in \Omega_{(2)}$ and there is a $\theta_{(2)} \neq {}_o\theta_{(2)}$ such that $P_H^\perp A(\theta_{(2)}) \neq 0$ then Condition 5.1 is not satisfied.

Proof. Suppose, to the contrary, that $P_H^\perp f_n(\theta) = c(\theta)\eta$ for all $\theta \in \Omega_K$. Then $P_H^\perp A(\theta_{(2)}) \theta'_{(1)} = c(\theta)\eta$ for all $\theta \in \Omega_K$. Since $\theta_{(1)}$ ranges over R^{p_1} and $c(\theta) \geq 0$ we have $P_H^\perp A(\theta_{(2)}) = 0$ for all $\theta_{(2)} \neq {}_o\theta_{(2)}$. ‖

Remark 5.1. As a consequence of Proposition 5.1 there do not exist UMP tests in such common situations as H: $\theta_2^o = {}_o\theta_2$ vs. K: $\theta_2^o \neq {}_o\theta_2$ when $f(x,\theta) = \theta_1 e^{\theta_2 x}$ or H: $\theta_3^o = {}_o\theta_3$ vs. K: $\theta_3^o \neq {}_o\theta_3$ when $f(x,\theta) = \theta_1 + \theta_2 e^{\theta_3 x}$ (under reasonable choices of $\{x_t\}$) .

6. Likelihood ratio test for Problem A. Under the assumption of normally distributed errors, σ^2 known, the likelihood is

$$L(\theta) = (2\pi\sigma^2)^{-n/2} \exp\{-\tfrac{1}{2}\sigma^{-2} \|y - f_n(\theta)\|^2\} .$$

Theorem 8. Let a regression model with normally distributed errors, σ^2 known, satisfy Assumptions 1, 2, 3. The Likelihood Ratio test for Problem A is

$$\varphi_1(y) = \begin{cases} 1 & t_1(y) \geq c_1 \\ 0 & t_1(y) < c_1 \end{cases}$$

where

$$t_1(y) = \sigma^{-2}\|y - f_n(\theta_o)\|^2 - \sigma^{-2}\|y - f_n(\hat{\theta}_n)\|^2$$

and c_1 is chosen so that $E_{\theta_o}\varphi = \alpha$ (provided such a c_1 exists).

Proof. $\sup_{\Omega_H} L(\theta) = L(\theta_o)$ and by Theorem 1 $\sup_{\Omega} L(\theta) = L(\hat{\theta}_n)$. ‖

The asymptotic distribution of $t_1(y)$ is given by

Theorem 9. Let a regression model with normally distributed errors, σ^2 known, satisfy Assumptions 1 through 11 and 15.

Under the hypothesis H: $\theta^o = \theta_o$

$$t_1(y) \xrightarrow{\mathcal{L}} \chi^2(p) .$$

($\chi^2(p)$ denotes the distribution function of a chi-squared random variable with p degrees freedom.)

Let $\delta = f_n(\theta^o) - f_n(\theta_o)$. Then there is an M such that for all $n > M$

$$t_1(y) = X_n + Y_n + o_p(1)$$

where:

a) X_n and Y_n are independent.

b) X_n is distributed as a non-central χ^2 with p degrees freedom and non-centrality $\lambda = \sigma^{-2}\delta' P_n(\theta^\circ)\delta$ (Graybill, 1961, p. 83).

c) Y_n is distributed as a normal with mean $\sigma^{-2}\delta' P_n^{\perp}(\theta^\circ)\delta$ and variance $4\sigma^{-2}\delta' P_n^{\perp}(\theta^\circ)\delta$.

Proof. The first conclusion follows from the second by putting $\delta = 0$. Apply Theorem 3 and

$$t_1(y) = \sigma^{-2}\{\ \|y - f_n(\theta_o)\|^2 - n\,\hat{\sigma}_n^2\}$$

$$= \sigma^{-2}\{\ \|e + \delta\|^2 - \|P_n^{\perp}(\theta^\circ)e\|^2 - o_p(1)\}$$

$$= \sigma^{-2}\ \|P_n(\theta^\circ)(e + \delta)\|^2 + \sigma^{-2}\{\ \|P_n^{\perp}(\theta^\circ)(e + \delta)\|^2 - \|P_n^{\perp}e\|^2\} - \sigma^{-2}o_p(1)$$

$$= X_n + Y_n + o_p(1)\ .$$

Set $\tilde{\alpha}_n = \tilde{\beta}_n = \theta^\circ$ in Lemma 3.3 then $\det(n^{-1}F_n'(\theta^\circ)F_n(\theta^\circ)) \to \det(F'F)$. By Assumption 11 $\det(F'F) > 0$ and there is an M such that for all $n > M$ $\operatorname{rank}(F_n(\theta^\circ)) = \operatorname{rank}(P_n(\theta^\circ)) = p$. It is easy to verify that X_n and Y_n have the required distributional properties. ▮

When Theorem 9 applies we denote the large sample approximation of the critical point c_1 in Theorem 8 by c_1^*; that is, $P[X \geq c_1^*] = \alpha$ when $X \sim \chi^2(p)$.

7. Likelihood ratio test for Problem B. For Problem B we have that

$$\hat{\hat{\theta}}_n(y) = ((y - a_H)'A_H[A_H'A_H]^+\ ,\ \ {}_o\theta_{(2)})$$

is $(\mathcal{B}_n)$ measurable and satisfies $Q_n(\hat{\hat{\theta}}_n) = \inf_{\Omega_H} Q_n(\theta)$ for every $y \in R^n$. As a consequence we have

Theorem 10. Let the regression model of Problem B have normally distributed errors, σ^2 known, and satisfy Assumptions 1, 2, 3. The Likelihood Ratio test is

$$\varphi_2(y) = \begin{cases} 1 & t_2(y) \geq c_2 \\ 0 & t_2(y) < c_2 \end{cases}$$

where

$$t_2(y) = \sigma^{-2} \|y - f_n(\hat{\hat{\theta}}_n)\|^2 - \sigma^{-2} \|y - f_n(\hat{\theta}_n)\|^2$$

and c_2 is chosen so that $\sup_{\Omega_H} E_\theta \varphi_2 = \alpha$ (provided such a c_2 exists).

Proof. $\sup_{\Omega_H} L(\theta) = L(\hat{\hat{\theta}}_n)$ and by Theorem 1 $\sup_\Omega L(\theta) = L(\hat{\theta}_n)$. ‖

The asymptotic null distribution of $t_2(y)$ is given by

Theorem 11. Let the regression model of Problem B have normally distributed errors, σ^2 known, and satisfy Assumptions 1 through 11 and 15. Under the hypothesis H: $\theta^\circ \in \Omega_H$

$$t_2(y) \xrightarrow{\mathcal{L}} \chi^2(p_2) .$$

Proof. For $\theta^\circ \in \Omega_H$ the matrix $F_n(\theta^\circ)$ is of the form $[A_H \mid F_{(2)}(\theta^\circ)]$. As in the proof of Theorem 9 there is an M such that $n > M$ implies rank $(F_n(\theta^\circ)) = p$ hence rank $(A_H) = p_1$ and rank $(F_{(2)}(\theta^\circ)) = p_2$. All statements in the proof are for $n > \max\{M, N\}$. By Theorem 3 and omitting the subscript n we have

$$t_2(y) = \sigma^{-2} \|e + f(\theta^\circ) - f(\hat{\hat{\theta}})\|^2 - \sigma^{-2} \|P^\perp e\|^2 - \sigma^2 o_p(1)$$

$$= \sigma^{-2} \|e + A_H \theta^{\circ\prime}_{(1)} - A_H \hat{\hat{\theta}}^{\prime}_{(1)}\|^2 - \sigma^{-2}\|P^\perp e\|^2 + o_p(1)$$

$$= \sigma^{-2} \|e + A_H \theta^{\circ\prime}_{(1)} - P_H(e + A_H \theta^{\circ\prime}_{(1)})\|^2 - \sigma^{-2} \|P^\perp e\|^2 + o_p(1)$$

$$= \sigma^{-2} \|P_H^\perp e\|^2 - \sigma^{-2} \|P^\perp(\theta^\circ)e\|^2 + o_p(1)$$

$$= \sigma^{-2} e'(P(\theta^\circ) - P_H)e + o_p(1) .$$

Since $(P(\theta^\circ) - P_H)$ is a symmetric, idempotent matrix with rank p_2 the result follows. ∎

When Theorem 11 applies we denote the large sample approximation of the critical point c_2 in Theorem 10 by c_2^*; that is, $P[X \geq c_2^*] = \alpha$ where $X \sim \chi^2(p_2)$.

8. The case when σ^2 is unknown. When σ^2 is unknown, the obvious approach is to replace σ^2 in the test statistics $t_i(y)$ $i = 1, 2$ by $\hat{\sigma}_n^2$ (provided $\hat{\sigma}_n^2$ exists and is non-zero). The test statistics thus obtained are:

$$s_1(y) = (\hat{\sigma}_n^2)^{-1} [\ \|y - f_n(\theta_0)\|^2 - \|y - f_n(\hat{\theta}_n)\|^2]$$

$$s_2(y) = (\hat{\sigma}_n^2)^{-1} [\ \|y - f_n(\hat{\hat{\theta}}_n)\|^2 - \|y - f_n(\hat{\theta}_n)\|^2] .$$

The corresponding tests are:

$$\psi_i(y) = \begin{cases} 1 & s_i(y) \geq d_i \\ 0 & s_i(y) < d_i \end{cases}$$

where d_i is chosen such that $\sup_{\Omega_H} E_\theta \psi_i = \alpha$ (provided such a d_i exists).

The same conditions which allowed us to approximate the critical points c_i of $\varphi_i(y)$ by c_i^* allow us to approximate the critical points d_i of $\psi_i(y)$ by c_i^*. To see this note that if Assumptions 1 through 8 are satisfied for all $n > N$ then $\hat{\sigma}_n^2 \xrightarrow{a.s.} \sigma^2$. Then if $t_i(y) \xrightarrow{\mathcal{L}} \chi^2(f_i)$ when θ° satisfies H it follows that

$$s_i(y) = [\sigma^2/\hat{\sigma}_n^2]\, t_i(y) \xrightarrow{\mathcal{L}} \chi^2(f_i) .$$

9. Grafted polynomials. At times, a desirable choice of a response function is to take $f(x,\theta)$ to be joined polynomial submodels constrained to be continuous and once differentiable in x ; sometimes called poly- nomial splines. Instances of their use when the join points have been fitted by visual inspection of the data are found in Fuller (1966) and Eppright (1972). We will consider a case when all parameters, including the abscissae of join points, are estimated by least squares; fitting methods are discussed in Gallant and Fuller (1973).

Letting $T_k(r) = r^k$ if $r \geq 0$ and 0 if $r < 0$, a quadratic-quadratic model will join point abscissa θ_5 constrained to be once continuously differentiable in x may be put in the form

$$f(x,\theta) = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 T_2(\theta_5 - x) .$$

Note that the model is of the form described in Problem B with $\theta_{(1)} = (\theta_1, \theta_2, \theta_3, \theta_4)$, $\theta_{(2)} = \theta_5$, $a(\theta_5) = 0$, and $A(\theta_5)$ the $n \times 4$ matrix with rows $(1, x_t, x_t^2, T_2(\theta_5 - x_t))$. Since $T_2(r)$ is only once differentiable with $T_2'(r) = 2T_1(r)$ the second partial derivatives of $f(x,\theta)$ do not exist at all points in $X \times \Omega$.

Let us further specify the model by taking $X = [a, b]$ and $\Omega = R^4 \times [c, d]$ where $-\infty < a < c < d < b < \infty$. The inputs $\{x_t\}_{t=1}^{\infty}$ from X will be chosen as "near" replicates according to the following construction. Choose $q\ (\geq 5)$ points $\{Z_i\}_{i=1}^{q}$ from X so that the first five satisfy $Z_1 < Z_2 < c < d < Z_3 < Z_4 < Z_5$. Choose q sequences $\{Z_{ij}\}_{j=1}^{\infty}$ from X converging to the Z_i at the rate $\Sigma_{j=1}^{\infty}[Z_{ij} - Z_i]^2 < \infty$ for $i = 1, 2, \ldots, q$ and such that $Z_{11} < Z_{21} < c < d < Z_{31} < Z_{41} < Z_{51}$. Assign the inputs according to $x_1 = Z_{11}$, $x_2 = Z_{21}, \ldots, x_q = Z_{q1}$, $x_{q+1} = Z_{12}, \ldots$. Lastly, take the errors $\{e_t\}$ to satisfy Assumption 3, take θ^o interior to Ω, and assume $\theta_4^o \neq 0$.

The quadratic-quadratic model thus specified can be shown to satisfy Assumptions 1 through 13 for all $n \geq 5$; a detailed verification may be found in Gallant (1971). In the next few paragraphs we will sketch the verification of Assumptions 8, 12, 13 for this response function.

Assumption 8. The sequence of measures $\{\mu_n\}$ determined by $\{x_t\}$ converge weakly to the measure defined by $\mu(A) = q^{-1}\, \Sigma_{i=1}^{q}\, I_A(Z_i)$. The matrix $n^{-1}A'(\theta_5)A(\theta_5)$ is positive definite for $n \geq 5$ and using Lemma 3.2 its elements can be shown to converge uniformly to the elements of

$$G(\theta_5) = \left[\int \frac{\partial}{\partial\theta_i} f(x,\theta) \frac{\partial}{\partial\theta_j} f(x,\theta)\, d\mu(x) \right]_{4 \times 4}$$

provided $c \leq \theta_5 \leq d$. Since $n^{-1} \Sigma_{t=1}^{n} f^2(x_t, \theta) = n^{-1}\theta_{(1)}A'(\theta_5)A(\theta_5)\theta'_{(1)}$ it follows that $n^{-1} \Sigma_{t=1}^{n} f^2(x_t, \theta) \leq M$ implies $\Sigma_{i=1}^{4} \theta_i^2 \leq$ trace $(n^{-1}A'(\theta_5)A(\theta_5))^{-1}M$. $G(\theta_5)$ is positive definite for $c \leq \theta_5 \leq d$ and trace $(n^{-1}A'(\theta_5)A(\theta_5))^{-1}$ converges uniformly to trace $G^{-1}(\theta_5)$ as $n \to \infty$ provided $c \leq \theta_5 \leq d$. Then given $\epsilon > 0$ there is an N such that for $n > N$, $n^{-1} \Sigma_{t=1}^{n} f^2(x_t, \theta) \leq M$ implies

$$\|\theta\|^2 \leq \text{trace } (n^{-1}A'(\theta_5)A(\theta_5))^{-1} M + c^2 + d^2$$

$$\leq \sup_{[c,d]} \text{trace } G^{-1}(\theta_5) M + M\epsilon + c^2 + d^2$$

$$< \infty .$$

Assumption 12. Given the point (r_0, s_0) the function $s \cdot T_2(r)$ can be put in the form

$$s \cdot T_2(r) = s_0 T_2(r_0) + T_2(r_0)(s - s_0) + s_0 2T_1(r_0)(r - r_0)$$

$$+ \beta(r, r_0, s, s_0) \|(s, r) - (s_0, r_0)\|^2$$

where $|\beta(r, r_0, s, s_0)| \leq |r_0| + |s|$. Thus

$$f(x,\theta) = f(x,\theta^\circ) + \nabla' f(x,\theta^\circ)(\theta - \theta^\circ)'$$

$$+ \beta(\theta_5 - x, \theta_5^\circ - x, \theta_4, \theta_4^\circ) \|(\theta_4, \theta_5) - (\theta_4^\circ, \theta_5^\circ)\|^2 .$$

The last term above can be put in the form required by Assumption 12.

Assumption 13. If $\bar{\Omega}^\circ$ is a closed and bounded sphere containing θ° there is a finite bound K such that

$$\sup_{\bar{\Omega}^\circ} [f'_k(Z_{ij}, \theta) - f'_k(Z_i, \theta)]^2 \leq K^2(Z_{ij} - Z_i)^2$$

where we have written $f'_k(x,\theta)$ for $\frac{\partial}{\partial\theta_k} f(x,\theta)$ $(k = 1, 2, \ldots, p)$. Given $\tilde{\theta}_n$ with range in $\bar{\Omega}^\circ$ and $\tilde{\theta}_n \xrightarrow{a.s.} \theta^\circ$

$$n^{-\frac{1}{2}} \Sigma_{t=1}^{n} \{f'_k(x_t, \tilde{\theta}_n) - f'_k(x_t, \theta^\circ)\} e_t$$

$$= n^{-\frac{1}{2}} \Sigma_{i=1}^{q} \Sigma_{j=1}^{m_{in}} \{f'_k(Z_{ij}, \tilde{\theta}_n) - f'_k(Z_{ij}, \theta^\circ)\} e_{ij}$$

$$= n^{-\frac{1}{2}} \Sigma_i \Sigma_j \{f'(Z_{ij}, \tilde{\theta}_n) - f'(Z_i, \tilde{\theta}_n)\} e_{ij}$$

$$+ \Sigma_i \left(\frac{m_{in}}{n}\right)^{\frac{1}{2}} \{f'(Z_i, \tilde{\theta}_n) - f'(Z_i, \theta^\circ)\} (m_{in})^{-\frac{1}{2}} \Sigma_j e_{ij}$$

$$+ \Sigma_i \left(\frac{m_{in}}{n}\right)^{\frac{1}{2}} (m_{in})^{-\frac{1}{2}} \Sigma_j \{f'(Z_i, \theta^\circ) - f'(Z_{ij}, \theta^\circ)\} e_{ij}$$

$$= X_n + Y_n + Z_n \ .$$

Now

$$|X_n| \leq n^{-\frac{1}{2}} \Sigma_i \Sigma_{j=1}^{M} |f'(Z_{ij}, \tilde{\theta}_n) - f'(Z_i, \tilde{\theta}_n)| \quad |e_{ij}|$$

$$+ (\Sigma_i \Sigma_{j=M}^{\infty} K^2 (Z_{ij} - Z_i)^2)^{\frac{1}{2}} (n^{-1} \Sigma_i \Sigma_j e_{ij}^2)^{\frac{1}{2}}$$

$$\xrightarrow{a.s.} \sigma K (\Sigma_i \Sigma_{j=M}^{\infty} (Z_{ij} - Z_i)^2)^{\frac{1}{2}}$$

The last term on the right can be made arbitrarily small by varying M hence $X_n \xrightarrow{a.s.} 0$. Since $(m_{in})^{-\frac{1}{2}} \Sigma_j e_{ij} \xrightarrow{\mathcal{L}} N_1$ and $(\frac{m_{in}}{n})^{\frac{1}{2}} \{f'(Z_i, \tilde{\theta}_n) - f'(Z_i, \theta^0)\} \xrightarrow{a.s.} q^{-\frac{1}{2}} \cdot 0 = 0$ we have $Y_n \xrightarrow{a.s.} 0$. Finally, $Z_n \xrightarrow{P} 0$ by Chebishev's inequality.

12. Acknowledgement. I wish to express my appreciation to Professor Wayne A. Fuller for his encouragement and guidance during the preparation of my dissertation of which this work is part.

REFERENCES

[1] Billingsly, P., and Topsoe, F. (1967) Uniformity in weak convergence. Z. Wahrscheinlichkeitstheorie und Verw. Geb. 7, 1-16.

[2] Eppright, E. S., et. al. (1972) Nutrition of infants and preschool children in the north central region of the United States of America. World Review of Nutrition and Dietetics 14, 270-332.

[3] Fuller, W. A. (1969) Grafted polynomials as approximating functions. Austral. J. Ag. Econ. 13, 35-46.

[4] Gallant, A. R. (1971) Statistical inference for nonlinear regression models. Ph. D. Dissertation, Iowa State University.

[5] Gallant, A. R. and Fuller, W. A. (1973) Fitting segmented polynomial models whose join points have to be estimated. J. Amer. Statist. Assoc. 68, 144-147.

[6] Graybill, F. A. (1961) An introduction to linear statistical models, Vol. 1. Wiley, New York.

[7] Jennrich, R. I. (1969) Asymptotic properties of non-linear least squares estimators. Ann. Math. Statist. 40, 633-643.

[8] Lehmann, E. L. (1959) Testing statistical hypotheses. Wiley, New York.

[9] Malinvaud, E. (1970) The consistency of nonlinear regressions. Ann. Math. Stat. 41, 956-969.

[10] Pratt, J. W. (1959) On a general concept of "In Probability." Ann. Math. Statist. 30, 549-558.

[11] Rao, C. R. (1965) Linear statistical inference and its applications. Wiley, New York.