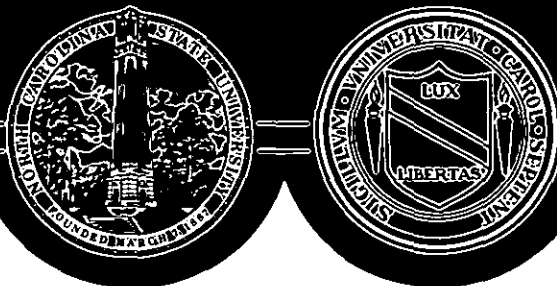


# THE INSTITUTE OF STATISTICS

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

	Anticipated Completion Date -----
1. Univariate Nonlinear Regression	Completed
1.0 Preface	
1.1 Introduction	
1.2 Taylor's Theorem and Matters of Notation	
1.3 Statistical Properties of Least Squares Estimators	
1.4 Methods of Computing Least Squares Estimators	
1.5 Hypothesis Testing	
1.6 Confidence Intervals	
1.7 References	
1.8 Index	
2. Univariate Nonlinear Regression: Special Situations	December 1985
3. A Unified Asymptotic Theory of Nonlinear Statistical Models	Completed
3.0 Preface	
3.1 Introduction	
3.2 The Data Generating Model and Limits of Cesaro Sums	
3.3 Least Mean Distance Estimators	
3.4 Method of Moments Estimators	
3.5 Tests of Hypotheses	
3.6 Alternative Representations of a Hypothesis	
3.7 Independently and Identically Distributed Regressors	
3.8 Constrained Estimation	
3.9 References	
3.10 Index	
4. Univariate Nonlinear Regression: Asymptotic Theory	Completed
4.0 Preface	
4.1 Introduction	
4.2 Regularity Conditions	
4.3 Characterizations of Least Squares Estimators and Test Statistics	
4.4 References	
4.5 Index	
5. Multivariate Linear Models: Review	Deletion likely

Table of Contents (continued)

	Anticipated Completion Date -----
6. Multivariate Nonlinear Models	Completed
6.0 Preface	
6.1 Introduction	
6.2 Least Squares Estimators and Matters of Notation	
6.3 Asymptotic Theory	
6.4 Hypothesis Testing	
6.5 Confidence Intervals	
6.6 Maximum Likelihood Estimators	
6.7 An Illustration of the Bias in Inference Caused by Misspecification	
6.8 References	
6.9 Index	
7. Linear Simultaneous Equations Models: Review	Deletion likely
8. Nonlinear Simultaneous Equations Models	December 1985
8.0 Preface	
8.1 Introduction	
8.2 Three-Stage Least Squares	
8.3 The Dynamic Case: Generalized Method of Moments	
8.4 Hypothesis Testing	
8.5 Maximum Likelihood Estimation	
8.6 Choice of Instruments	
8.7 References	
8.8 Index	
9. Dynamic Nonlinear Models	Completed
9.0 Preface	
9.1 Introduction	
9.2 A Uniform Strong Law and a Central Limit Theorem for Dependent, Nonstationary Random Variables.	
9.3 Data Generating Process	
9.4 Least Mean Distance Estimators	
9.5 Method of Moments Estimators	
9.6 Hypothesis Testing	
9.7 References	
9.8 Index	

## 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

$$y_t = f(x_t, \theta^0) + e_t \quad t = 1, 2, \dots, n$$

with  $\theta^0$  known to lie in some compact set  $\Theta^*$ . The functional form of  $f(x, \theta)$  is known,  $x$  is  $k$ -dimensional,  $\theta$  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

$$y = f(\theta^0) + e$$

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

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

We are interested in testing the hypothesis

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

which we assume can be given the equivalent representation

$$H: \theta^0 = g(\rho^0) \text{ for some } \rho^0 \text{ against } A: \theta^0 \neq g(\rho) \text{ for any } \rho$$

where  $h: \mathbb{R}^p \rightarrow \mathbb{R}^q$ ,  $g: \mathbb{R}^r \rightarrow \mathbb{R}^p$ , and  $p = r + q$ . The correspondence with the notation of Chapter 3 is as follows.

## NOTATION 1

General (Chapter 3)

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

$$\gamma \in \Gamma$$

$$y = Y(e, x, \gamma)$$

$$s(y_t, x_t, \hat{\tau}_n, \lambda)$$

$$\lambda \in \Lambda^*$$

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

$$s_n^\circ(\lambda) = (1/n) \sum_{t=1}^n \int_{\mathcal{E}} s[Y(e, x_t, \gamma_n^\circ), x_t, \tau_n^\circ, \lambda] dP(e)$$

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

$$\hat{\lambda}_n \text{ minimizes } s_n(\lambda)$$

$$\tilde{\lambda}_n \text{ minimizes } s_n(\lambda)$$

$$\text{subject to } h(\lambda) = 0$$

$$\lambda_n^\circ \text{ minimizes } s_n^\circ(\lambda)$$

$$\lambda_n^* \text{ minimizes } s_n^\circ(\lambda)$$

$$\text{subject to } h(\lambda) = 0$$

$$\lambda^* \text{ minimizes } s^*(\lambda)$$

Specific (Chapter 4)

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

$$\theta \in \Theta^*$$

$$y = f(x, \theta) + e$$

$$[y_t - f(x_t, \theta)]^2$$

$$\theta \in \Theta^*$$

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

$$s_n^\circ(\theta) = \sigma^2 + (1/n) \sum_{t=1}^n [f(x_t, \theta_n^\circ) - f(x_t, \theta)]^2$$

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

$$\hat{\theta}_n \text{ minimizes } s_n(\theta)$$

$$\tilde{\theta}_n = g(\hat{\rho}_n) \text{ minimizes } s_n(\theta)$$

$$\text{subject to } h(\theta) = 0$$

$$\theta_n^\circ \text{ minimizes } s_n^\circ(\theta)$$

$$\theta_n^* = g(\rho_n^\circ) \text{ minimizes } s_n^\circ(\theta)$$

$$\text{subject to } h(\theta) = 0$$

$$\theta^* \text{ minimizes } s^*(\theta)$$



## 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

$$y_t = f(x_t, \theta_n^0) + e_t \quad t = 1, 2, \dots, n$$

Assumptions 1 through 5 of Chapter 3 read as follows.

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

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

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

ASSUMPTION 4'. (Identification) The parameter  $\theta^0$  is indexed by  $n$  and the sequence  $\{\theta_n^0\}$  converges to  $\theta^*$ .

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

has a unique minimum over  $\Theta^*$  at  $\theta^*$ .  $\square$

ASSUMPTION 5'.  $\Theta^*$  is compact,  $[e + f(x, \theta^0) - f(x, \theta)]^2$  is dominated by  $b(e, x)$ ;  $b(e, x)$  is that of Assumption 3.  $\square$

The sample objective function is

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

with expectation

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

By Lemma 1 of Chapter 3, both  $s_n(\theta)$  and  $s_n^o(\theta)$  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  $\theta_n^o$  of the unknown parameter is also a minimizer of  $s_n^o(\theta)$  so that our use of  $\theta_n^o$  to denote them both is not ambiguous. We may apply Theorem 3 of Chapter 3 and conclude that

$$\begin{aligned} \lim_{n \rightarrow \infty} \theta_n^o &= \theta^*, \\ \lim_{n \rightarrow \infty} \hat{\theta}_n &= \theta^* \text{ almost surely.} \end{aligned}$$

Assumption 6 of Chapter 3 may be restated as follows.

ASSUMPTION 6!  $\Theta^*$  contains a closed ball  $\Theta$  centered at  $\theta^*$  with finite, nonzero radius such that

$$\begin{aligned} (\partial/\partial\theta_i) s[Y(e, x, \theta^o), x, \theta] &= -2[e + f(x, \theta^o) - f(x, \theta)] (\partial/\partial\theta_i) f(x, \theta) \\ (\partial^2/\partial\theta_i \partial\theta_j) s[Y(e, x, \theta^o), x, \theta] &= 2[(\partial/\partial\theta_i) f(x, \theta)] [(\partial/\partial\theta_j) f(x, \theta)] \\ &\quad - 2[e + f(x, \theta^o) - f(x, \theta)] (\partial^2/\partial\theta_i \partial\theta_j) f(x, \theta) \\ \{(\partial/\partial\theta_i) s[Y(e, x, \theta^o), x, \theta]\} \{(\partial/\partial\theta_j) s[Y(e, x, \theta^o), x, \theta]\} \\ &= 4[e + f(x, \theta^o) - f(x, \theta)]^2 [(\partial/\partial\theta_i) f(x, \theta)] [(\partial/\partial\theta_j) f(x, \theta)] \end{aligned}$$

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

Moreover,

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

is nonsingular.  $\square$

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).

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

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

$$u^* = 0$$

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

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

$$u_n^\circ = 0$$

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

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

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

Noting that

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

we have from Theorem 4 of Chapter 3 that

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

and from Theorem 5 that

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

$$\lim_{n \rightarrow \infty} Q_n^\circ = Q.$$

The Pitman drift assumption is restated as follows.

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

Noting that

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

we have from Theorem 6 that

$$\lim_{n \rightarrow \infty} \tilde{\theta}_n = \theta^* \text{ almost surely}$$

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

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

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

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

Assumption 13 of Chapter 3 is restated as follows.

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

## PROBLEMS

1. Use the derivatives given in Assumption 6 to compute  $\bar{\mathcal{J}}(\theta)$ ,  $\bar{\mathcal{J}}'(\theta)$ ,  $\bar{u}(\theta)$  and  $\bar{\mathcal{J}}(\theta)$ ,  $\bar{\mathcal{J}}'(\theta)$ ,  $\bar{u}(\theta)$  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  $\theta_n^\circ$  are in  $\Theta$  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  $n \geq N$  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

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

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

The next characterization that needs justification is

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

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

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

From the preceding 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

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

A straightforward argument using Taylor's theorem yields

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

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

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

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

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

We next show that

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

where

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

Fix a realization of the errors  $\{e_t\}$  for which  $\lim_{n \rightarrow \infty} s^2 = \sigma^2$  and  $\lim_{n \rightarrow \infty} e'(I - P_F)e/(n-p) = \sigma^2$ ; 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  
 $\lim_{n \rightarrow \infty} [(n-p)/e'(I - P_F)e]^2 = 1/\sigma^4$ . 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

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

$$\lambda = h'(\theta_n^0) \{H(\theta_n^0) [F'(\theta_n^0) F(\theta_n^0)]^{-1} H'(\theta_n^0)\}^{-1} h(\theta_n^0) / (2\sigma^2) .$$

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

$$\begin{aligned} \sqrt{n} h(\hat{\theta}_n) &= \sqrt{n} h(\theta_n^0) + \sqrt{n} H(\theta_n^0) [F'(\theta_n^0) F(\theta_n^0)]^{-1} F'(\theta_n^0) e + o_p(1) \\ &= \mu + U + o_p(1) \end{aligned}$$

$$\begin{aligned} \{ \hat{H}[(1/n)F'(\hat{\theta}_n)F(\hat{\theta}_n)]^{-1} \hat{H}' \}^{-1} &= \{ H(\theta_n^0) [(1/n)F'(\theta_n^0)F(\theta_n^0)]^{-1} H'(\theta_n^0) \}^{-1} + o_p(1) \\ &= A^{-1} + o_p(1) \end{aligned}$$

whence

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

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

$$\begin{aligned} \lambda &= \mu' A^{-1} \mu / (2\sigma^2) \\ &= n h'(\theta_n^\circ) \{ H(\theta_n^\circ) [(1/n) F'(\theta_n^\circ) F(\theta_n^\circ)]^{-1} H'(\theta_n^\circ) \}^{-1} h(\theta_n^\circ) / (2\sigma^2) . \end{aligned}$$

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

$$\begin{aligned} (\text{SSE}_{\text{full}})/n &= e' P_F^\perp e / n + o_p(1/n) \\ n / (\text{SSE}_{\text{full}}) &= n / e' P_F^\perp e + o_p(1/n) \end{aligned}$$

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

$$\begin{aligned} \text{SSE}_{\text{full}} &= \|y - f(\hat{\theta}_n)\|^2 \\ \text{SSE}_{\text{reduced}} &= \|y - f(\tilde{\theta}_n)\|^2 = \|y - f[g(\hat{\rho}_n)]\|^2 . \end{aligned}$$

The claim that

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

with

$$\begin{aligned} \delta &= f(\theta_n^\circ) - f(\theta_n^*) = f(\theta_n^\circ) - f[g(\rho_n^\circ)] \\ P_{FG}^\perp &= I - P_{FG} = I - F(\theta_n^\circ) G(\rho_n^\circ) [G'(\rho_n^\circ) F'(\theta_n^\circ) F(\theta_n^\circ) G(\rho_n^\circ)]^{-1} G'(\rho_n^\circ) F'(\theta_n^\circ) \end{aligned}$$

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

$$\begin{aligned}\bar{H} \sqrt{n} (\tilde{\theta}_n - \theta_n^*) &= o_s(1) \\ \sqrt{n} (\tilde{\theta}_n - \theta_n^*) &= \bar{g}^{-1} \sqrt{n} (\partial/\partial\theta)_{s_n}(\tilde{\theta}_n) - \bar{g}^{-1} \sqrt{n} (\partial/\partial\theta)_{s_n}(\theta_n^*) + o_s(1)\end{aligned}$$

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

$$\begin{aligned}H \sqrt{n} (\tilde{\theta}_n - \theta_n^*) &= o_s(1) \\ \sqrt{n} (\tilde{\theta}_n - \theta_n^*) &= g^{-1} [\sqrt{n} (\partial/\partial\theta)_{s_n}(\tilde{\theta}_n) - \sqrt{n} (\partial/\partial\theta)_{s_n}(\theta_n^*)] + o_p(1)\end{aligned}$$

with  $g = g_n^\circ$  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

$$\begin{aligned}\sqrt{n} [(\partial/\partial\theta)_{s_n}(\tilde{\theta}_n)]' \sqrt{n} (\tilde{\theta}_n - \theta_n^*) &= o_p(1) \\ \sqrt{n} (\tilde{\theta}_n - \theta_n^*) &= -g^{-1} [g - H'(Hg^{-1}H')^{-1}H]g^{-1} \sqrt{n} (\partial/\partial\theta)_{s_n}(\theta_n^*) + o_p(1).\end{aligned}$$

Then using Taylor's theorem

$$\begin{aligned}n[s_n(\tilde{\theta}_n) - s_n(\theta_n^*)] &= -n[(\partial/\partial\theta)_{s_n}(\tilde{\theta}_n)](\tilde{\theta}_n - \theta_n^*) - (n/2)(\tilde{\theta}_n - \theta_n^*)' [g + o_s(1)](\tilde{\theta}_n - \theta_n^*) \\ &= (-n/2)(\tilde{\theta}_n - \theta_n^*)' g(\tilde{\theta}_n - \theta_n^*) + o_p(1) \\ &= (-n/2)[(\partial/\partial\theta)_{s_n}(\theta_n^*)]' [g^{-1} - g^{-1}H'(Hg^{-1}H')^{-1}Hg^{-1}] [(\partial/\partial\theta)_{s_n}(\theta_n^*)].\end{aligned}$$

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

$$g^{-1} - g^{-1}H'(Hg^{-1}H')^{-1}Hg^{-1} = G(G'gG)^{-1}G'$$

whence

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

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

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

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

$$(\text{SSE}_{\text{reduced}})/(\text{SSE}_{\text{full}}) = (e + \delta)' P_{FG}^{\perp} (e + \delta) / e' P_F^{\perp} e = o_p(1/n)$$

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

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

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

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

$$\begin{aligned} \tilde{D}' \tilde{F}' \tilde{F} \tilde{D} &= [y - r(\tilde{\theta}_n)]' \tilde{F}' (\tilde{F}' \tilde{F})^{-1} \tilde{F}' [y - r(\tilde{\theta}_n)] \\ &= (n/4) [(\partial/\partial\theta) s_n(\tilde{\theta}_n)]' [(1/n) \tilde{F}' \tilde{F}]^{-1} [(\partial/\partial\theta) s_n(\tilde{\theta}_n)] \\ &= (n/2) [(\partial/\partial\theta) s_n(\tilde{\theta}_n)]' [\mathcal{J} + o_s(1)]^{-1} [(\partial/\partial\theta) s_n(\tilde{\theta}_n)] \\ &= (n/2) [(\partial/\partial\theta) s_n(\theta_n^*)]' \mathcal{J}^{-1} H' (H \mathcal{J}^{-1} H')^{-1} H \mathcal{J}^{-1} [(\partial/\partial\theta) s_n(\theta_n^*)] + o_p(1) \\ &= (n/2) [(\partial/\partial\theta) s_n(\theta_n^*)]' [\mathcal{J}^{-1} - G(G' \mathcal{J} G)^{-1} G'] [(\partial/\partial\theta) s_n(\theta_n^*)] + o_p(1) \\ &= (1/n) (e + \delta)' F(\theta_n^0) [(Q_n^0)^{-1} - G(G' Q_n^0 G)^{-1} G'] F'(\theta_n^0) (e + \delta) + o_p(1) \\ &= (1/n) (e + \delta)' F(\theta_n^0) [(Q_n^0)^{-1} - G(G' Q_n^0 G)^{-1} G'] F'(\theta_n^0) (e + \delta) + o_p(1) \\ &= (e + \delta)' (P_F - P_{FG}) (e + \delta) + o_p(1) . \end{aligned}$$

## 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  $\lim_{n \rightarrow \infty} s^2 = \sigma^2$  almost surely and prove directly that  $\lim_{n \rightarrow \infty} e'(I - P_F)e/(n-p) = \sigma^2$  almost surely.

3. Show in detail that  $(\partial/\partial\theta)s_n(\theta_n^*) = (-2/n)F'(\theta_n^0)(e+\delta) + o_p(1/\sqrt{n})$  suffices to reduce  $(n/2)[(\partial/\partial\theta)s_n(\theta_n^*)]'G(G'G - G)^{-1}G[(\partial/\partial\theta)s_n(\theta_n^*)]$  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

Assumption 1', 4-2-1  
 Assumption 2', 4-2-1  
 Assumption 3', 4-2-1  
 Assumption 4', 4-2-1  
 Assumption 5', 4-2-1  
 Assumption 6', 4-2-2  
 Assumption 7', 4-2-4  
 Assumption 13', 4-2-4  
 Characterizations  
   estimator of variance, 4-3-2  
   Lagrange multiplier test statistic, 4-3-8  
   least squares estimator, 4-3-1  
   likelihood ratio test statistic, 4-3-6  
   nonlinear function of the least squares estimator, 4-3-3  
   reciprocal of the estimator of variance, 4-3-3  
   restricted estimator of variance, 4-3-5  
   Wald test statistic, 4-3-4  
 Least squares estimators for univariate models  
   asymptotic normality, 4-2-4  
   characterization as a linear function of the errors, 4-3-1  
   consistency, 4-2-2  
   defined, 4-1-1  
   data generating model, 4-2-1  
   identification, 4-2-1  
   sample objective function, 4-1-1, 4-2-2  
   score function, 4-2-4  
   Pitman drift, 4-2-4  
 Notation 1, 4-1-2  
 Notation 2, 4-1-3  
 Univariate nonlinear regression model  
   defined, 4-1-1  
   identification, 4-2-1  
   Pitman drift, 4-2-4