UNIVERSIDADE NOVA DE LISBOA

Faculdade de Economia

NONLINEAR ESTIMATION: SOME STANDARD RESULTS

Fernando M. L. Chau

Working Paper nº 12

UNIVERSIDADE NOVA DE LISBOA
Faculdade de Economia
Campo Grande, 185
1700 LISBOA

Dezembro, 1983

NONLINEAR ESTIMATION:

SOME

STANDARD RESULTS

by

Fernando M. L. Chau

Nonlinear models arise naturally in economics. Both least squares and maximum-likelihood estimators are considered; strong consistency and asymptotic normality for the estimators are showed. A statistic for hypothesis testing is presented.

Keywords: Nonlinear estimation; maximum likelihood estimator; least squares estimator; asymptotic distribution; hypothesis testing.

I'm very grateful to Prof. J.D. Coelho for encouragement and sustaining interest for this study. Remaining errors are, of course, mine.

1.

WITH few exceptions, nonlinear estimation is considered a hard subject by the econometric textbooks. Since research on it had reached promising developments, our aim is to introduce some standard results to one not acquainted with this area.

Prior to the statistical results, the computational aspects has its own interest. Burguette et all (1981) showed that a very large methods of estimation subsumes as an optimization problem; then, the recourse to nonlinear programming algorithms produce numerical content to the nonlinear estimator. For the single equation case, Hartley (1961) proposed a modified Gauss-Newton method for the nonlinear Least Square (LS) estimator. Beauchamp and Cornell(1966) used a multivariate modified Gauss-Newton method for the multivariate model, but Gallant (1975) sinmplified their approach with an iterated generalized LS.

For the simultaneous equation model, Eisenpress and Greenstad (1966) and Chow (1973) proposed the Newton method; Berndt et all (1974) presented the gradient method which seems to be the most efficient one.

On the comparative advantage of nonlinear LS versus Maximum Likelihood (ML) estimators see Belsley (1979,1980).

As we shall see, explicit forms for the nonlinear estimators do not exist and so, finite sample properties are awkward to work out. Therefore, analysis is concerned about their asymptotic properties.

When the sample size increases, if the estimator converges to its true value we say that it's a consistent estimator. Moreover, if that convergence holds almost surely, i.e. the sets where the estimator and the true value are different have null probability, then it's a strong consistent estimator; and if it holds in probability then it's a weak consistent estimator.
As it turns out, strong consistency is more easy to obtain rather than weak consistency.

Jennrich (1969) provides the first rigorous proof for the consistency of the nonlinear estimator, Malinvaud (1970) obtained weak consistency while former dealed with strong consistency.

We conclude this introduction by outlining the remaining sections. Next section the model and dhe notations are introduced. Section 3 discuss the sufficient condidions for the existence of the nonlinear estimator. Next we show that it is strong consistent. Section 4 its asympotic distribution and a statistic for hypothesis testing ars presented.

2.

We restrict our atention to the scalar (single equation) model with an additive disturbance component. We have

A.1

(1) $$y_t = f(x_t,\theta_o) + e_t \qquad t = 1,\ldots,n$$

where e_t is the error term, following i.i.d. with zero mean and variance σ^2. $f(.)$ is a (known) continuous function on a compact set $\textcircled{H}$, a subset of an Euclidean space.

The LS estimator $\hat{\theta}_{LS}$ is the solution of

(2) $$\min_{\theta\in\textcircled{H}} Q_n(\theta) = \min_{\theta\in\textcircled{H}} 1/n \sum_{t=1}^{n} (y_t - f(x_t,\theta))^2$$

The conditional density function of y_t is denoted as $h(y_t|x_t,\theta)$ and its log-likelihood is

$$L_n(\theta) = \sum_{t=1}^{n} \log h(y_t|x_t,\theta)$$

The ML estimator of (1) $\hat{\theta}_{ML}$ is the solution of

(3) $$\min_{\theta\in\textcircled{H}} -L_n(\theta) = \min_{\theta\in\textcircled{H}} -1/n \sum_{t=1}^{n} \log h(y_t|x_t,\theta)$$

Let

(4) $$g(x_t,\theta) = \begin{cases} f(x_t,\theta) \\ -h(y_t|x_t,\theta) \end{cases} \quad \text{and} \quad d(y_t,\theta) = \begin{cases} (y_t - f(x_t,\theta))^2 \\ -\log h(y_t|x_t,\theta) \end{cases}$$

where the upper case is for the LS and the bottom for the ML estimator.

To conclude this section we define

(5) $$D_n(\theta) = \sum_{t=1}^{n} d(y_t,\theta)$$

3.

Now we use $\hat{\theta}$ for the estimator of θ without distinguishing it as the LS or the ML estimator. As an estimator, $\hat{\theta}$ is a function of the random sample and, on the other hand, it's a random variable. Being a function of random variables doesn't imply that it's a random variable.

Suppose that $\hat{\theta}$ is an estimator of θ, then (by the definition of random variable)

(6) $\quad P[\hat{\theta} < t] = G(t) \qquad$ for all $t \in R$

where G(.) is a distribution function. But (6) is equivalent do say that the set

$$\{ w \mid \hat{\theta}(w) < t \}$$

is measurable and that $\hat{\theta}$ is a measurable function (Rudin, pg 311). Thus to find an estimator is equivalent to say that $\hat{\theta}$ is a measurable function. The following proposition provides sufficient conditions for the existence of $\hat{\theta}$.

Proposition 1

Let $D_n(y,\theta)$ be a measurable function on a measurable space Y and for each y in Y a continuous function for θ in a compact set (H). Then there exists a measurable function $\hat{\theta}_n(y)$ such that

$$D_n(y,\hat{\theta}_n(y)) = \inf_{\theta \in (H)} D_n(y,\theta) \qquad \text{for all } y \text{ in } Y$$

Proposition 1 is the first part of Lema 3 of Amemiya (1973,pg.1002). For proof see Jennrich (1969,pg.637).

For the $\hat{\theta}_{LS}$ to be well defined, $Q_n(\theta)$ of (2) must be measurable in Y and continuous in Ⓗ . f(.) is continuous, the sums, powers and minimum of continuous functions are continuous so $Q_n(\theta)$ is continuous on both Y and Ⓗ. As continuous functions are measurable (Bierens,1981,Theorem 2.1.4,pg 11) then $\hat{\theta}_{LS}$ exists.

Now, by assumption h(.) is a continuous function. Using similar arguments of the above paragraph, we conclude that $\hat{\theta}_{ML}$ is well defined.

4.

Our interest is centered about the behaviour of the (limit of the) sequence $\hat{\theta}_n(y)$. Since the latter is a solution of (2) or (3), it's convergence depends on the sequence of random functions $Q_n[\theta(y)]$ or $L_n[\theta(y)]$.

What kind of convergence is needed for $D_n(.)$? As $n\to\infty$ the sequence $\{x_t, e_t\}$ gives all the information it contains. Clearly, the type of convergence of $D_n(.)$ depends on the nature of Ⓗ. Suppose, for a while, that it contains just two points; then pointwise convergence suffices because we can evaluate $D_n(.)$ at these points and pick up the desired one (i.e. which gives the minimum). Under A.1 Ⓗ is compact, hence there are an infinity of random functions $D_n(\theta)$ in an infinitesimal neighborhood of θ_o ; in this case, the convergence must be uniform on the set Ⓗ.

$D_n(\theta)$ is the sums and/or power or log of $g(x_t,\theta)$, so the convergence of the former depends on the (conditions imposed for the) convergence of the latter. This is known as the indirect approach. The direct approach is due to Malinvaud (1970) where the conditions are imposed on the variables x_t . We shall follow here the indirect tradition. Recall that $D_n(.)$ depends g(.) through d(.).

The next proposition gives sufficient conditions for the convergence of $D_n(\theta)$.

Proposition 2

Let $x_1, x_2, \ldots$ be a sequence of independent random vectors with distribution F. Let $d(x_t,\theta)$ be a function on $X \times \textcircled{H}$ where X is an Euclidean space and $\textcircled{H}$ a compact subset of an Euclidean space. Let $d(x_t,\theta)$ be a continuous function of θ for each x and a measurable function of x for each θ. Suppose $|d(x_t,\theta)| \leq m(x)$ for all x and θ, where m is integrable with respect to F. Then, for almost every sequence $\{x_t\}$

$$1/n \sum_{t=1}^{n} d(x_t,\theta) \rightarrow \int d(x,\theta)\, dF(x)$$

uniformly for all θ in $\textcircled{H}$.

This is Jennrich's Theorem 2 (pg. 636) known as Mickey theorem.

For the fixed regressors case, F is a degenerate distribution function.

If the conditions of proposition 2 are fulfilled can we say that the $\lim \tilde{\theta}_n$ as $n\rightarrow\infty$ exists? The answer is no. Because $D_n(.)$ can have more than one solution. When this happens, we say the model is not identifiable.

In LS estimation, Wu(1981) provides a very intuitive discussion related to this problem. Recall that θ_o is the true parameter, and $\tilde{\theta}_{LS}$ is the solution of

$$(2) \qquad \min_{\theta \in \textcircled{H}} 1/n \sum_{t=1}^{n} (f(x_t,\theta_o) - f(x_t,\theta))^2$$

Then, as $n\rightarrow\infty$ and for $\theta \neq \theta_o$ we expect that the sum of squares diverges to infinity. This is equivalent to say that θ_o is the unique minimum as $n\rightarrow\infty$. If there is an $\theta \neq \theta_o$ such that (2) is finite as $n\rightarrow\infty$ than the model is not identifiable and there are no consistent estimators for θ_o (see Wu, Theorem 1 pg. 503). The following example (brought freely from Wu,1981,pg 503) would clarify the above discussion.

Consider the model $y_t = \exp(-\alpha t) + e_t$, $t = 1, \ldots$, $\alpha \in (0, 2\pi)$. For $\alpha \neq \beta \in (0, 2\pi)$ we see that $\sum_{t=1}^{n} (\exp(-\alpha t) - \exp(-\beta t))^2$ is finite as $n \to \infty$, and so any estimator of α is inconsistent.

The following assumptions are sufficient for the strong consistency of $\tilde{\theta}$.

A.2

(i) The family of distribution $F(x,\theta)$ has Radon-Nikodým densities $s(x,\theta) = dF(x,\theta)/d\mu$ which are measurable in x for every θ in (H), a compact subset of an Euclidean space, and continuous in θ for every x in X.

(ii) $f(x_t,\theta)$ is a continuous function of θ for each x in X and a measurable function of x for each θ in (H), a compact subset of an Euclidean space.

A.3

(i) $E[d(x,\theta)]$ exists and $|d(x,\theta)| \leq m(x)$ for all θ in (H) where m is integrable with respect to F;

(ii) $D_n(\theta)$ has an unique minimum at θ_o in (H).

This comment regards the uniqueness assumption. In practise, any algorithm needs a starting point and, if convergence is attained, that is, the problem with that starting point provide a convergent series, than the limit is the required solution. By suitably changing (some components of) the starting point, we get (some) assurance about the uniqueness of the solution. Of course, global uniqueness for all θ is not necessary since one can restrict (H) such that there is only one (stable) solution.

Proposition 3

Given assuptions A.1 - A.3, then $\tilde{\theta}_n \to \theta_o$ a.s..

Proof:

(The proof is due to Jennrich and Amemiya(1973); we added some minor steps.)

With A.1 and A.2, conditions of proposition 1 are fulfilled so $\hat{\theta}_n$ exists and is measurable. Conditions of proposition 2 are met under A.3 so that $D_n(x,\theta) \to D(x,\theta)$ $\left\{= \int d(x,\theta)dF(x) = E[d(x,\theta)]\right\}$ for almost every sequence $\{x_t\}$ and uniformly for all θ in Ⓗ. Let N be an open neighborhood around θ_o such that $\overline{N}$, its complement in Ⓗ, is compact. Therefore, $\min_{\theta \in \overline{N}} D(\theta)$ exists. Let $D(\theta^*) = \min_{\theta \in \overline{N}} D(\theta)$. Denote $\varepsilon = D(\theta^*) - D(\theta_o)$.

The difference $D_n(\hat{\theta}_n) - D(\theta^*)$ can be written as

$$D_n(\hat{\theta}_n) - D(\theta^*) = D_n(\hat{\theta}_n) - D(\theta_o) + D(\theta_o) - D(\theta^*)$$

Since $D_n(\hat{\theta}_n) = \min_{\theta \in Ⓗ} D(\theta) \leq \min_{\theta \in \overline{N}} D(\theta) = D(\theta^*)$, we have

$$| D_n(\hat{\theta}_n) - D(\theta_o)| \leq \varepsilon$$

by the triangle inequality

$$|D_n(\hat{\theta}_n) - D(\theta^*)| \leq \varepsilon/2$$

but this means that $\hat{\theta}_n \in N$. Since N is arbitrary (close to θ_o), $\hat{\theta}_n$ converges to θ_o a.s..

For the ML estimation, the variance σ^2 is estimated as an element of the vector θ. This is not the case for the LS. By (1) we can rewrite (2) as

$$(2') \qquad \min_{\theta \in Ⓗ} \; 1/n \sum_{t=1}^{n} e_t^2$$

and under A.1-A.3, Kolmogorov's strong law of large numbers applies and therefore (2') is a strong consistent estimator for σ^2. To see this, firs note that by A.1 e_t^2 follows i.i.d., on the other hand $E\, e_t^2$ exists and is finite, therefore, $E|e_t| < \infty$ (Loève, pg.121) and the Kolmogorov result applies.

Formally, we have

Proposition 4

Under A.1

$$Q_n(\theta) \to \sigma^2 \text{ a.s..}$$

As a byproduct of the numerical optimization of (2), one has an estimate for σ^2 given by $Q_n(\bar{\theta}_n)$.

5.

Consistency of the estimators is the first part of our task. For purposes of statistical inference, knowledge about the asymptotic distribution is needed.

$\bar{\theta}_n$ is the solution of (2) or (3), then it is true that

$$\left.\frac{\partial D_n(\theta)}{\partial \theta}\right|_{\bar{\theta}_n} = 0 \tag{7}$$

As $\bar{\theta}_n \to \theta_o$ a.s., eventually, $\bar{\theta}_n$ is located in a convex compact neighborhood of θ_o. If θ_o is an interior point of Ⓗ, then by Taylor expansion, we have

$$\left.\frac{\partial D_n(\theta)}{\partial \theta}\right|_{\bar{\theta}_n} = 0 = \left.\frac{\partial D_n(\theta)}{\partial \theta}\right|_{\theta_o} + \left.\frac{\partial^2 D_n(\theta)}{\partial \theta \, \partial \theta'}\right|_{\theta_1} (\bar{\theta}_n - \theta_o)$$

or

$$(\bar{\theta}_n - \theta_o) \left.\frac{\partial^2 D_n(\theta)}{\partial \theta \, \partial \theta'}\right|_{\theta_1} = - \left.\frac{\partial D_n(\theta)}{\partial \theta}\right|_{\theta_o} \tag{8}$$

where $\theta_1 = \bar{\theta}_n + \lambda(\theta_o - \bar{\theta}_n)$ for $\lambda \in [0,1]$. First, we note that $\left.\frac{\partial D_n(\theta)}{\partial \theta}\right|_{\theta_o}$

follows i.i.d. under A.1. Next, $E \left. \frac{\partial D_n(\theta)}{\partial \theta} \right|_{\theta_o} = 0$ (White, 1981 pg 429).

Under appropriate conditions to be introduced, we can apply Lindeberg-Lévy central limit theorem to the right hand side of (8). If the matrix $\left. \frac{\partial^2 D_n(\theta)}{\partial \theta \partial \theta'} \right|_{\theta_1}$ converges to $\left. \frac{\partial^2 D_n(\theta)}{\partial \theta \partial \theta'} \right|_{\theta_o}$ a.s. and if its inverse exists, then the asymptotic distribution of $(\bar{\theta}_n - \theta_o)$ follows at once.

Before we introduce the appropriate assumptions, let us consider some notation. When the partial derivatives exist, we have the matrices

$$A_n(\theta) = \left\{ 1/n \sum_{t=1}^{n} \partial^2 d(x_t,\theta)/\partial\theta_i \partial\theta_j \right\} \qquad \text{all } i \text{ and } j$$

$$B_n(\theta) = \left\{ 1/n \sum_{t=1}^{n} \partial d(x_t,\theta)/\partial\theta_i \cdot \partial d(x_t,\theta)/\partial\theta_j \right\} \qquad " \ i \ " \ j$$

If the expectations exist, we have

$$A(\theta) = \left\{ E[\partial^2 d(x_t,\theta)/\partial\theta_i \partial\theta_j] \right\} \qquad " \ i \ " \ j$$

$$B(\theta) = \left\{ E[\partial d(x_t,\theta)/\partial\theta_i \cdot \partial d(x_t,\theta)/\partial\theta_j] \right\} \qquad " \ i \ " \ j$$

And if the inverses exist, let

$$C_n(\theta) = A_n(\theta)^{-1} B_n(\theta) A_n(\theta)^{-1}$$

(9)

$$C(\theta) = A(\theta)^{-1} B(\theta) A(\theta)^{-1}$$

The following assumptions are sufficient for the application of Lindeberg-Lévy central limit theorem to $n^{-1/2} \left. \frac{\partial D_n(\theta)}{\partial \theta} \right|_{\theta_o}$.

A.4

$\partial d(x_t,\theta)/\partial\theta_i$ for all i is a measurable function of x for each θ in (H) and continuously differentiable function of θ for each x in X.

A.5

$|\partial^2 d(x_t,\theta)/\partial\theta_i\partial\theta_j|$ and $|\partial d(x_t,\theta)/\partial\theta_i \cdot \partial d(x_t,\theta)/\partial\theta_j|$ for all i and j are dominated by functions integrable with respect to F for all x in X and θ in (H).

A.6

(i) θ_o is an interior of (H);

(ii) $A(\theta_o)$ and $B(\theta_o)$ are nonsingular.

First, we rearrange expression (8) as

$$(10) \qquad n^{-1/2}(\tilde{\theta}_n - \theta_o) = - \left[\frac{\partial^2 D_n(\theta)}{\partial\theta\,\partial\theta'}\bigg|_{\theta_1}\right]^{-1} n^{-1/2} \frac{\partial D_n(\theta)}{\partial\theta}\bigg|_{\theta_o}$$

if the inverse exists. As $n\to\infty$, $\tilde{\theta}_n \to \theta_o$ a.s., then θ_1 converges to θ_o a.s.. Hence, the inverse in the r.h.s. of (10) converges to $\left[\frac{\partial^2 D_n(\theta)}{\partial\theta\,\partial\theta'}\bigg|_{\theta_o}\right]^{-1}$.

Following the above discussion and assumptions, we have

Proposition 5

Under A.1-A.6

$$(11) \qquad n^{-1/2}(\tilde{\theta}_n - \theta_o) \overset{A}{\frown} N(0, C(\theta_o))$$

Moreover, $C_n(\tilde{\theta}_n) \to C(\theta_o)$ a.s., element by element.

This is White theorem 3.3 (1981,pg 430). For the LS estimator with fixed regressors, the asymptotic distribution can be obtained using (11). With fixed regressors, e_t is independent from f(.). Using

$$f_{ti} = \partial f(x_t,\theta)/\partial\theta_i \quad \text{and}$$

$$f_{tij} = \partial^2 f(x_t,\theta)/\partial\theta_i\partial\theta_j$$

then

$$A_n(\theta) = \left\{ 2/n \sum_{t=1}^{n} (f_{ti} f_{tj} - e_t f_{tij}) \right\}$$

$$B_n(\theta) = \left\{ 4/n^2 \sum_{t=1}^{n} e_t^2 f_{ti} f_{tj} \right\}$$

applying expectation

$$A(\theta) = \left\{ 2/n \sum_{t=1}^{n} f_{ti} f_{tj} \right\}$$

$$B(\theta) = \left\{ 4/n \sigma^2 \sum_{t=1}^{n} f_{ti} f_{tj} \right\}$$

because under independence of e_t with f(.),

$$\sum_{t=1}^{n} E(e_t f_{tij}) = \sum_{t=1}^{n} E e_t \sum_{t=1}^{n} f_{tij} = 0 \cdot \sum_{t=1}^{n} f_{tij} = 0$$

$$1/n \sum_{t=1}^{n} E(e_t^2 f_{ti} f_{tj}) = 1/n \sum_{t=1}^{n} E e_t^2 \sum_{t=1}^{n} f_{ti} f_{tj} = \sigma^2 \sum_{t=1}^{n} f_{ti} f_{tj}$$

so

$$C(\theta) = \sigma^2 [2/n \sum_{t=1}^{n} f_{ti} f_{tj}]^{-1} [4/n \sum_{t=1}^{n} f_{ti} f_{tj}] [2/n \sum_{t=1}^{n} f_{ti} f_{tj}]^{-1}$$

$$= \sigma^2 [1/n \sum_{t=1}^{n} f_{ti} f_{tj}]^{-1}$$

which is Jennrich's theorem 7.

For statistical testing, (11) is a valuable result. The next one is taken from White (1980).

Let the null be given as

H_o: $s(\theta) = 0$

against

H_1: $s(\theta) \neq 0$

where

$s: \textcircled{H} \rightarrow R^r$

Under the following assumption, we had the desired result.

A.7

$s : \textcircled{H} \rightarrow R^r$ is a continuously differentiable function on $\textcircled{H}$ such that the Jacobian at θ_o is finite and has full row rank r.

Then, the statistic

(12) $$n\, s(\hat{\theta}_n)' [\nabla s(\hat{\theta}_n) \hat{A}_n^{-1} \hat{B}_n \hat{A}_n^{-1} \nabla s(\hat{\theta}_n)]^{-1} s(\hat{\theta}_n) \overset{A}{\sim} \chi^2_r$$

under the null if A.1-A.7 hold, where

$$\hat{A}_n^{-1} \hat{B}_n = A_n(\hat{\theta}_n)^{-1} B_n(\hat{\theta}_n)$$

Models, by definition, are approximation to "reality". Misspecification problems, then, are serious ones. The true model is often unknown and the residual component could have other characteristics rather than the i.i.d..

Vigorous research is being carry on by White (1980,1981), Domowitz and White (1982), Bierens (1981, 1982) to mention a few.

Test of misspecification of wrong model, alternative characteristics for e_t and combinations had been proposed. With more experience and further research more stronger results and easier computationally statistics would appear in the near future, we hope.

Appendix

θ_n is a weak consistent estimator of θ_o, if for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P[\,|\theta_n - \theta_o| \geq \varepsilon\,] = 0$$

and we write

$$\text{plim}\ \theta_n = \theta_o$$

θ_n is a strong consistent estimator of θ_o, if

$$P[\lim_{n\to\infty} \theta_n = \theta_o] = 1$$

and we write

$$\theta_n \longrightarrow \theta_o \quad \text{a.s.}$$

A sequence of functions $\{g_\nu\}$ converge pointwise to a function g with domain Ω if the sequence $\{g_\nu(w)\}$ converge to g(w) for each $w \in \Omega$.

A sequence of functions $\{g_\nu(\theta,w)\}$ defined on $\textcircled{H} \times \Omega$ converge almost surely uniformly on $\textcircled{H}$ to $g(\theta)$, if there is a null set N and a number $n_o(w,\varepsilon)$ such that for every $\varepsilon > 0$ and every $w \in \Omega - N$,

$$\sup_{\theta \in \textcircled{H}} |g_\nu(\theta,w) - g(\theta,w)| \leq \varepsilon \qquad \text{if } \nu \geq n_0(w,\varepsilon)$$

Kolmogorov strong law of large numbers (Loève,volI,pg 251)

If the independent random variables X_n are identically distributed with a commom law $\mathcal{L}(X)$, then $1/n \sum_{i=1}^{n} X_i \longrightarrow c$ a.s., where c is finite, if, and only if, $E|X| < \infty$; and then $c=E(X)$.

Lindeberg-Lévy central limit theorem

Let $\{x_t\}$ be a sequence of i.i.d. random vectors with mean μ and covariance matrix Σ. Then $n^{-1/2} \sum_{t=1}^{n} (x_t - \mu)$ converges in distribution to the normal law $\mathcal{L}(0,\Sigma)$.

References:

Amemiya, T., Regression analysis when the dependent variable is truncated normal, Econometrica, 1973, 41, 997-1016

-------, The nonlinear two-stage Least Squares estimator, Jornal of Econometrics, 1974, 2, 105-110

-------, The maximum likelihood estimator and the nonlinear three-stage least squares estimator in the general nonlinear simu-taneous equation model Econometrica, 1977, 45, 955-968

-------, Nonlinear regression models, Technical report nº 330, 1981, Institute for Mathematical Studies in the Social Sciences, Stanford University Stanford, CA.

Barnett,W.A., Maximum likelihood and iterated Aitken estimation of nonlinear systems of equations, Journal of ths American Statiscal Association, 1976, 71, 354-360

Beauchamp,J.J. and R.G. Cornell, Simultaneous nonlinear estimation, Technometrics 1966, 8, 319-326

Belsley, D.A., On the computational competitiveness of ful-information maximum likelihood and three-stage least squares in the estimation of nonli near, simultaneous equation models, Journal of Econometrics, 1979,9, 315-342

-------, On the efficient computation of the nonlinear ful-information maximum likelihood estimation, Journal of Econometrics, 1980,14,203-225

Berndt,E.R., B.H. Hall, R.E. Hall, and J.A. Hausman, Estimation and inference in nonlinear structural models, Annals of Economic and Social Measurement, 1974, 3, 653-665

Bierens, H.J., Robust methods and asymptotic theory in nonlinear econometrics, Lecture notes in economics and mathematical systems, vol. 192, 1981 Spring-Verlag, Heidelberg

-------, Consistent model specification tests, Journal of Econometrics, 1982, 20, 105-134

Burguete, J.F., A.R. Gallant, and G. Souza, On unification of the asymptotic theory of nonlinear econometric models, Econometric Reviews, 1982, 1, 151-190

Chow, G.C., On the computation of ful information maximum likelihood estimates for nonlinear equation systems, Review of Economics and Statistics, 1973, 55, 104-109

Domowitz,I and H. White, Misspecified models with dependent observations, Journal of Econometrics,1982,20,35-58

Eisenpress, H. and J. Grennstad, The estimation of nonlinear econometric systems, Econometrica,1966,44,851-861

Gallant,A.R., Inference for nonlinear models,1973,Institute of Statistics mimeograph series n2875(North Carolina State University, Raleigh,NC)

-------, Seemingly unrelated nonolinear aegressions, Journal of Econometrics, 1975,3,35-50

---, and D.W. Jorgenson, Statistical inference for a system of simultaneous, nonlinear, implicit equations in the context of instrumental variable estimation, Journal of Econometrics,1979,11,275-302

Hannan, E.J., Nonlinear time series regression, Journal of Applied Probability, 1971,8,767-780

Hartley, H.O., The modified Gauss-Newton method for the fitting of nonlinear regression functions by least squares, Technometrics, 1961,3,269-280

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

Loève,M., Probability Theory, vol.1, 1977, 4th ed., Springer-Verlag, Heidelberg

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

Phillips, P.C.B. The iterated minimum distance estimator and the quasi-maximum likelihood estimator, Econometrica,1976,44,449-460

Robinson, P.M., Non-linear regression for multiple time series, Journal of Applied Probability,1972,9,758-768

Rudin, W., Principles of Mathematical Analysis, 2nd ed., McGraw-Hill,New York,1962

White, H., Nonlinear regression on cross-section data, Econometrica,1980,48,721-746

-------, Consequences and detection of misspecified nonlinear regression models, Journal of the American Statistical Association,1981,76,419-433

-------, Maximum likelihood estimation of misspecified models, Econometrica,1982, 50,1-25

Wu, C.-F., Asymptotic theory of nonlinear least squares estimation, Annals of Statistics,1981,9,501-513