Cowles Foundation Paper 549

Econometrica, Vol. 50, No. 5 (September, 1982)

ON THE CONSISTENCY OF NONLINEAR FIML

BY P. C. B. PHILLIPS[1]

Examples are given which show that normality is not necessary for the consistency of the quasi-maximum likelihood estimator in the nonlinear simultaneous equations model (nonlinear FIML) even when there are major departures from linearity. A possibility theorem is proved which demonstrates that when nonlinear FIML is consistent under normality it is always possible to find a nonnormal error distribution for which the consistency of nonlinear FIML is maintained. The procedure that is developed for finding a class of error distributions which preserve the consistency of nonlinear FIML can be applied more generally and may be useful in other contexts.

1. INTRODUCTION

RECENT THEORETICAL WORK on the nonlinear simultaneous equations model seems to have emphasized the importance of the normality assumption and, more generally, correct distributional assumptions about the equation errors in establishing the consistency of the nonlinear full information maximum likelihood (FIML) estimator. For example, in [2], Amemiya argued that the proof of consistency depends crucially on the assumption of normality of the error term; and, in [3] Amemiya stated that if the true distribution of the error term is not normal then nonlinear FIML[2] is not even consistent. In this respect, the general nonlinear model appears very different from the linear simultaneous equations model, where it is known that the consistency of FIML based on the hypothesis of normally distributed errors is maintained for a wide class of alternative error distributions.

As a result, it now appears to be a fairly common belief in the profession that normality of the errors is necessary for the consistency of nonlinear FIML. Some authors have been led to act on this belief in applied work. For example, Fair and Parke [5] have recently proposed the Hausman [10] specification test to test the hypothesis that the errors are normally distributed by comparing the nonlinear FIML and three stage least squares (3SLS) estimates. This test might be appropriate in a nonlinear model if the FIML estimates were, indeed, inconsistent and the 3SLS estimates consistent when the errors on the equations were not normally distributed but belonged to a certain wider class of distributions. However, such a result has not actually been proved in the literature.

In addition to the common belief that normality is necessary for the consistency of nonlinear FIML, there appears to be another widely held view, albeit rather loosely expressed, that consistency of FIML in nonlinear models is the

[1] I am grateful to William Brainard for helpful discussions and the referees and Kenneth Wallis for their comments. My thanks to Glena Ames for her skill and effort in preparing the typescript. The research reported here was supported by the National Science Foundation under Grant Number SES 800 7571.

[2] Throughout the rest of this paper we will use the term nonlinear FIML to describe the estimator obtained by maximizing what would be the likelihood if the normality assumption were correct.

exception rather than the rule when the likelihood is misspecified.[3] This is a view on which the statistical literature does not seem to provide complete guidance. It has, on the one hand, been argued [**15**] that the cases in which maximum likelihood (ML) estimates have been proved to have good properties are extremely restricted; while elsewhere [**11**] it has been shown that the true distribution underlying the observations does not need to belong to the parametric family defining the ML estimator for the estimator to be consistent. These results are not contradictory. The ML estimator can be consistent for a family of distributions other than the true distribution, while the overall class for which consistency obtains can itself be relatively narrow. The extent of this latter class will depend on the form of the data generating process and on other components of the model such as the assumed behavior of the exogenous variable sequence. At present, the nature of this dependence is relatively unexplored. But it is, in my opinion, going too far to assert that consistency of FIML in nonlinear models is the exception rather than the rule. For, while it is recognized that the linear simultaneous equations model is a very important exception, any major departure from linearity, such as the presence of levels and logarithms of the same endogenous variable in the model, is thought by many to put us in a different theoretical arena where normality of the errors or the correct specification of the likelihood becomes critical for the consistency of FIML. One aim of the present paper is to argue that this is not the case. We take an example which does involve both levels and logarithms of the variables and illustrate a procedure for finding an alternative class of error distributions other than the normal for which a nonlinear FIML is consistent. This and other examples are discussed in Section 3 of the paper and, since the type of nonlinearity considered is common in applied work, it should be of some relevance in practical econometric work where such nonlinearities in the variables actually occur.

The fact that a specific example has been used in Section 3 naturally leads to the following questions: (a) how important is this example to an applied investigator who is working with what might be a complicated nonlinear model and who wants to know if normality is, indeed, necessary for nonlinear FIML to be consistent in his model; and (b) how generally applicable is the procedure for finding nonnormal error distributions that is illustrated in Section 3. In an attempt to resolve some of these issues, a Possibility Theorem is given in Section 4 which demonstrates essentially that, when nonlinear FIML is consistent under normality, it is always possible to find a nonnormal error distribution for which the consistency of nonlinear FIML is maintained.

2. ON THE GRADIENT OF THE LOG-LIKELIHOOD

As in [**2**] we write the nonlinear simultaneous equations model in the form

$$f_i(y_t, x_t, \alpha_i) = u_{it} \qquad (i = 1, \dots, n) \tag{1}$$

[3] Such a view was expressed by one referee of a preliminary version of Section 2 of this paper.

where y_t is an $n \times 1$ vector of endogenous variables, x_t is a vector of exogenous variables, and α_i is a vector of parameters. The disturbance vector $u_t = (u_{it})$ is assumed to be independent and identically distributed $N(0, \Sigma)$ where Σ is a positive definite matrix.

In deriving the asymptotic properties of the maximum likelihood estimators of the parameters in (1), Amemiya makes extensive use of the following lemma:

LEMMA: *If $u_1, \ldots, u_n$ are jointly normal with mean zero and covariance matrix (σ_{ij}) and $h(u_1, \ldots, u_n)$ is such that $E(h)$ and $E(\partial h/\partial u_i)$ are finite, then $E(\partial h/\partial u_i) = E(h\sum_{j=1}^{n}\sigma^{ij}u_j)$.*

This lemma is used to establish that there is a consistent root of the likelihood equation, provided a number of other more usual assumptions are made concerning the existence and nature of convergence of certain summations that appear in the likelihood and its first two derivatives. In particular, if L denotes the log likelihood function concentrated with respect to the α_i $(i = 1, \ldots, n)$, then we have in the notation of [**2**, equation (3.8)],

$$(2)\qquad T^{-1}\frac{\partial L}{\partial \alpha_i}\bigg|_{\alpha_0} = T^{-1}\sum_{t=1}^{T}\left[\frac{\partial g_{it}}{\partial u_{it}} - g_{it}u_t'\sigma^i\right] - \left(T^{-1}\sum_{t=1}^{T} g_{it}u_t'\right)\left[\left(T^{-1}\sum_{t=1}^{T} u_t u_t'\right)_i^{-1} - \sigma^i\right]$$

where σ^i is the ith column of Σ^{-1}, $(\ )_i^{-1}$ denotes the ith column of the inverse of the matrix within the bracket, $g_{it} = \partial f_{it}/\partial \alpha_i$ and α_0 is the true value of $\alpha = (\alpha_i)$. When the conclusion of the Lemma holds, we deduce from (2) that the mean of the first term on the right side of (2) is zero. It then follows by elementary arguments that $\text{plim}_{T\to\infty}[T^{-1}\partial L/\partial \alpha_i]_{\alpha_0} = 0$. This conclusion is of vital importance in establishing that there is a weakly consistent root of the likelihood equation.

Similar arguments apply in the case of a misspecified likelihood function. In this case the conclusion of the Lemma can be replaced by the direct condition that the expectation of the gradient of the misspecified likelihood vanishes when it is evaluated at the true values of the parameters and when the expectation is taken with respect to the true distribution. Such a condition has been assumed by Huber [**11**] and Inagaki [**12**] in their discussion of consistency in misspecified situations. Their approach has already been mentioned in the econometrics literature by Hatanaka [**9**]. An alternative approach which involves no differentiability assumptions and which extends Wald's consistency proof [**20**] beyond the correctly specified case has also been developed by Huber in [**11**]. This approach relies on assumptions which ensure that a suitably standardized objective function (on which the estimator is based by optimization) has a limit with a unique optimum at the true value of the parameters. Related work in a general framework for nonlinear econometric models has recently been done by

Burguete, Gallant, and Souza [4]. These authors require, for the purpose of identifying a certain point in their estimation space, that at this point the limit of their objective criterion has a unique optimum. In Sections 3 and 4 below we consider the extent to which the conclusion of the Lemma continues to hold as we move away from the normality hypothesis. A similar approach can be taken with regard to the results of [4] by considering the extent to which the unique optimum property of the limit of the objective function is retained as we move away from the assumed error distribution on which that limit is based.

Some difficulties arise in the proof of the Lemma which appeared in [2]. We can illustrate with the scalar case in which $n = 1$ and $\phi(u)$ represents the normal density. The conclusion of the Lemma is now derived [2, equation (3.7)] from the equation $\int_a^b \{d(h\phi)/du\}\,du = [h\phi]_a^b$ by allowing the limits $b \to \infty$ and $a \to -\infty$. For this to be valid we need $h\phi$ to be absolutely continuous over $[-\infty, \infty]$. If we also assume that $E|h'(u)|$ is finite, these conditions are sufficient to ensure that $h(u)\phi(u) \to 0$ as $|u| \to \infty$ and the formula $E(h') = \sigma^{-2}E(hu)$ is valid. A complete statement and proof of this Lemma has recently been published by Stein [18].[4]

The fact that the restrictions on the class of allowable h functions for the validity of the lemma ensure that $h(u)\phi(u)$ tends to zero as $|u| \to \infty$ has a meaningful interpretation in terms of the model (1). In using the lemma to develop an asymptotic theory for the model (1) we want to be able to set $h(u)$ equal to the functions that appear in the structural specification and their derivatives. The fact that $h(u)\phi(u)$ tends to zero as $|u| \to \infty$ means that the maximum allowable growth of these functions as u becomes large is controlled by the rate at which the probability density of the error decays to zero in the tails. This condition therefore moderates the influence of outliers in the error distribution on the behavior of the endogenous variables. Such moderation would seem to be necessary in the development of an asymptotic theory in which certain summations involving the structural functions and their derivatives are assumed to converge in some stochastic sense as the sample size grows large. For, if the structural functions take on values which become very large for certain realizations of the errors relative to the probability that the errors actually assume these realizations, then the summations which involve these structural functions may not converge as the sample size tends to infinity. It is apparent that this problem is also relevant to estimators other than nonlinear FIML. The problem also arises when we relax the assumptions that the errors are normally distributed and follow instead a law which gives a different and possibly greater probability to outliers. It will be referred to again in the next section.

3. AN EXAMPLE IN WHICH NONLINEAR FIML IS CONSISTENT WITH NONNORMAL ERRORS

It is not difficult to construct models with minor departures from linearity in which nonlinear FIML is consistent for a wide class of error distributions. One

[4] In Lemma 1 of [18] Stein requires h to be an indefinite integral of h'. This is equivalent to requiring h to be absolutely continuous (AC) and since ϕ is AC the product $h\phi$ is AC also [19, p. 375], as required above. A multivariate version of this result is proved in Lemma 2 of [18].

such example is given (although this feature of the example is not discussed) by Malinvaud [**16**] (see, in particular, page 732); another by Phillips and Wickens [**17**, problem and solution 6.22]. The following example involves what may be regarded as a major departure from linearity and is based on an example used by Gallant [**6**] and Gallant and Holly [**7**] to illustrate the verification of the conditions they used in the development of an asymptotic theory for nonlinear FIML and 3SLS.

The structural model is

$$\ln y_{1t} + a_1 = u_{1t}, \tag{3}$$

$$y_{2t} + b_1 y_{1t} = u_{2t}, \tag{4}$$

and its reduced form

$$y_{1t} = e^{-a_1 + u_{1t}}, \tag{5}$$

$$y_{2t} = -b_1 e^{-a_1 + u_{1t}} + u_{2t}. \tag{6}$$

The concentrated log likelihood (or quasi likelihood) function is

$$\begin{aligned} L(a_1, b_1) &= -\frac{T}{2} \ln\left[\left\{T^{-1}\sum_t (\ln y_{1t} + a_1)^2\right\}\left\{T^{-1}\sum_t (y_{2t} + b_1 y_{1t})^2\right\}\right. \\ &\qquad \left. - \left\{T^{-1}\sum_t (y_{2t} + b_1 y_{1t})(\ln y_{1t} + a_1)\right\}^2\right] \\ &= -\frac{T}{2} \ln A_T(a_1, b_1) \qquad \text{say} \end{aligned} \tag{7}$$

and its first derivatives

$$\begin{aligned} \frac{\partial L}{\partial a_1} = -\frac{1}{2} T A_T^{-1} &\left[\left\{2T^{-1}\sum_t (\ln y_{1t} + a_1)\right\}\left\{T^{-1}\sum_t (y_{1t} + b_1 y_{1t})^2\right\}\right. \\ &- 2\left\{T^{-1}\sum_t (y_{2t} + b_1 y_{1t})\right\} \\ &\left. \times \left\{T^{-1}\sum_t (y_{2t} + b_1 y_{1t})(\ln y_{1t} + a_1)\right\}\right], \end{aligned} \tag{8}$$

$$\begin{aligned} \frac{\partial L}{\partial b_1} = -\frac{1}{2} T A_T^{-1} &\left[\left\{T^{-1}\sum_t (\ln y_{1t} + a_1)^2\right\}\left\{2T^{-1}\sum_t (y_{2t} + b_1 y_{1t}) y_{1t}\right\}\right. \\ &- 2\left\{T^{-1}\sum_t (y_{2t} + b_1 y_{1t})(\ln y_{1t} + a_1)\right\} \\ &\left. \times \left\{T^{-1}\sum_t y_{1t}(\ln y_{1t} + a_1)\right\}\right]. \end{aligned} \tag{9}$$

It follows that at the true values a_1^0, b_1^0 of the parameters we have

(10) $$\operatorname*{plim}_{T\to\infty} T^{-1}\partial L(a_1^0, b_1^0)/\partial a_1 = 0,$$

(11) $$\operatorname*{plim}_{T\to\infty} T^{-1}\partial L(a_1^0, b_1^0)/\partial b_1$$
$$= -\frac{1}{2}A^{-1}\left[a\sigma_{11}E(u_{2t}e^{-a_1+u_{1t}}) - 2\sigma_{12}E(u_{1t}e^{-a_1+u_{1t}})\right],$$

where the disturbance vector $u_t = (u_{it})$ is assumed, as in (1), to be independent and identically distributed with zero mean and covariance matrix $\Sigma = (\sigma_{ij})$ for all t; but not necessarily normal. For the expectations in (11) to be finite, we also require that the moment generating function of u_t,

(12) $$\mathrm{mgf}(s) = E(e^{s'u_t}),$$

exist for certain nonzero values of the vector $s' = (s_1, s_2)$. A precise region within which we will require (12) to exist will be specified later. In (11) A is given by $A = \mathrm{plim}_{T\to\infty} A_T(a_1^0, b_1^0) = \det\Sigma$. Note that (10) and (11) can also be derived from (2) by setting $g_1 = 1$ and $g_2 = e^{-a_1+u_1}$.

To prove that nonlinear FIML applied to (3) and (4) gives consistent estimates we set $\alpha' = (a_1, b_1)$ and it will be sufficient to show that the following two conditions hold:

(13i) $$\operatorname*{plim}_{T\to\infty}\left(T^{-1}\partial L(a_1^0, b_1^0)/\partial\alpha\right) = 0; \quad \text{and}$$

(13ii) $$\operatorname*{plim}_{T\to\infty}\left(T^{-1}\partial L(a_1^0, b_1^0)/\partial\alpha\partial\alpha'\right) \quad \text{is negative definite.}$$

If (13i) and (13ii) hold and the convergence in (13ii) is uniform in a neighborhood of (a_1^0, b_1^0), then it follows from the argument in the Appendix of [2] that nonlinear FIML is consistent.[5]

To establish (13i) and (13ii) we need to make explicit distributional assumptions so that the expectations that appear in the limits can be evaluated. It is easy to verify (13) when u_t is multivariate normal. Since we are interested in specifying a nonnormal error distribution for which (13) continue to hold, a convenient point of departure is to specify a class, such as the following mixtures of the multivariate normal, which includes the normal as a special case. Specifically, we consider the class of probability densities given by

(14) $$\mathrm{pdf}(u_t) = \int_0^\infty (2\pi w)^{-1}(\det\tilde{\Sigma})^{-1/2}\exp\left(-\frac{1}{2}u_t'\tilde{\Sigma}^{-1}u_t/w\right)dG(w)$$

[5] More accurately, this approach establishes that there is a consistent root of the likelihood equation We recognize the validity of Wald's argument in **[20]** that this does not strictly imply that the root corresponding to nonlinear FIML is consistent. An alternative approach is possible based on the line of argument in **[20]** and involves extensions of the work in **[11** and **4]**. It has already been suggested in Section 2

where $G(w)$ is a distribution function supported on the half line $[0,\infty)$ and $\tilde{\Sigma}=(\tilde{\sigma}_{ij})$ is a positive definite matrix. One immediate restriction on $G(w)$ is that the moment generating function (12) exist and since

$$\text{mgf}(s)=E(e^{s'u_t})=\int_0^\infty e^{ws'\tilde{\Sigma}s/2}\,dG(w) \tag{15}$$

we have

$$\partial\,\text{mgf}(s)/\partial s=\left\{\int_0^\infty we^{ws'\tilde{\Sigma}s/2}\,dG(w)\right\}\tilde{\Sigma}s \tag{16}$$

and

$$\partial^2\text{mgf}(s)/\partial s\partial s'=\left\{\int_0^\infty we^{ws'\tilde{\Sigma}s/2}\,dG(w)\right\}\tilde{\Sigma} + \left\{\int_0^\infty w^2e^{ws'\tilde{\Sigma}s/2}\,dG(w)\right\}\tilde{\Sigma}ss'\tilde{\Sigma}. \tag{17}$$

We deduce that $E(u_t)=0$ and $E(u_tu_t')=\{\int_0^\infty w\,dG(w)\}\tilde{\Sigma}$. For compatibility with (1), we then require that

$$\left\{\int_0^\infty w\,dG(w)\right\}\tilde{\Sigma}=\Sigma. \tag{18}$$

Nonlinear FIML will now be consistent for every error distribution in the class (14) for which conditions (13) hold. From (11), we see for the present example that condition (13i) requires that

$$E(u_{1t}^2)E(u_{2t}e^{u_{1t}})-E(u_{1t}u_{2t})E(u_{1t}e^{u_{1t}})=0. \tag{19}$$

That is

$$\begin{aligned}(-\sigma_{12},\sigma_{11})\begin{pmatrix}\partial/\partial s_1\\ \partial/\partial s_2\end{pmatrix}\text{mgf}(s)\Big|_{\substack{s_1=1\\ s_2=0}} &=(-\sigma_{12},\sigma_{11})\tilde{\Sigma}\begin{pmatrix}1\\0\end{pmatrix}\left\{\int_0^\infty we^{w\tilde{\sigma}_{11}/2}\,dG(w)\right\}\\ &=(-\sigma_{12},\sigma_{11})\Sigma\begin{pmatrix}1\\0\end{pmatrix}\left\{\int_0^\infty we^{w\tilde{\sigma}_{11}/2}\,dG(w)\right\}\Big/\left\{\int_0^\infty w\,dG(w)\right\}\\ &=0\end{aligned} \tag{20}$$

provided $G(w)$ is such that the integral

$$\int_0^\infty we^{w\tilde{\sigma}_{11}/2}\,dG(w)<\infty. \tag{21}$$

Taking the restrictions (18) and (21) together we have a whole class of non-normal error distributions given by (14) for which condition (13i) holds. (13ii) imposes only a very mild additional restriction on $G(w)$ as we see in the special case below. This procedure therefore provides a very natural way of determining a class of nonnormal error distributions for which nonlinear FIML is consistent.

We end this section by illustrating a nontrivial nonnormal member of the class prescribed by (14), (18), and (21) for which nonlinear FIML is consistent. We take the continuous exponential distribution for $G(w)$ with density

$$(22) \qquad g(w) = G'(w) = \lambda e^{-\lambda w}, \qquad \lambda > 0.$$

We deduce from (14) that the corresponding density of the error vector u_t is

$$(23) \qquad \text{pdf}(u_t) = \lambda(2\pi)^{-1}(\det\tilde{\Sigma})^{-1/2}\int_0^{\infty} w^{-1}\exp\{-\lambda w - \beta(u_t)/w\}\,dw$$

where $\beta(u_t) = \frac{1}{2}u_t'\tilde{\Sigma}^{-1}u_t$. Let $v = \lambda w$ and transform variables in (23) to get

$$\begin{aligned}(24) \qquad & \lambda(2\pi)^{-1}(\det\tilde{\Sigma})^{-1/2}\int_0^{\infty} v^{-1}\exp\{-v - \lambda\beta(u_t)/v\}\,dv \\ &= \lambda(2\pi)^{-1}(\det\tilde{\Sigma})^{-1/2}2K_0\left(2\sqrt{\lambda\beta}\right) \\ &= \lambda\pi^{-1}(\det\tilde{\Sigma})^{-1/2}K_0\left(2\left(\frac{\lambda u_t'\tilde{\Sigma}^{-1}u_t}{2}\right)^{1/2}\right)\end{aligned}$$

where $K_0(z)$ is the modified Bessel function of the third kind.[6] We note from (18) and (22) that $\lambda^{-1}\tilde{\Sigma} = \Sigma$ so that we can write the density of u_t as

$$(25) \qquad \text{pdf}(u_t) = \pi^{-1}(\det\Sigma)^{-1/2}K_0\left(2\left(\frac{u_t'\Sigma^{-1}u_t}{2}\right)^{1/2}\right).$$

In the general case, where u_t is an $n \times 1$ vector the corresponding density is given by

$$\begin{aligned}(26) \qquad \text{pdf}(u_t) &= 2(2\pi)^{-n/2}(\det\Sigma)^{-1/2} \\ &\quad \times K_{n/2-1}\left(2\left(\frac{u_t'\Sigma^{-1}u_t}{2}\right)^{1/2}\right)\left(\left(\frac{u_t'\Sigma^{-1}u_t}{2}\right)^{1/2}\right)^{-n/2+1}\end{aligned}$$

and the moment generating function of u_t is simply

$$(27) \qquad \text{mgf}(s) = \left[1 - \tfrac{1}{2}s'\Sigma s\right]^{-1}.$$

[6] See, for example, Lebedev **[14]** and, in particular, equation (5.10.25) on page 119 of **[14]** for the representation of the integral in (24).

This can be viewed as a multivariate generalization of the Laplace distribution and we note that in the case $n = 1$ (26) reduces to

$$\text{(28)} \qquad \mathrm{pdf}(u_t) = (2\gamma)^{-1}e^{-|u|/\gamma}, \qquad \gamma^2 = \sigma_{11}/2,$$

which is the univariate Laplace with variance equal to $2\gamma^2 = \sigma_{11}$ and moment generating function equal to $[1 - \gamma^2 s^2]^{-1}$.

The tail behavior of the density (26) can be determined from the following asymptotic expansion of the function $K_\nu(z)$ for any real ν:

$$\text{(29)} \qquad K_\nu(z) \sim \left(\frac{\pi}{2z}\right)^{1/2} e^{-z}$$

as $|z| \to \infty$ [**12**, page 123]. We deduce from (26) and (29) that

$$\text{(30)} \qquad \mathrm{pdf}(u_t) \sim 2^{-n/2}\pi^{-(n-1)/2}(\det\Sigma)^{-1/2}\left(\tfrac{1}{2}u_t'\Sigma^{-1}u_t\right)^{-(n+1)/4+1/2} \cdot \exp\left\{-\left(2u_t'\Sigma^{-1}u_t\right)^{1/2}\right\}$$

as $\|u_t\| \to \infty$. Thus, although (26) has exponentially thin tails as described by (30) these tails are thicker than those of the multivariate normal distribution.

We now return to the verification of conditions (13) for the density (25). For (13i) to hold it remains only to check (21). We have

$$\int_0^\infty w e^{w\tilde{\sigma}_{11}/2}\, dG(w) = \lambda \int_0^\infty w e^{\lambda\sigma_{11}w/2} e^{-\lambda w}\, dw$$

which will be finite provided $\sigma_{11} < 2$. This condition can also be obtained directly from the moment generating function (27). Setting $s_2 = 0$ in (27) we require, for the moment generating function to exist,

$$\text{(31)} \qquad s_1^2\sigma_{11} < 2,$$

and if $E(e^{u_{1t}})$ is to be finite this requires $\sigma_{11} < 2$ as stated.[7,8] This verifies condition (13i).

For condition (13ii) to hold we need $\operatorname{plim}_{T\to\infty}\partial^2 L(a_1^0, b_1^0)/\partial\alpha\partial\alpha'$ to exist and be negative definite. Calculations show that the probability limit will exist provided

$$\text{(32)} \qquad E(e^{2u_{1t}}) < \infty,$$

[7] Note that this condition is also needed, at least as far as the proofs in [**1**] and [**6**] are concerned, for the consistency of the nonlinear 3SLS estimator as some simple manipulations will show

[8] This condition will also ensure that $e^{u_{1t}}\mathrm{pdf}(u_t) \to 0$ as $|u_{1t}| \to \infty$ as can be verified directly from the asymptotic form (30).

that is, from (31), provided $\sigma_{11} < 1/2$. If this restriction on $\operatorname{var}(u_{1t})$ holds we find

$$\begin{aligned}
&\operatorname*{plim}_{T\to\infty} \frac{\partial^2 L(a_1^0, b_1^0)}{\partial\alpha\partial\alpha'} \\
&\quad = -(\det\Sigma)^{-1} \\
&\qquad \times \left[\begin{array}{c|c} \sigma_{22} & \sigma_{12}e^{-a_1}E(e^{u_{1t}}) \\ \hline \sigma_{12}e^{-a_1}E(e^{u_{1t}}) & \sigma_{11}e^{-2a_1}E(e^{2u_{1t}}) - e^{-2a_1}(E(e^{u_{1t}}u_{1t}))^2 \end{array}\right]
\end{aligned}$$

$$= -(\det\Sigma)^{-1}\left[\begin{array}{c|c} \sigma_{22} & \dfrac{2\sigma_{12}e^{-a_1}}{2-\sigma_{11}} \\ \hline \dfrac{2\sigma_{12}e^{-a_1}}{2-\sigma_{11}} & e^{-2a_1}\left\{\dfrac{\sigma_{11}}{1-2\sigma_{11}} - \left(\dfrac{4\sigma_{11}}{(2-\sigma_{11})^2}\right)^2\right\} \end{array}\right] \tag{33}$$

which is certainly negative definite for a range of values of $\sigma_{11} < 1/2$. This verifies (13ii).

It follows that nonlinear FIML applied to (10) will be consistent for the nonnormal error density (25) provided $\sigma_{11} < 1/2$ and (33) is negative definite. The inequality constraint $\sigma_{11} < 1/2$ is really innocuous because it is just a necessary condition for $\operatorname{var}(y_{1t})$ to be finite when errors driving the equation system (3) and (4) follow the distribution (25). Moreover, this condition is also needed in justifying the asymptotic normality, although not the consistency, of nonlinear 3SLS according to the proofs of [**1** and **6**]. In this connection we may refer back to the final comment of Section 2.

The example just discussed may be regarded by some readers as highly specialized. We note, for example, that it has a recursive structure, only a single nonlinearity and there are no exogenous variables in the system. When any or all of these special features of the example are removed, rather different results might be expected. For example, does the class of nonnormal error distributions for which nonlinear FIML is consistent substantially contract as we move away from the simplicity of (3) and (4)? This question is difficult to answer in generality because much will depend on the form of the departures taken. The following two models, in which x_t is a truly exogenous variable, help to throw some light on the question.

Model A:
$$\begin{aligned} \ln y_{1t} + a_1 + a_2x_t &= u_{1t}, \\ y_{2t} + b_1y_{1t} + b_2x_t &= u_{2t}. \end{aligned}$$

Model B:
$$\begin{aligned} \ln y_{1t} + a_1\ln y_{3t} + a_2 &= u_{1t}, \\ y_{1t}y_{2t} + b_1y_{1t} + b_2x_t &= u_{2t}, \\ \ln y_{3t} + c_1\ln y_{1t} &= u_{3t}, \\ y_{4t} + d_1y_{2t} &= u_{4t}. \end{aligned}$$

Model A is simply the model of the above example with an additive exogenous input. Model B involves an additional nonlinearity in the second equation, an exogenous input, and the recursive structure has been removed. An analysis of these models along the same lines of the above example will show that nonlinear FIML is consistent for a similarly broad class of mixtures of normals. Only very weak supplementary conditions on the exogenous variables are required for this to be true; it will be sufficient, for instance, that the sample mean and variance of x_t converge to a constant and positive constant respectively as $T \to \infty$ and, in the case of Model A, that in addition, as $T \to \infty$, $T^{-1}\sum_{t=1}^{T} e^{-2a_2 x_t}$ converges to a finite positive constant.

An alternative way of addressing the question asked in the previous paragraph is to start with a general nonlinear system for which nonlinear FIML is known to be consistent and, without specifying functional forms, ask if it is possible to find nonnormal errors for which this consistency is preserved. This is the approach we take in the next section.

4 A POSSIBILITY THEOREM

As is clear from equation (2) and the working of the example in the previous section, the critical condition we need to verify for nonlinear FIML to be consistent is that the probability limit of the standardized gradient of the log likelihood is zero at the true values of the parameters being estimated (that is, condition (i) of Section 3). If the conclusion of the Lemma holds, viz. that $E[\partial g_{it}/\partial u_{it} - g_{it} u_t' \sigma^i] = 0$ where $g_{it} = \partial f_{it}/\partial \alpha_i$ and the f_{it} are the structural functions in (1), then condition (i) will usually follow directly from the law of large numbers for suitable exogenous sequences $\{x_t\}$. But, since the functions g_{it} are in general dependent on the exogenous variables x_t as well as the errors in the model (1) it is not in fact necessary that the conclusion of the lemma hold for each t in order that

(34) $$\operatorname*{plim}_{t \to \infty} T^{-1} \sum_{t=1}^{T} \xi_t = 0 \qquad \text{where} \qquad \xi_t = \xi_t(u_t, x_t; \alpha, \Sigma) = \partial g_{it}/\partial u_{it} - g_{it} u_t' \sigma^i .$$

It will instead be enough that realizations of the (possibly dependent) exogenous variable sequence $\{x_t\}$ and the i.i.d sequence $\{u_t\}$ be compatible with (34), as they will be if the following two conditions hold:

(a) $$T^{-1} \sum_{t=1}^{T} E(\xi_t) \to 0 \qquad \text{as} \qquad T \to \infty, \qquad \text{and}$$

(b) $$E\left[\left\{ T^2 + \left(\sum_{t=1}^{T} (\xi_t - E\xi_t) \right)^2 \right\}^{-1} \times \left\{ \sum_{t=1}^{T} (\xi_t - E\xi_t) \right\}^2 \right] \to 0 \qquad \text{as} \qquad T \to \infty$$

[**8**, p. 250]. In (a) and (b) above, expectations are now taken with respect to the joint probability distribution of (u_t, x_t). In order to explore the implications of these conditions in terms of an allowable class of error distributions it will be helpful to make more specific assumptions about the process generating the sequence $\{x_t\}$. Condition (a) is the essential requirement for consistency and, for certain classes of exogenous sequences $\{x_t\}$ such as when the x_t are i.i.d. from a certain probability distribution, it will indeed be sufficient that $E(\xi_t) = 0$ for all t, where the expectation is taken with respect to the joint probability distribution of (u_t, x_t). While this is far from being the most general situation, it is an important special case.

To provide a framework for our own discussion we start off by assuming that the $\{x_t\}$ form a random sample from a probability distribution with distribution function $F(x;\theta)$ on a Euclidean space X, where θ is a vector indexing a parametric family of possible probability distributions and that the sequence $\{x_t\}$ is truly exogenous. Then, by Theorem 2 of Jennrich [**13**], it follows that for almost all sequences $\{u_t, x_t\}$

$$T^{-1}\sum_{t=1}^{T}\xi_t \rightarrow \int_X dF(x;\theta)\int_u \xi(u,x;\alpha,\Sigma)\mathrm{pdf}(u)\,du \qquad \text{as} \quad T\rightarrow\infty \tag{35}$$

uniformly in α and Σ. For (35) to be valid as in [**13**] we require that $\xi(u,x;\alpha,\Sigma)$ be dominated by a function independent of the parameters and integrable with respect to the joint probability distribution of (u,x). We also require that ξ be a continuous function of the parameters for each (u,x) and a measurable function of (u,x) for each parameter value. An alternative approach is to assume as in Gallant [**6**] that the joint sequence $\{u_t, x_t\}$ generates Cesaro summable sequences with respect to the joint probability distribution on (u,x). Either approach leads to a representation of the limit as in (35) above. It is this case, where the limit representation is of the form (35) for a parametric family of distributions $\{F(x;\theta)\}$, that we now consider. We allow θ to lie in some subset Θ of R^q where q is the dimension of θ and require $F(x;\theta)$ to be a continuous function of $\theta \in \Theta$ for each x.

The possibility of finding nonnormal error distributions for which the consistency of nonlinear FIML is preserved is illustrated by the construction in the following theorem. To make the essential ideas behind the construction as clear as possible we confine ourselves to the single equation and single parameter case (i.e. $n = 1$ and $\alpha =$ scalar in (1)) and eliminate the corresponding subscripts on the variables and functions.

THEOREM (Possibility of Nonnormality and Consistent Nonlinear FIML): *Suppose the following conditions hold*: (i) $\{u_t\}$ *is i.i.d.* $(0,\sigma^2)$ *with probability density*

$$\mathrm{pdf}(u_t) = \int_0^\infty (2\pi w)^{-1/2}\tilde{\sigma}^{-1}\exp\{-u_t^2/2w\tilde{\sigma}^2\}\,dG(w) \tag{36}$$

where $G(w)$ is a distribution function supported on $[0,\infty)$, $\tilde{\sigma}^2>0$ *and* $\sigma^2 = \tilde{\sigma}^2\int_0^\infty wG(w)$.

(ii) $g(u_t,x_t)\mathrm{pdf}(u_t)$ *is absolutely continuous on* $[-\infty,\infty]$ *for each* x_t, *where* $g(u_t,x_t)=\partial f(y_t,x_t;\alpha)/\partial\alpha$.

(iii) $E_u(|g'|)<\infty$ *for each* x_t *where* E_u *indicates expectation with respect to* u_t *and* $g'=\partial g(u_t,x_t)/\partial u_t$ *is continuous on* $-\infty<u_t<\infty$.

(iv) *All summations that appear in the standardized (normal) log likelihood and its first derivative are assumed to converge almost surely and uniformly in the parameters. In particular*

$$T^{-1}\sum_{t=1}^{T}(g'-\sigma^{-2}gu_t)\rightarrow\int_X dF(x;\theta)\int_u(g'-\sigma^{-2}gu)\mathrm{pdf}(u)\,du \tag{37}$$

a.s. and uniformly in (α,σ^2) *as* $T\rightarrow\infty$ *where* $F(x;\theta)$ *is a parametric family of probability distributions on* X *indexed by* $\theta\in\Theta$ *just as in* (35) *above.*

(v) *Nonlinear FIML is consistent when the probability density* (36) *is normal (i.e. when* $G(w)=0,1;\ w<1,\ w\geq 1$).

Then, given θ, *there exists a nonnormal error distribution of the form* (36) *involving a mixing distribution* $G_\theta(w)$ *which depends on* θ *(and, in fact, an uncountable infinity of such distributions) for which the consistency of nonlinear FIML is preserved.*

PROOF: It follows from (2) and (iv) that

$$T^{-1}[\partial L/\partial\alpha]_{\alpha_0}\rightarrow\int_X dF(x;\theta)\int_u(g'-\sigma^{-2}gu)\mathrm{pdf}(u)\,du \qquad \text{a.s.} \tag{38}$$

as $T\rightarrow\infty$. Since $g(u)\mathrm{pdf}(u)$ is absolutely continuous, it follows that

$$\int_a^b g'\,\mathrm{pdf}(u)\,du+\int_a^b g\,\mathrm{pdf}'(u)\,du=[\,g\,\mathrm{pdf}(u)]_a^b \tag{39}$$

over any interval $[a,b]$. We let $a\rightarrow-\infty$ and $b\rightarrow\infty$ in (39). The first integral is finite by (iii) and it is easy to show that $[g\,\mathrm{pdf}(u)]_a^b\rightarrow 0$ as $b\rightarrow\infty$ and $a\rightarrow-\infty$. Hence, the second integral in (39) remains finite and we have

$$\begin{aligned}&\int_u g'\,\mathrm{pdf}(u)\,du\\ &\quad=-\int_u g\,\mathrm{pdf}'(u)\,du\\ &\quad=\sigma^{-2}\int_u gu\int_0^\infty(2\pi w)^{-1/2}\tilde{\sigma}^{-1}\exp\{-u^2/2w\tilde{\sigma}^2\}m_w w^{-1}\,dG(w)\end{aligned}$$

where $m_w = \int_0^\infty w\, dG(w)$. The limit of (38) can now be written as

$$\sigma^{-2}\int_x dF(x;\theta)\int_u gu\, du$$
$$\times \int_0^\infty (2\pi w)^{-1/2}\tilde{\sigma}^{-1}\exp\{-u^2/2w\tilde{\sigma}^2\}\left[m_w w^{-1}-1\right]dG(w)$$
$$= \sigma^{-3}m_w^{1/2}\int_0^\infty (2\pi w)^{-1/2}\left[m_w w^{-1}-1\right]dG(w)$$
$$\times \int_x dF(x;\theta)\int_u gu\exp\{-u^2/2w\tilde{\sigma}^2\}\, du$$
$$= \int_0^\infty (2\pi w)^{-1/2}\left[m_w w^{-1}-1\right]h(w;\alpha,\sigma^2,\theta)\, dG(w), \quad \text{say} \tag{40}$$

where h is a continuous function of w and of the parameters α, σ^2, θ.

We now require a mixing distribution $G(w)$ for which (40) is zero, so that (38) tends to a zero limit as $T \to \infty$. In the normal error case, this follows directly since we have $G(w) = 0, 1$ for $w < 1$, $w \geq 1$, and $m_w = 1$. To construct a nonnormal error distribution for which (40) is zero we can proceed as follows.

We take a discrete mixture of normals with mixing distribution given by

$$G(w) = \begin{cases} 0, & w < w_1, \\ \alpha, & w_1 \leq w < w_2, \\ 1-\epsilon+\alpha, & w_2 \leq w < w_3, \\ 1, & w_3 \leq w, \end{cases} \tag{41}$$

with $0 < \alpha < \epsilon$. Then as $\epsilon \to 0$ or as $(w_3 - w_1) \to 0$ the mixture density (36) approaches normality and (40) is zero because, in the limit, $w = m_w$ with probability one. We now wish to show that we can move away from this case of a degenerate mixture in a systematic way so that, for every mixing distribution of the form (41) on this path, the limit function (40) maintains the value zero and nonlinear FIML remains consistent within this class of nonnormal error densities.

Take $\eta > 0$ and suppose the mass points w_i lie respectively in the intervals $1-\eta < w_1 < 1$, $w_1 < w_2 < w_3$, $1 < w_3 < 1+\eta$. Suppose also that we select w_2 so that $m_w = \alpha w_1 + (1-\epsilon)w_2 + (\epsilon-\alpha)w_3 = 1$, which implies that we can write $w_2 = \{1 - [\alpha w_1 + (\epsilon - \alpha)w_3]\}/(1-\epsilon)$. For any $\epsilon^* > 0$, however small, we can select $\epsilon_a > 0$ in such a way that for all $0 < \epsilon \leq \epsilon_a$ we have $|w_2 - 1| < \epsilon^*$. Given $w_1 < 1$ and $w_2 > 1$ we can also select $\epsilon_b > 0$ so that for all $0 < \epsilon \leq \epsilon_b$ we have $w_1 < w_2 < w_3$. In what follows, we take $\epsilon \leq \epsilon_c = \min(\epsilon_a, \epsilon_b)$. The limit (40) is now

$$(2\pi)^{-1/2}\alpha h(w_1)(w_1^{-1}-1)w_1^{-1/2}$$
$$+ (2\pi)^{-1/2}(1-\epsilon)h(w_2)(w_2^{-1}-1)w_2^{-1/2}$$
$$+ (2\pi)^{-1/2}(\epsilon-\alpha)h(w_3)(w_3^{-1}-1)w_3^{-1/2}. \tag{42}$$

Let us assume that $h(w;\alpha,\sigma^2,\theta) \neq 0$ for $1-\eta \leq w \leq 1+\eta$ and in some neighborhood of the true values of the parameters (α,σ^2,θ). If, on the other hand, $h(w)$ were zero in some interval of $[0,\infty)$ then the construction of a mixing distribution $G(w)$ for which (40) were zero would become straightforward. Note that in the example of Section 3 $h(w)$ is identically zero, since from (20)

$$h(w) = (-\sigma_{12}\sigma_{11})\Sigma\begin{pmatrix}1\\0\end{pmatrix} we^{w\bar{\sigma}_{11}/2} m_w^{-1}$$

$$\equiv 0$$

for all w on $[0,\infty)$.

(42) will be zero if we select α in such a way that

$$\text{(43)} \qquad \alpha\left[h(w_1)(w_1^{-1}-1)w_1^{-1/2} - h(w_3)(w_3^{-1}-1)w_3^{-1/2}\right]$$

$$= -\epsilon h(w_3)(w_3^{-1}-1)w_3^{-1/2} - (1-\epsilon)h(w_2)(w_2^{-1}-1)w_2^{-1/2}.$$

It remains to show that if α is the solution of (43) for certain $w_1 < 1$ and $w_3 > 1$ then $0 < \alpha < \epsilon$, as required for $G(w)$. Suppose $h(1) > 0$ (the case $h(1) < 0$ can be dealt with in the same way). By continuity, we can select $\eta > 0$ so that the coefficient of α in square brackets on the left side of (43) is positive. The second term on the right side of (43) can be made arbitrarily small since $|w_2 - 1| < \epsilon^*$ for all $\epsilon \leq \epsilon_c$. The first term on the right side is positive and the modulus of the coefficient of ϵ in this term is less than the coefficient of α. It follows that we can select $\eta > 0$ and $\epsilon > 0$ in such a way that for any w_1 and w_3 satisfying $1-\eta < w_1 < 1 < w_3 < 1+\eta$ there exists a unique α satisfying (43) for which $0 < \alpha < \epsilon$. A mixing distribution $G(w)$ constructed according to (41) with these mass points and weights will annihilate the limit function (40). Note that (43) gives us not one but an infinite class of mixing distributions $G(w)$ for which (40) is annihilated. As w_1 and w_2 move away from unity (43) will determine the path of mixtures of normals for which the limit of the gradient of the (normal) log likelihood is zero at the true values of the parameters. We might also note that to the extent that h is a continuous function of θ we can regard α in (43) and hence $G(w)$ as dependent on θ. Thus we have an implied parametric family $G_\theta(w)$ of mixing distributions corresponding to the family of exogenous variable distributions $\{F(x;\theta)\}$.

Finally, we observe that by (v) nonlinear FIML is consistent when the error density (36) is normal. It follows that the limit of the standardized log likelihood has a global maximum at the true values of the parameters. When the error density is of the form (36) with the mixing distribution $G_\theta(w)$ constructed as above it follows by continuity that the limit of the standardized (quasi-) log likelihood retains a global maximum at the true values of the parameters. This proves that there exists a family of nonnormal error densities of the form (36) involving a mixing distribution $G_\theta(w)$ constructed as above for which the consistency of nonlinear FIML is preserved. □

Generalization of the construction in the proof of the Theorem to cover the multi-equation, multi-parameter case will involve the use of multiple mass points in the discrete mixture of normals, with the number of points required depending on the number of parameters and equations. The essential ideas behind the construction stay the same however and we end up with a system of linear equations in place of (43). Other more general methods of generating nonnormal error distributions could have been used and will be necessary if we are to investigate the class of nonnormal errors for which nonlinear FIML is consistent rather than simply establish the possibility as we have done here.

5. FINAL REMARKS

The examples in Section 3 of the paper show that normality is not necessary for the consistency of nonlinear FIML even when there are major nonlinearities in the structural functions and the analysis suggests a general procedure for constructing nonnormal error distributions for which the consistency of nonlinear FIML is maintained. The analysis also demonstrates the intimate relationship that exists between the form of the nonlinear functions admitted into the structural specification of the model and the tail behavior of the error distribution which is permissible if an asymptotic theory is to be developed. This compatibility between the nonlinearities in the structure and the probability of outliers in the error distribution prevents the influence of outliers interfering with the operation of the law of large numbers and is, in large part, independent of the estimation technique that is being used. In the main example of Section 3, the presence of both a level and a logarithm of an endogenous variable in the model is seen to substantially curtail the class of allowable error distributions. In particular, if the mean and variance of this endogenous variable are to exist, which in many cases will turn out to be near minimal conditions for an asymptotic theory to be developed for many estimators, not just FIML, we find ourselves already confined to an error distribution with exponentially thin tails. The example further demonstrates that nonlinear FIML will indeed be consistent for a Laplace-type error density with exponential tails. In this case, therefore, the predominant influence in curtailing the class of allowable error densities turns out to be the form of the nonlinearities in the structural functions. This is likely to be the case in many nonlinear econometric models where both levels and logarithms of the same variables appear in different parts of the equation system. This is not to say, however, that nonlinear FIML and 3SLS will be consistent for an identical class of error distributions in the main example of Section 3 or for the other models. On the contrary, the asymptotic properties of the nonlinear 3SLS procedure are likely to be more robust than those of nonlinear FIML. The extent to which this is so deserves further investigation.

The approach I have taken to the subject of nonlinear FIML in this paper is rather different from the existing literature. In part, this has been motivated by my unease over the common belief that normality is necessary for the consistency of nonlinear FIML. Intuition suggests that if nonlinear FIML is consistent for a given model and a given constellation of exogenous variable sequences under normally distributed errors then the same should be true for certain departures from normality, at least locally. This is now established by the Possibility Theorem in Section 4. As we increase the complexity of the model, we can expect a narrowing in the range of allowable alternatives and those departures from normality which are allowable can be expected to sustain a corresponding increase in their own complexity. This approach, of course, invites the critique that the allowable nonnormal alternatives are model and exogenous variable dependent and as such might be regarded as artificial or even pathological. While there is some force to this argument, it is worth bearing in mind that compatibility between the nonlinearities in the structure and the distribution of the errors is always necessary, as I have pointed out in the last paragraph, in order to prevent the tail behavior of the errors interfering with the operation of the law of large numbers. The necessity of such compatibility is by itself a major force in narrowing down a feasible class of error distributions, as the examples in Section 3 demonstrate.

This paper has left untouched a host of interesting problems associated with the asymptotic distribution of nonlinear FIML and other estimators under plausible alternative error distributions such as the normal mixture (14). One relevant aspect of the asymptotic distribution of nonlinear FIML is that although the Possibility Theorem shows that the limit of the gradient $T^{-1}\partial L(\alpha_0)/\partial\alpha$ is zero at the true parameter value for a certain class of nonnormal errors it does not necessarily follow from the construction that $-T^{-1}\partial^2 L(\alpha^0)/\partial\alpha\partial\alpha'$ and $T^{-1}(\partial L(\alpha^0)/\partial\alpha)(\partial L(\alpha^0)/\partial\alpha')$ converge in probability to the same limit. On the other hand, the construction could be made in such a way that there was equivalence, but this equivalence would then apply only to (what may then be) a narrower class of error distributions. This will have an important bearing on the suitability of the formulae employed for the asymptotic covariance matrix of nonlinear FIML. Moreover, when the likelihood is misspecified, the usual asymptotic efficiency ordering of nonlinear FIML and 3SLS no longer necessarily holds under the true distribution. These issues deserve further investigation before the nature of the Hausman specification test in the context suggested by Fair and Parke [**6**] is fully understood.

Yale University

Manuscript received August, 1980; revision received September, 1981.

REFERENCES

[1] AMEMIYA, T.: "The Nonlinear Two Stage Least Squares Estimator," *Journal of Econometrics*, 2(1974), 105–110.

[2] ———: "The Maximum Likelihood and Non-linear Three Stage Least Squares Estimator in the General Non-linear Simultaneous Equations Model," *Econometrica*, 45(1977), 955–968.

[3] ———: "L'Estimation Des Modèles a Équations Simultanées Non Linéaires," in *Cahiers Du Seminaire D'Econométrie*, ed. by E. Malinvaud. No 19, 1978.

[4] BURGUETE, J F., A. R. GALLANT, AND G. SOUZA: "On Unification of the Asymptotic Theory of Nonlinear Econometric Models," *Econometric Reviews* (forthcoming).

[5] FAIR, R. C., AND W. R. PARKE: "Full Information Estimates of a Non-Linear Macroeconometric Model," *Journal of Econometrics*, 13(1980), 269–291.

[6] GALLANT, A. R.: "Three Stage Least Squares Estimation for a System of Simultaneous Nonlinear Implicit Equations," *Journal of Econometrics*, 5(1977), 71–88.

[7] GALLANT, A. R. AND A. HOLLY: "Statistical Inference in an Implicit Non-linear Simultaneous Equations Model in the Context of Maximum Likelihood Estimation," *Econometrica*, 48(1980), 697–720.

[8] GNEDENKO, B. V.: *The Theory of Probability*. New York: Chelsea, 1962.

[9] HATANAKA, M.: "On the Efficient Estimation Methods for the Macro Economic Models Non linear in Variables," *Journal of Econometrics*, 8(1978), 323–356.

[10] HAUSMAN, J. A.: "Specification Tests in Econometrics," *Econometrica*, 46(1978), 1251–1271.

[11] HUBER, P. J.: "The Behaviour of Maximum Likelihood Estimates under Non Standard Conditions," *Fifth Berkeley Symposium*, 1965/66, Vol. 1, pp. 221–233.

[12] INAGAKI, N.: "Asymptotic Relations between the Likelihood Estimating Function and the Maximum Likelihood Estimator," *Annals of the Institute of Statistical Mathematics*, 25(1973), 1–26.

[13] JENNRICH, R I.: "Asymptotic Properties of Nonlinear Least Squares Estimators," *Annals of Mathematical Statistics*, 40(1969), 633–643.

[14] LEBEDEV, N. N.: *Special Functions and Their Applications*. Dover: Prentice-Hall, 1965.

[15] LECAM, L.: "Locally Asymptotically Normal Families of Distributions," *University of California Publications in Statistics*, 3(1960), 37–98.

[16] MALINVAUD, E.: *Statistical Methods of Econometrics*, 3rd Edition. Amsterdam: North-Holland, 1980

[17] PHILLIPS, P. C B, AND M R. WICKENS: *Exercises in Econometrics, Vol. II*. Cambridge, Mass.: Ballinger-Philip Allan, 1978.

[18] STEIN, C. M.: "Estimation of the Mean of a Multivariate Normal Distribution," *Annals of Statistics*, 9(1981), 1135–1151.

[19] TITCHMARSH, E. C.: *The Theory of Functions*. Oxford Oxford University Press, 1939.

[20] WALD, A.: "Note on the Consistency of the Maximum Likelihood Estimate," *Annals of Mathematical Statistics*, 20(1949), 595–601.