COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 573

ON THE CONSISTENCY OF NON-LINEAR FIML

P. C. B. Phillips

December 1980

ON THE CONSISTENCY OF NON-LINEAR FIML*

by

P. C. B. PHILLIPS

YALE UNIVERSITY

0. ABSTRACT

Examples are given which show that: (i) normality is not necessary for the consistency of the quasi maximum likelihood estimator in the non-linear simultaneous equations model (non-linear FIML) even when there are major departures from linearity; and (ii) the lemma which is used extensively by Amemiya [2] in the theoretical development of the properties of non-linear FIML under the assumption of normality is, as presently stated, incorrect.

---

*I am grateful to William C. Brainard for helpful discussions. A preliminary version of Section 2 of this paper was first circulated under the title "On a Lemma of Amemiya". The research reported in this paper was supported by the National Science Foundation under Grant Number SES 8007571.

1. INTRODUCTION

Recent theoretical work on the non-linear simultaneous equations model has emphasized the importance of the normality assumption or, more generally, correct distributional assumptions about the equation errors in establishing the consistency of the non-linear full information maximum likelihood (FIML) estimator. In particular, Amemiya [2] argued that the proof of consistency depends crucially on the assumption of normality of the error term. In this respect, the general non-linear 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 [2].

As a result, it is now a widely held view in the profession that normality of the errors is necessary for the consistency of non-linear FIML.[1] Some authors have been led to act on this belief in applied work. For example, Fair and Parke [4] have recently proposed the Hausman [8] specification test to test the hypothesis that the errors are normally distributed by comparing the non-linear FIML and three stage least squares (3SLS) estimates. This test might be appropriate in a non-linear 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. Instead, it seems to be a rather loose but nevertheless widely held view that consistency of FIML in non-linear models is the exception rather than

---

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

the rule when the likelihood is misspecified. 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 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 show by example that this is simply 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 non-linear FIML is consistent. The example is discussed in Section 3 of the paper and, since the type of non-linearity in this example is common in applied work, it should be of some relevance in practical econometric work where such non-linearities in the variables actually occur.

A further aim of the present paper is to show that under the presently stated regularity conditions in [2] normality is not, by itself, sufficient for the consistency of non-linear FIML. Specifically, we need to exclude from the structural specification of the model certain functions which can grow as fast or faster than the rate at which the true probability density of the errors decays to zero in the tails. Such functions are not at present completely excluded by the integrability conditions in [2].

2. AMEMIYA'S LEMMA

As in [2] we write the non-linear simultaneous equations model as

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

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 likelihood function concentrated with respect to the α_i $(i = 1, \ldots, n)$, then we have in the notation of [2] (see particularly, (3.8) in [2])

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

where σ^i is the i^{th} column of Σ^{-1} , $(\ )_i^{-1}$ denotes the i^{th} 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 Lemma holds, we deduce from (2) that $E\{[T^{-1}\partial L/\partial\alpha_i]_{\alpha_0}\} = 0$ and this conclusion is of vital importance in establishing that there is a weakly consistent root of the likelihood equation.

As presently stated, however, the lemma is incorrect and a supplementary condition on the function $h(u_1, \ldots, u_n)$ is needed to ensure that the result goes through. To see this we note that the proof of the lemma depends on the following application of integration by parts (equation (3.7) of [2]):

$$(3) \qquad \int_{-\infty}^{\infty} \frac{\partial h}{\partial u_i}\phi du_i = [h\phi]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} h\sum_j \sigma^{ij}u_j\phi du_i$$

where $\phi(u)$ is the joint density of $u' = (u_1, \ldots, u_n)$. Amemiya argues that:

> "the first term on the right side of (1) is zero because E(h) is finite. Therefore, integrating both sides of (1) with respect to the remaining n-1 components of u we get the desired result." (Amemiya [2], page 958)

This argument is invalid since the finiteness of $E(h)$ is not sufficient to ensure that the first term on the right side of (3) is zero. Simple counterexamples to this argument can be constructed in the present case using one of the standard examples in analysis of convergent improper integrals whose integrands do not tend to zero at infinity (see, for example, #12, page 45 in Gelbaum and Olmsted [7]). A complete counterexample to the lemma itself is more complicated since h is such that both $E(h)$ and $E(\partial h/\partial u_i)$ are finite. Moreover, the use of the lemma in Amemiya's paper involves setting h equal to certain derivatives of the functions that

that appear in the structural specification of the non-linear model.[1] Since the latter are assumed by Amemiya to be continuously differentiable, we will require the same of the function h . In view of these extra complications and, since no standard counterexamples seem to be available in the analysis literature to cover this case, it seems worthwhile to develop a complete counterexample here.

COUNTEREXAMPLE TO THE LEMMA. *It is sufficient to deal with the scalar case so we set* $n = 1$ *in the lemma. A function which satisfies the conditions of the lemma but not its conclusion is then*

$$h(u) = \ell(u)/\phi(u) \tag{4}$$

with

$$\ell(u) = \begin{cases} \{1+(u-k)\}^{-1}\exp\left\{-\dfrac{k^4(u-k)^2}{1-4(u-k)^2}\right\}\exp\left\{-\tfrac{1}{2}k^4(u-k)^2\right\}, \\ \qquad k-\tfrac{1}{2} < u < k+\tfrac{1}{2} \quad (k = 1, 2, 3, \dots) \\ 0 \quad \text{otherwise.} \end{cases} \tag{5}$$

The function $\ell(u)$ has been constructed so that at $u = k$ it takes on the value unity and at $u = k - \frac{1}{2}, k + \frac{1}{2}$ it takes on the value zero. As k becomes large the function is negligible except at the spikes in the immediate neighborhood of $u = k$ and the second exponential factor in $\ell(u)$ is a smudge function which is designed to smudge the function down to zero at the ends of the interval $\left[k-\frac{1}{2}, k+\frac{1}{2}\right]$ while retaining the continuous differentiability property. The first factor is included to provide

---

[1] For example, $h = g_{it} = \partial f_{it}/\partial\alpha_i$ in (2) above.

some asymmetry in the function about $u = k$ within each interval.

To verify that $E(h)$ exists we need only check that $\ell(u)$ is integrable. We have, in fact,

$$\text{(6)} \qquad \int_{-\infty}^{\infty} \ell(u)du = \int_{1/2}^{\infty} \ell(u)du = \sum_{k=1}^{\infty} \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} \ell(u)du$$

provided the series converges. It is bounded by

$$\text{(7)} \qquad \sum_{k=1}^{\infty} \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} |\ell(u)|du < \sum_{k=1}^{\infty} \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} 2 \exp\left\{- \frac{1}{2}k^4(u-k)^2\right\}du$$

$$= 2\sqrt{2\pi} \sum_{k=1}^{\infty} k^{-2} \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} \frac{k^2}{\sqrt{2\pi}} \exp\left\{- \frac{1}{2}k^4(u-k)^4\right\}du$$

$$< 2\sqrt{2\pi} \sum_{k=1}^{\infty} k^{-2} \int_{-\infty}^{\infty} \frac{k^2}{\sqrt{2\pi}} \exp\left\{- \frac{1}{2}k^4(u-k)^2\right\}du$$

$$= 2\sqrt{2\pi} \sum_{k=1}^{\infty} k^{-2} = 2^{1/2}\pi^{3/2}/3 .$$

It follows that $\ell(u)$ is absolutely integrable. However, $\ell(u)$ does not have a limit as $u \to \infty$ (the sequences $\{\ell(k)\}$ and $\{\ell(k+\frac{1}{4})\}$, for instance, tend to unity and zero respectively). This contradicts the argument in Amemiya's lemma, noted above, that the first term on the right side of (3) is zero.

Finally, in order to verify that all the conditions of the stated lemma hold we need to show that $E(\partial h/\partial u)$ is finite. Now, $\partial h/\partial u = \ell'(u)/\phi(u) + (u/\sigma)\ell(u)/\phi(u)$ so it will suffice to show that ℓ' is integrable. We have

$$(8) \qquad \ell'(u) = -\{1+(u-k)\}^{-2}\exp\left\{-\frac{k^4(u-k)^2}{1-4(u-k)^2} - \frac{1}{2}k^4(u-k)^2\right\}$$

$$+\{1+(u-k)\}^{-1}\left\{\frac{2k^4(u-k)[8(u-k)^2-1]}{(1-4(u-k)^2)^2}\right\}\exp\left\{-\frac{k^4(u-k)^2}{1-4(u-k)^2} - \frac{1}{2}k^4(u-k)^2\right\}$$

$$-\{1+(u-k)\}^{-1}k^4(u-k)\exp\left\{-\frac{k^4(u-k)^2}{1-4(u-k)^2} - \frac{1}{2}k^4(u-k)^2\right\}$$

for $k - \frac{1}{2} < u < k + \frac{1}{2}$ $(k = 1, 2, 3, \ldots)$ and zero elsewhere. Using the same argument as in (6) and (7) above we can establish that the first term on the right side of (8) is absolutely integrable. For the second term, we have a series whose k^{th} term is given by

$$(9) \qquad -\int_{-1/2}^{1/2}(1+z)^{-1}2k^4z(8z^2-1)(1-4z^2)^{-2}\exp\left\{-k^4\left[\frac{z^2}{1-4z^2}+\frac{z^2}{2}\right]\right\}dz \ .$$

As $k \to \infty$ the region in which the exponential factor in the integrand of (9) is significantly different from zero becomes a smaller and smaller neighborhood of the origin. At this point the integrand itself is zero because of the factor z ; and the function $[z^2/(1-4z^2)+z^2/2]$ in the exponent has a minimum. This latter function has a non-zero second derivative at $z = 0$ so that, using Laplace's method to represent (9) as $k \to \infty$ (see, for instance, equation (5-1.21) on page 185 of Bleistein and Handelsman [3]) we find that (9) is bounded by a quantity of $O(k^4(k^4)^{-3/2}) = O(k^{-2})$. This series involving (9) as its k^{th} term is bounded by an absolutely convergent series and therefore converges by comparison. It follows that the second term on the right side of (8) is absolutely integrable. A similar argument verifies that the third term on the right side of (8) is absolutely integrable.

This proves that the function $h(u)$ defined by (4) and (5) satisfies the conditions of the lemma. However, as noted above $h(u)\phi(u) = \ell(u)$ does not approach a limit of zero at infinity and the lemma is false.

☐

One apparent way of tightening up the lemma is to restrict the class of allowable h functions so that the argument following (1) is valid. It is too much to require that $h(u)$ tends to zero at the limits of its domain since we want to set $h(u)$ equal to the functions that appear in the structural specification of the model and their derivatives. It will often be unrealistic to require these functions to vanish at infinity. On the other hand, we may suitably bound the growth of these functions to ensure that $h(u)\phi(u)$ tends to zero as $u \to \infty$. This requirement then allows the structural functions that are present in the model to be unbounded over the whole space of realizations of the random elements in the model (just as they are in the linear simultaneous equations framework or the simple regression model). But the requirement will also prevent the structural functions from taking on values which become too large for certain realizations of the random elements relative to the probability that the random elements actually assume these realizations.

A correct statement of the lemma therefore obtains if we simply add the requirement that " $h(u)\phi(u) \to 0$ as $\|u\| \to \infty$."

3. AN EXAMPLE IN WHICH NON-LINEAR FIML IS CONSISTENT WITH NON-NORMAL ERRORS

It is not difficult to construct models with minor departures from linearity in which non-linear FIML is consistent for a wide class of error distributions. One such example is given by Malinvaud [11] (see, in particular, page 732); another by Phillips and Wickens [12] (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 [5] and Gallant and Holly [6] to illustrate the verification of conditions used in the development of an asymptotic theory for non-linear FIML and 3SLS.

The structural model is

$$\ell n\ y_{1t} + a_1 = u_{1t} \tag{10a}$$

$$y_{2t} + b_1 y_{1t} = u_{2t} \tag{10b}$$

and its reduced form

$$y_{1t} = e^{-a_1+u_{1t}}$$

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

The concentrated log likelihood (or quasi likelihood)[1] function is

---

[1] What would be the likelihood if the errors were normally distributed.

$$(11)\qquad L(a_1,\ b_1) = -\frac{T}{2}\ell n[\{T^{-1}\sum_t(\ell n\ y_{1t}+a_1)^2\}\{T^{-1}\sum_t(y_{2t}+b_1y_{1t})^2\}$$

$$-\{T^{-1}\sum_t(y_{2t}+b_1y_{1t})(\ell n\ y_{1t}+a_1)\}^2]$$

$$= -\frac{T}{2}\ell n\ A_T(a_1,\ b_1)\quad \text{say}$$

and its first derivatives

$$\frac{\partial L}{\partial a_1} = -\frac{1}{2}TA_T^{-1}[\{2T^{-1}\sum_t(\ell n\ y_{1t}+a_1)\}\{T^{-1}\sum_t(y_{1t}+b_1y_{1t})^2\}$$

$$-2\{T^{-1}\sum_t(y_{2t}+b_1y_{1t})\}\{T^{-1}\sum_t(y_{2t}+b_1y_{1t})(\ell n\ y_{1t}+a_1)\}]$$

$$\frac{\partial L}{\partial b_1} = -\frac{1}{2}TA_T^{-1}[\{T^{-1}\sum_t(\ell n\ y_{1t}+a_1)^2\}\{2T^{-1}\sum_t(y_{2t}+b_1y_{1t})y_{1t}\}$$

$$-2\{T^{-1}\sum_t(y_{2t}+b_1y_{1t})(\ell n\ y_{1t}+a_1)\}\{T^{-1}\sum_t y_{1t}(\ell n\ y_{1t}+a_1)\}]\ .$$

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

$$(12a)\qquad \operatorname*{plim}_{T\to\infty} T^{-1}\partial L(a_1^0,\ b_1^0)/\partial a_1 = 0$$

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

where the disturbance vector $u_t = (u_{it})$ in (10) is assumed, as in (1), to be independent and identically distributed with zero mean and covariance matrix $\sum = (\sigma_{ij})$ for all t ; but not necessarily normal. For the expectations in (12b) to be finite, we also require that the moment generating function of u_t

$$(13) \qquad \mathrm{mgf}(s) = E(e^{s'u_t})$$

exist for certain non-zero values of the vector $s' = (s_1 s_2)$. A precise region within which we will require (13) to exist will be specified later. In (12b) A is given by $A = \mathrm{plim}_{T\to\infty} A_T(a_1^0, b_1^0) = \det \Sigma$.

To prove that non-linear FIML applied to (10) gives consistent estimates it will be sufficient to show that, setting $\alpha' = (a_1, b_1)$, the following two conditions hold:

(i) $\mathrm{plim}_{T\to\infty}(T^{-1}\partial L(a_1^0, b_1^0)/\partial\alpha) = 0$; and

(ii) $\mathrm{plim}_{T\to\infty}(T^{-1}\partial L(a_1^0, b_1^0)/\partial\alpha\partial\alpha')$ is negative definite.

If (i) and (ii) hold and the convergence in (ii) is uniform in a neighborhood of (a_1^0, b_1^0), then it follows from the argument in the Appendix of Amemiya's [2] paper that the non-linear FIML estimates are consistent.

To establish (i) and (ii) we need to make explicit distributional assumptions so that the expectations that appear in the limits can be evaluated. It is easy to verify (i) and (ii) when u_t is multivariate normal. Since we are interested in specifying a non-normal error distribution for which (i) and (ii) 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

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

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 (13) exist and since

$$\text{(15)} \qquad \mathrm{mgf}(s) = E(e^{s'u_t}) = \int_0^\infty e^{\frac{1}{2}ws'\tilde{\Sigma}s}\,dG(w)$$

we have

$$\text{(16)} \qquad \partial\,\mathrm{mgf}(s)/\partial s = \{\int_0^\infty we^{\frac{1}{2}ws'\tilde{\Sigma}s}\,dG(w)\}\tilde{\Sigma}s$$

and

$$\text{(17)} \qquad \partial^2\mathrm{mgf}(s)/\partial s\partial s' = \{\int_0^\infty we^{\frac{1}{2}ws'\tilde{\Sigma}s}\,dG(w)\}\tilde{\Sigma} + \{\int_0^\infty w^2e^{\frac{1}{2}ws'\tilde{\Sigma}s}\,dG(w)\}\tilde{\Sigma}ss'\tilde{\Sigma} .$$

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

$$\text{(18)} \qquad \{\int_0^\infty wdG(w)\}\tilde{\Sigma} = \Sigma .$$

Non-linear FIML will now be consistent for every error distribution in the class (14) for which conditions (i) and (ii) hold. From (12b), we see for the present example that condition (i) requires that

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

That is

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

$$= (-\sigma_{12}, \sigma_{11})\Sigma\begin{pmatrix}1\\ 0\end{pmatrix}\{\int_0^\infty we^{\frac{1}{2}w\tilde{\sigma}_{11}}dG(w)\}/\{\int_0^\infty wdG(w)\}$$

$$= 0$$

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

$$(21) \qquad \int_0^\infty w e^{\frac{1}{2}w\tilde{\sigma}_{11}} dG(w) < \infty .$$

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

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

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

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

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

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

$$(24)\qquad \lambda(2\pi)^{-1}(\det \tilde{\Sigma})^{-1/2}\int_0^\infty v^{-1}\exp\{-v-\lambda\beta(u_t)/v\}dv$$

$$= \lambda(2\pi)^{-1}(\det \tilde{\Sigma})^{-1/2}2K_0(2\sqrt{\lambda\beta})$$

$$= \lambda\pi^{-1}(\det \tilde{\Sigma})^{-1/2}K_0\left\{2\left(\frac{\lambda u_t'\tilde{\Sigma}^{-1}u_t}{2}\right)^{1/2}\right\}$$

where $K_0(z)$ is the modified Bessel function of the third kind.[1] 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

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

and the moment generating function of u_t is simply

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

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

---

[1] See, for example, Lebedev [10] and, in particular, equation (5.10.25) on page 119 of [10] for the representation of the integral in (24).

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

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 $|z| \to \infty$ ([10], page 123). We deduce from (26) and (29) that

$$\text{(30)} \qquad \mathrm{pdf}(u_t) \sim 2^{-n/2} \pi^{-(n-1)/2} (\det \textstyle\sum)^{-1/2} \left(\frac{1}{2} u_t' \textstyle\sum^{-1} u_t\right)^{-(n+1)/4+1/2}$$

$$\cdot \exp\{-(2u_t' \textstyle\sum^{-1} u_t)^{1/2}\}$$

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 (i) and (ii) for the density (25). For condition (i) to hold it remains only to check (21). We have

$$\int_0^\infty w e^{\frac{1}{2} w \tilde{\sigma}_{11}} dG(w) = \alpha \int_0^\infty w e^{\frac{1}{2}\alpha\sigma_{11} w} e^{-\alpha w} dw$$

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,

$$s_1^2\sigma_{11} < 2 \tag{31}$$

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

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

$$E(e^{2u_{1t}}) \tag{32}$$

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

$$\underset{T\to\infty}{\mathrm{plim}}\ \frac{\partial^2L(a_1^0,\ b_1^0)}{\partial\alpha\partial\alpha'} = -(\det \textstyle\sum)^{-1}\left[\begin{array}{c|c} \sigma_{22} & \sigma_{12}e^{-a_1}E(e^{u_{1t}}) \\ \hline \sigma_{12}e^{-a_1}E(e^{u_{1t}}) & \sigma_{11}e^{-2a_1}E(e^{2u_{1t}}) - e^{-2a_1}(E(e^{u_{1t}}u_{1t}))^2 \end{array}\right]$$

$$= -(\det \textstyle\sum)^{-1}\begin{bmatrix} \sigma_{22} & \dfrac{2\sigma_{12}e^{-a_1}}{2-\sigma_{11}} \\ \dfrac{2\sigma_{12}e^{-a_1}}{2-\sigma_{11}} & e^{-2a_1}\left\{\dfrac{\sigma_{11}}{1-2\sigma_{11}} - \left(\dfrac{4\sigma_{11}}{(2-\sigma_{11})^2}\right)^2\right\} \end{bmatrix} \tag{33}$$

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

It follows that non-linear FIML applied to (10) will be consistent for the non-normal error density (25) provided $\sigma_{11} < 1/2$ and (33) is negative

---

[1] Note that this condition is also needed, at least as far as the proofs in [1] and [5] are concerned, for the consistency of the non-linear 3SLS estimator as some simple manipulations will show.

definite. The inequality constraint $\sigma_{11} < 1/2$ is really innocuous because it is just a necessary condition for $\mathrm{var}(y_{1t})$ to be finite when the errors driving the equation system (10) follow the distribution (25).[1]

4. FINAL REMARKS

The example of the previous section shows that normality is not necessary for the consistency of non-linear FIML and the analysis suggests a general procedure for constructing non-normal error distributions for which the consistency of non-linear FIML is maintained. The analysis also demonstrates that there is an intimate relationship between the form of the non-linear 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 non-linearities in the structure and the probability of outliers in the error distribution prevents the influence of outliers interferring with the operation of the law of large numbers and is, in large part, independent of the estimation technique that is being used. This is not to say, however, that non-linear FIML and 3SLS will be consistent for an identical class of error distributions. On the contrary, it does seem likely that the asymptotic properties of the non-linear 3SLS procedure will be more robust than those of non-linear FIML, but this has not been the subject of investigation in the present paper.

---

[1] Moreover this condition is also needed in justifying the asymptotic normality, although not the consistency, of non-linear 3SLS according to the proofs of [1] and [5].

REFERENCES

[1] Amemiya, T., "The Nonlinear Two Stage Least Squares Estimator," Journal of Econometrics, 2, 1974, pp. 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, pp. 955-968.

[3] Bleistein, N. and R. A. Handelsman, Asymptotic Expansions of Integrals, Holt Rinehart and Wilson.

[4] Fair, R. C. and W. R. Parke, "Full Information Estimates of a Non-Linear Macroeconometric Model," Journal of Econometrics, 13, 1980, pp. 269-291.

[5] Gallant, A. R., "Three Stage Least Squares Estimation for a System of Simultaneous Non-linear Implicit Equations," Journal of Econometrics, 5, 1977, pp. 71-88.

[6] ________, and A. Holly, "Statistical Inference in an Implicit Non-linear Simultaneous Equation Model in the Context of Maximum Likelihood Estimation," Econometrica, 48, 1980, pp. 697-720.

[7] Gelbaum, B. R. and J. R. Olmsted, Counterexamples in Analysis, Holden Day, 1964.

[8] Hausman, J. A., "Specification Texts in Econometrics," Econometrica, 46, pp. 1251-1271.

[9] Koopmans, T. C., H. Rubin and R. B. Leipnik, "Measuring the Equation Systems of Dynamic Economics," in T. C. Koopmans, (ed.), Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph 10, Wiley, 1950, pp. 53-237.

[10] Lebedev, N. N., Special Functions and Their Applications, New York: Dover, 1965.

[11] Malinvaud, E., Statistical Methods of Econometrics, (Third Edition), Amsterdam: North Holland, 1980.

[12] Phillips, P. C. B. and M. R. Wickens, Exercises in Econometrics, Vol. II, Ballinger-Philip Allan, 1978.