ECONOMETRIC MODELLING WITH NONNORMAL DISTURBANCES*

by

Stephen M. Goldfeld
Richard E. Quandt

Econometric Research Program
Research Memorandum No. 265

June 1980

*We are indebted to Mark Plant for useful comments and to NSF Grant No. SOC77/07680 for support.

Econometric Research Program
Princeton University
207 Dickinson Hall
Princeton, New Jersey

1. INTRODUCTION

In the majority of cases, the distribution of disturbances in econometric equations is assumed to be normal. This is by and large true for single equation and simultaneous equation models, for linear and nonlinear models as well as for some models that do not fit exactly into the traditional equation-fitting mold such as those involving discrete choice. There is good reason for the pervasiveness of the normal distribution. First, to the extent that error terms are thought to represent the effects of omitted variables which are assumed to be independent and additive, a vague appeal to some central limit theorem may justify the assumption of normality. Secondly, in at least the simplest model such as the single equation regression model, the assumption of normality causes maximum likelihood estimates and the BLU least squares estimates to coincide and allows easy derivation of the finite-sample distributions of the estimators.

None of these points is fully convincing. Neither the additivity, nor the independence of omitted variables is established sufficiently rigorously to make appeal to the central limit theorem comfortable. Given the advances in computing, the advantages of dealing with linear estimators in the simplest of models may not be overwhelming. Finally, in many models currently in use, the normality of disturbances yields neither linear estimators nor tractable finite-sample distributions for the estimators in any event, and frequently only asymptotic procedures are available.

These considerations may create modest doubt whether the arguments in favor of assuming error normality are overwhelming. Much more serious doubt about the usefulness of normality is raised by recent developments in qualitative choice models, disequilibrium models and certain other models involving unobservables. In these models, evaluation of the likelihood function typically requires the computation of multiple integrals of density functions which is a difficult task if normality is assumed. Consider a choice situation in which individuals must choose among m alternatives. Let the ith individual's utility from choosing alternative j be given by

$$U_{ij} = V(x_{ij},\beta + \varepsilon_{ij} \qquad (1\text{-}1)$$

where the x_{ij} are measurable characteristics of individuals and alternatives, β represents parameters and the ε_{ij} are error terms. The probability that the kth alternative is chosen by the individual is

$$P_{ik} = \Pr\{\varepsilon_{ij} - \varepsilon_{ik} < V_{ik} - V_{ij} \ \forall i \neq k\}$$

where we use V_{ij} to denote $V(x_{ij},\beta)$. If the joint pdf of the $m-1$ $\eta_{ijk} = \varepsilon_{ij} - \varepsilon_{ik}$ is given by $h(\eta_{i1k},\ldots,\eta_{imk})$, we have

$$P_{ik} = \int_{-\infty}^{V_{ij}-V_{i1}} \cdots \int_{-\infty}^{V_{ik}-V_{im}} h(\eta_{i1k},\ldots,\eta_{imk})d\eta_{i1k}\ldots k\eta_{imk} \ . \qquad (1\text{-}2)$$

Maximum likelihood estimation of the parameters β is then based on maximizing the likelihood function

$$L = \prod_{i=1}^{n} P_{i1}^{y_{i1}} \ldots P_{im}^{y_{ij}} \ , \qquad (1\text{-}3)$$

where n is the number of individuals in the sample and $y_{ij} = 1$ if individual i chooses alternative j and $y_{ij} = 0$ otherwise. If $h(\ )$

is taken to be a multivariate normal density, (1-2) requires the evaluation of an (m-1)-fold multiple integral of the multivariate normal.[1] This problem may be made more tractable by assuming the errors to be independently Weibull distributed which is a notable precedent for employing distributions other than the normal (Domencich and McFadden (1975)).

The simplest disequilibrium model is given by

$$D_t = \beta_1' x_{1t} + u_{1t}$$

$$S_t = \beta_2' x_{2t} + u_{2t} \qquad (1\text{-}4)$$

$$Q_t = \min(D_t, S_t)$$

where D_t, S_t are the unobserved demand and supply, Q_t the observed traded quantity in period t, and u_{1t}, u_{2t} error terms. It is well known (Maddala and Nelson (1974)) that the pdf of Q_t is given by

$$h(Q_t) = \int_{Q_t}^{\infty} f(D_t, Q_t)\,dD_t + \int_{Q_t}^{\infty} f(Q_t, S_t)\,dS_t \qquad (1\text{-}5)$$

where $f(D_t, S_t)$ is the joint pdf of D_t, S_t obtained from (1-4). If u_{1t}, u_{2t} are jointly normal, (1-5) and the likelihood function based on it require the evaluation of normal integrals. If m markets are in disequilibrium and they are "connected" by spillovers, the pdf corresponding to (1-5) has 2^m terms each of which involves an m-fold integral (Gourieroux, Laffont, Monfort (1980), Ito (1980), Goldfeld and Quandt (forthcoming)). Similar problems arise in simultaneous probit or tobit models.

---

1. See Hausman and Wise (1978).

The common difficulty in all these cases is the need for multiple integrals of the normal pdf which is not obtainable in closed form. Numerical methods are extremely accurate and fast for one dimension, but in spite of many ingenious approaches to integrating multivariate normal densities[2], the problem must be considered difficult if more than a double integral is required and very difficult if a four- or five-tuple integral is needed. The present paper explores the consequences in some well-known estimation problems of using an error distribution for which the integrals of type (1-5) or (1-2) are obtainable in closed form. Section 2 introduces the density in question and briefly discusses some of its properties. Section 3 applies the density to the case of estimating the parameters of a regression model by OLS. Section 4 applies it to the simple disequilibrium model. Section 5 contains some conclusions and suggestions for further applications.

2. AN ALTERNATIVE DENSITY FUNCTION

To render the computational problem tractable we want the pdf of the random variable in question to be easily integrable and to be crudely similar to the normal. Hence, we shall require the random variable u to have a pdf which is unimodal, symmetric, with support $(-\infty,\infty)$ and with integral computable in closed form. A general class of such densities is given by

$$f(u) = Ke^{-\alpha|u|}(1+\sum_{j=1}^{p}\gamma_j\alpha^j|u|^j) \qquad (2-1)$$

2. For a variety of approaches see Owen (1956), Clark (1961), Daganzo, Bouthelier, Sheffi (1977), Manski and Lerman (1978), Dutt (1976). For a review see Quandt (1980).

where $\alpha > 0$, $\gamma_j \geqq 0$, $j=1,\ldots,P$, and where K depends on α and the γ's.[3] The second-order density of this type is

$$f(u) = \frac{\alpha e^{-\alpha|u|}}{2(1+\gamma_1+2\gamma_2)}(1+\alpha\gamma_1|u|+\alpha^2\gamma_2 u^2) \qquad (2\text{-}2)$$

Is is easy to verify that its moment generating function is

$$\mu(\theta) = \frac{\frac{\alpha}{\alpha+\theta} + \frac{\alpha}{\alpha-\theta} + \frac{\gamma_1\alpha^2}{(\alpha+\theta)^2} + \frac{\gamma_1\alpha^2}{(\alpha-\theta)^2} + \frac{\gamma_2\alpha^3}{(\alpha+\theta)^3} + \frac{\gamma_2\alpha^3}{(\alpha-\theta)^3}}{2(1+\gamma_1+2\gamma_2)} \qquad (2\text{-}3)$$

The mean and variance are

$$\mu = 0 \qquad\qquad \sigma^2 = \frac{1}{\alpha^2}\left(\frac{2+6\gamma_1+12\gamma_2}{1+\gamma_1+2\gamma_2}\right)$$

A simpler version is the first order Sargan density

$$f(u) = \frac{\alpha e^{-\alpha|u|}}{2(1+\gamma_1)}(1+\alpha\gamma_1|u|) \qquad (2\text{-}4)$$

---

3. $f(u)$ is a generalization of the Laplace density which is obtained by setting $\gamma_j = 0$ for all j . We owe the suggestion and the general type of pdf given by (2-1) to Denis Sargan. Accordingly, we shall refer to this class of densities as Sargan densities.

It is obvious that $f(u)$ is continuous everywhere and that it has continuous first and second derivatives everywhere except, possibly, at $u=0$. Since numerical optimization will be necessary for maximum likelihood estimation, it will be desirable for $f(u)$ to have continuous first and second derivatives everywhere. In order to insure this, we have

<u>Theorem 1</u>. For all values of P in (2-1) greater than or equal to 1, $f'(u)$ is continuous if and only if $\gamma_1 = 1$.

<u>Proof</u>. It is sufficient to show that at $u = 0$ the left and right derivatives coincide if and only if $\gamma_1 = 1$. But evaluating the derivatives yields

$$f'(0) = \begin{cases} K\alpha(-1+\gamma_1) & \text{if } u \geqq 0 \\ K\alpha(1-\gamma_1) & \text{if } u \leqq 0 \end{cases}$$

from which the conclusion follows. We assume henceforth that $\gamma_1 = 1$.

<u>Theorem 2</u>. For all values of $P \geqq 1$, $f''(u)$ is continuous at $u = 0$.

<u>Proof</u>. Evaluating the right and left second derivatives at the origin yields

$$f''(0) = \begin{cases} K\alpha^2(1-2\gamma_1+2\gamma_2) \\ K\alpha^2(1-2\gamma_1+2\gamma_2) \end{cases}$$

which are identical for all values of the parameters. The extent to which

Sargan densities can resemble the normal density can be shown as follows. Consider the comparison of $N(0,1)$ and of a second-order Sargan pdf which has variance = 1 and which has the same density at the origin as $N(0,1)$. (Obviously there are many other ways in which "similar" Sargan pdf's can be found.) The implied values of the parameters are $\alpha = 2.11907$, $\gamma_1 = 1.0$, $\gamma_2 = .32807$. The densities of the normal and Sargan pdf's for selected values of u are displayed in Table 1. Obviously the similarity will be somewhat less marked for a first order density with $\gamma_1 = 1$. We now concentrate on this simple case.

TABLE 1.

Normal and Sargan Densities

u	Normal Densities	Sargan Densities
0	.399	.399
.5	.352	.336
1.0	.242	.220
1.5	.130	.124
2.0	.054	.064
2.5	.018	.031
3.0	.004	.014

The log likelihood function for a sample of n u_i's is

$$\log L = n\log\alpha - n\log 4 - \alpha\sum_i |u_i| + \sum_i \log(1+\alpha|u_i|) \qquad (2\text{-}5)$$

We now state

Theorem 3. The maximum likelihood estimator is a unique $\hat{\alpha} > 0$.

Proof. By differentiating,

$$\frac{\partial \log L}{\partial \alpha} = \frac{n}{\alpha} - \sum_i |u_i| + \sum_i \frac{|u_i|}{1+\alpha|u_i|}$$

$$\frac{\partial^2 \log L}{\partial \alpha^2} = -\frac{n}{\alpha^2} - \sum_i \frac{u_i^2}{(1+\alpha|u_i|)^2} \quad .$$

Setting the first derivative equal to zero, we can write $\frac{n}{\alpha} = \sum_i |u_i| - \sum_i (|u_i|/(1+\alpha|u_i|))$. As α goes to infinity, the left-hand side decreases monotonically to zero and the right-hand side is monotone increasing; hence the graphs of the two sides intersect. Since the second derivative is negative everywhere, the theorem follows. It is an immediate corollary that a sufficient condition for the uniqueness of the maximum likelihood estimator in the second order case with $\gamma_1 = 1$ and γ_2 known is that $2\gamma_2 < 1$.

3. THE ORDINARY REGRESSION MODEL

Consider the model

$$Y = X\beta + u \qquad (3\text{-}1)$$

where X is $n \times k$ of rank k and where u is a vector of iid error terms distributed according to the first order Sargan distribution with

$\gamma_1 = 1$. Three questions are of interest: (1) Are ML estimates routinely computable? (a) Do they yield reasonable results in finite samples? (3) How do they compare with OLS estimates in terms of efficiency and asymptotic efficiency?

We shall examine the last question first. We first prove

<u>Theorem 4</u>. The asymptotic distribution of the maximum likelihood estimator $\sqrt{n}(\hat{\beta}_{ML}-\beta)$ is $N(0,\frac{3.375}{\alpha^2}(\frac{X'X}{n})^{-1})$.

<u>Proof</u>. Writing the regression for the ith observation as $y_i = \beta' x_i + u_i$, and denoting by $\sum_+$ and $\sum_-$ summation over positive and negative terms of $y_i - \beta' x_i$ respectively, we have

$$\frac{\partial \log L}{\partial \alpha} = \frac{n}{\alpha} - \sum_+ (y_i-\beta' x_i) + \sum_- (y_i-\beta' x_i) + \sum_+ \frac{y_i-\beta' x_i}{1+\alpha(y_i-\beta' x_i)} - \sum_- \frac{y_i-\beta' x_i}{1-\alpha(y_i-\beta' x_i)}$$

$$\frac{\partial \log L}{\partial \beta_j} = \alpha \sum_+ x_{ij} - \alpha \sum_- x_{ij} + \sum_+ \frac{-\alpha x_{ij}}{1+\alpha(y_i-\beta' x_i)} + \sum_- \frac{\alpha x_{ij}}{1-\alpha(y_i-\beta' x_i)}$$

$$\frac{\partial^2 \log L}{\partial \alpha^2} = -\frac{n}{\alpha^2} - \sum_+ \frac{(y_i-\beta' x_i)^2}{[1+\alpha(y_i-\beta' x_i)]^2} - \sum_- \frac{(y_i-\beta' x_i)^2}{[1-\alpha(y_i-\beta' x_i)]^2}$$

$$\frac{\partial^2 \log L}{\partial \beta_j^2} = -\sum_+ \frac{\alpha^2 x_{ij}^2}{[1+\alpha(y_i-\beta' x_i)]^2} - \sum_- \frac{\alpha^2 x_{ij}^2}{[1-\alpha(y_i-\beta' x_i)]^2} \qquad (3-2)$$

$$\frac{\partial^2 \log L}{\partial \alpha \partial \beta_j} = \sum_+ x_{ij} - \sum_- x_{ij} - \sum_+ \frac{x_{ij}}{[1+\alpha(y_i-\beta' x_i)]^2} + \sum_- \frac{x_{ij}}{[1-\alpha(y_i-\beta' x_i)]^2}$$

$$\frac{\partial^2 \log L}{\partial \beta_j \partial \beta_k} = \sum_+ \frac{\alpha^2 x_{ij} x_{ik}}{[1+\alpha(y_i-\beta' x_i)]^2} + \sum_- \frac{\alpha^2 x_{ij} x_{ik}}{[1-\alpha(y_i-\beta' x_i)]^2} .$$

In order to find the asymptotic distribution of the maximum likelihood estimates, we require the probability limits of the second partial derivatives.

It can be shown with tedious algebra that, defining $\theta = (\alpha, \beta')$,

$$\text{plim}\left\| -\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\right\| = \begin{bmatrix} \frac{(2+.439e)n}{\alpha^3} & 0 \\ 0 & .109\alpha^2 e(X'X) \end{bmatrix}$$

Hence (3-3)

$$\text{plim}\left\| -\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\right\|^{-1} = \begin{bmatrix} \frac{\alpha^3}{(2+.439e)n} & 0 \\ 0 & \frac{(X'X)^{-1}}{.109\alpha^2 e} \end{bmatrix}$$

Hence the asymptotic distribution of $\sqrt{n}\,(\hat{\beta}_{ML}-\beta)$ is $N(0, \frac{3.375}{\alpha^2}(\frac{X'X}{n})^{-1})$.

Remark. The asymptotic efficiency of the OLS estimator is approximately .84. This can be shown by noting that the covariance matrix of the OLS estimator is $\hat{\sigma}^2(X'X)^{-1}$; hence the asymptotic distribution of $\sqrt{n}(\hat{\beta}_{OLS}-\beta)$ is $N(0, \frac{4}{\alpha^2}(\frac{X'X}{n})^{-1})$, since $\text{plim}\ \hat{\sigma}^2 = \frac{4}{\alpha^2}$. The asymptotic efficiency therefore is $\frac{3.375}{4} \cong .84$.

The behavior of the ML estimator in finite samples was examined through Monte Carlo experiments. Data were generated from the model $y_i = \beta_1 + \beta_2 x_i + u_i$ with

u_i having Sargan distribution with mean zero and $\alpha = .2$ and with $\beta_1 = \beta_2 = 1.0$. For sample sizes $N = 50, 100, 200, 400$ fifty replications each were generated and the mean biases and mean square errors determined. These are displayed in Table 2. Although the mean biases show no clear pattern, the MSE's of the ML estimates are uniformly smaller than those of the OLS estimator. Moreover, the ratios of the MSE's of ML and OLS estimates bracket the theoretical figure of .84 rather closely, ranging from .67 to .88 and averaging .78. Since the computation of the ML estimator requires nonlinear optimization techniques, it is of interest that no computational failures were encountered. Although in the regression context there may be no particular reason to assume Sargan-distributed errors, it is useful to know that the ML estimates based on this distribution are reasonable.

4. THE SIMPLE DISEQUILIBRIUM MODEL

The Likelihood Function and Its Properties. The simple disequilibrium model is stated in (1-4) and the density function of the observed variable Q in (1-5). Assuming that u_{1t} and u_{2t} have independent first order Sargan distributions with parameters α_1 and α_2 respectively, $g(D,S)$ is

$$g(D,S) = [\frac{\alpha_1}{4}e^{-\alpha_1|D-\beta_1'x_1|}(1+\alpha_1|D-\beta_1'x_1|)][\frac{\alpha_2}{4}e^{-\alpha_2|S-\beta_2'x_2|}(1+\alpha_2|S-\beta_2'x_2|)] \quad (4\text{-}1)$$

where the subscript t has been omitted for simplicity. Obtaining (1-5) is thus simply a matter of integrating (4-1).

Define

$$\xi(Q,\alpha_i,\beta_i,x_i) = \frac{\alpha_i}{4}e^{-\alpha_i|Q-\beta_i'x_i|}(1+\alpha_i'|Q-\beta_i'x_i|) \tag{4-2}$$

$$\psi_I(Q,\alpha_i,\beta_i,x_i) = \frac{1}{2}e^{-\alpha_i(Q-\beta_i'x_i)} + \frac{1}{4}\alpha_i(Q-\beta_i'x_i)e^{-\alpha_i(Q-\beta_i'x_i)} \tag{4-3}$$

$$\psi_{II}(Q,\alpha_i,\beta_i,x_i) = 1 - \frac{1}{2}e^{\alpha_i(Q-\beta_i'x_i)} + \frac{1}{4}\alpha_i(Q-\beta_i'x_i)e^{\alpha_i(Q-\beta_i'x_i)} \tag{4-4}$$

It follows directly that

$$h(Q) = \begin{cases} \xi(Q,\alpha_1,\beta_1,x_1)\psi_I(Q,\alpha_2,\beta_2,x_2) + \xi(Q,\alpha_2,\beta_2,x_2)\psi_I(Q,\alpha_1,\beta_1,x_1) & \text{if } Q \geqq \beta_1'x_1,\ Q \geqq \beta_2'x_2 \\ \xi(Q,\alpha_1,\beta_1,x_1)\psi_I(Q,\alpha_2,\beta_2,x_2) + \xi(Q,\alpha_2,\beta_2,x_2)\psi_{II}(Q,\alpha_1,\beta_1,x_1) & \text{if } Q < \beta_1'x_1,\ Q \geqq \beta_2'x_2 \\ \xi(Q,\alpha_1,\beta_1,x_1)\psi_{II}(Q,\alpha_2,\beta_2,x_2) + \xi(Q,\alpha_2,\beta_2,x_2)\psi_I(Q,\alpha_1,\beta_1,x_1) & \text{if } Q \geqq \beta_1'x_1,\ Q < \beta_2'x_2 \\ \xi(Q,\alpha_1,\beta_1,x_1)\psi_{II}(Q,\alpha_2,\beta_2,x_2) + \xi(Q,\alpha_2,\beta_2,x_2)\psi_{II}(Q,\alpha_1,\beta_1,x_1) & \text{if } Q < \beta_1'x_1,\ Q < \beta_2'x_2 \end{cases} \tag{4-5}$$

Restoring the subscript indexing the observations the likelihood function is simply

$$L = \prod_t h(Q_t) \quad . \tag{4-6}$$

Likelihood functions which come in pieces which depend on the current values of the parameters are frequently poorly behaved. Fortunately, major difficulties are avoided in the present case as a result of Theorem 5.

<u>Theorem 5</u>. The likelihood function (4-6) is continuous and has continuous first partial derivatives.

<u>Proof</u>. Problems may arise only on the boundaries $Q = \beta_1' x_1$ and $Q = \beta_2' x_2$ of the regions corresponding to the four parts of $h(Q)$. We thus confine ourselves to these boundaries. Since complete symmetry prevails, it is sufficient to examine a single boundary, say $Q = \beta_1' x_1$ and, correspondingly, the first and second pieces of $h(Q)$.

(a) The continuity of $h(Q)$ is established by noting that on the indicated boundary $h(Q) = \alpha_1 \psi_I(Q,\alpha_2,\beta_2,x_2)/4 + \xi(Q,\alpha_2,\beta_2,x_2)/2$ irrespective of whether the first or second piece of $h(Q)$ is employed in (4-5).

(b) In order to establish the continuity of the first partial derivatives we define $\theta_1' = (\beta_1',\alpha_1)$ and $\theta_2' = (\beta_2',\alpha_2)$. Then

$$\frac{\partial h(Q)}{\partial \theta_1} = \begin{cases} \dfrac{\partial \xi(Q,\alpha_1,\beta_1,x_1)}{\partial \theta_1}\psi_I(Q,\alpha_2,\beta_2,x_2) + \xi(Q,\alpha_2,\beta_2,x_2)\dfrac{\partial \psi_I(Q,\alpha_1,\beta_1,x_1)}{\partial \theta_1} & \text{if } Q \geqq \beta_1' x_1 \\[2ex] \dfrac{\partial \xi(Q,\alpha_1,\beta_1,x_1)}{\partial \theta_1}\psi_I(Q,\alpha_2,\beta_2,x_2) + \xi(Q,\alpha_2,\beta_2,x_2)\dfrac{\partial \psi_{II}(Q,\alpha_1,\beta_1,x_1)}{\partial \theta_1} & \text{if } Q \leqq \beta_1' x_1 \end{cases}$$

and

$$\frac{\partial h(Q)}{\partial \theta_2} = \begin{cases} \xi(Q,\alpha_1,\beta_1,x_1)\dfrac{\partial \psi_I(Q,\alpha_2,\beta_2,x_2)}{\partial \theta_2} + \dfrac{\partial \xi(Q,\alpha_2,\beta_2,x_2)}{\partial \theta_2}\psi_I(Q,\alpha_1,\beta_1,x_1) & \text{if } Q \geqq \beta_1' x_1 \\ \xi(Q,\alpha_1,\beta_1,x_1)\dfrac{\partial \psi_I(Q,\alpha_2,\beta_2,x_2)}{\partial \theta_2} + \dfrac{\partial \xi(Q,\alpha_2,\beta_2,x_2)}{\partial \theta_2}\psi_{II}(Q,\alpha_1,\beta_1,x_1) & \text{if } Q \leqq \beta_1' x_1 \end{cases}$$

Evaluating the various partial derivatives from (4-2), (4-3), (4-4) at $Q = \beta_1' x_1$ shows that the partial derivatives $\partial h(Q)/\partial \theta_1$ and $\partial h(Q)/\partial \theta_2$ are the same on the boundaries, irrespective of whether the first or second piece of (4-5) is employed, thus proving the assertion.

It is well-known that the disequilibrium likelihood function is unbounded in parameter space if the underlying errors are normal (Goldfeld and Quandt (1978)). It is in principle possible for (4-6) to become unbounded as well. The demonstration is contained in the following Theorem and Remarks.

Theorem 6. The disequilibrium likelihood function $L(\alpha_1,\beta_1,\alpha_2,\beta_2)$ with Sargan error densities is unbounded.

Proof. Define the sets

$$T_1 = \{t \mid Q_t \geqq \beta_1' x_{1t}, \quad Q_t \geqq \beta_2' x_{2t}\}$$
$$T_2 = \{t \mid Q_t < \beta_1' x_{1t}, \quad Q_t \geqq \beta_2' x_{2t}\}$$
$$T_3 = \{t \mid Q_t \geqq \beta_1' x_{1t}, \quad Q_t < \beta_2' x_{2t}\}$$
$$T_4 = \{t \mid Q_t < \beta_1' x_{1t}, \quad Q_t < \beta_2' x_{2t}\}$$

where membership in the sets clearly depends on how β_1 and β_2 are chosen in the estimation process.

The likelihood function is

$$L = \prod_{t\varepsilon T_1} [\xi(Q_t,\alpha_1,\beta_1,x_{1t})\psi_I(Q_t,\alpha_2,\beta_2,x_{2t}) + \xi(Q_t,\alpha_2,\beta_2,x_{2t})\psi_I(Q_t,\alpha_1,\beta_1,x_{1t})]$$

$$\times \prod_{t\varepsilon T_2} [\xi(Q_t,\alpha_1,\beta_1,x_{1t})\psi_I(Q_t,\alpha_2,\beta_2,x_{2t}) + \xi(Q_t,\alpha_2,\beta_2,x_{2t})\psi_{II}(Q_t,\alpha_1,\beta_1,x_{1t})]$$

$$\times \prod_{t\varepsilon T_3} [\xi(Q_t,\alpha_1,\beta_1,x_{1t})\psi_{II}(Q_t,\alpha_2,\beta_2,x_{2t})+ \xi(Q_t,\alpha_2,\beta_2,x_{2t})\psi_I(Q_t,\alpha_1,\beta_1,x_{1t})]$$

$$\times \prod_{t\varepsilon T_4} [\xi(Q_t,\alpha_1,\beta_1,x_{1t})\psi_{II}(Q_t,\alpha_2,\beta_2,x_{2t})+ \xi(Q_t,\alpha_2,\beta_2,x_{2t})\psi_{II}(Q_t,\alpha_1,\beta_1,x_{1t})] \quad .$$

Now assume that β_1 has k_1 elements and that the number of members of T_1 and T_3 together is $\leqq k_1$. Choose β_1 such that $Q_t = \beta_1' x_{1t}$ for all $t \varepsilon T_1 \cup T_3$. By the assumption about the number of elements in $T_1 \cup T_3$ this can always be done. Now let $\alpha_1 \to \infty$ with β_2, α_2 being bounded.

Then:

(1) For $t \varepsilon T_1 \cup T_3$, $\xi(Q_t,\alpha_1,\beta_1,x_{1t}) \to \infty$.

(2) For $t \varepsilon T_1 \cup T_3$, $\psi_I(Q_t,\alpha_2,\beta_2,x_{2t})$

and $\psi_{II}(Q_t,\alpha_2,\beta_2,x_{2t})$ are not equal to zero.

(3) For $t \varepsilon T_2 \cup T_4$, $\xi(Q_t,\alpha_2,\beta_2,x_{2t}) \neq 0$.

(4) For $t \varepsilon T_2 \cup T_4$, $\psi_{II}(Q_t,\alpha_1,\beta_1,x_{1t}) \to 1$.

It follows that $L(\alpha_1,\beta_1,\alpha_2,\beta_2) \to \infty$.

<u>Remark 1</u>. If the number of elements in $T_1 \cup T_3$ were greater than k_1, it would be impossible to find β_1 such that all $\xi(Q_t,\alpha_1,\beta_1,x_{1t}) \to \infty$ for $t \varepsilon T_1 \cup T_3$. There would be at least one value of t, say t^*, for which $\xi(Q_{t^*},\alpha_1,\beta_1,x_{1t^*}) \to 0$. Also, for that value t^*, $\psi_I(Q_{t^*},\alpha_1,\beta_1,x_{1t^*}) \to 0$.

Hence at least one term of the likelihood function goes to zero. Moreover, the unbounded terms are of order $0(\alpha_1^2)$ whereas the term going to zero is of order strictly less than $0(\alpha_1^{-2})$. Hence the likelihood is bounded. This remark shows, in effect, that the likelihood can become unbounded if and only if the number of elements in $T_2 \cup T_3$ is $\leqq k_1$.

<u>Remark 2</u>. Assume that the number of elements of β_2 is k_2 . Unboundedness cannot occur with α_1 and α_2 both $\to \infty$ if the number of elements in $T_2 \cup T_4$ is greater than k_2 . For in that case at least one $\xi(Q_t,\alpha_2,\beta_2,x_{2t}) \to 0$ for $t \in T_2 \cup T_4$ and all $\xi(Q_t,\alpha_1,\beta_1,x_{1t}) \to 0$ for $t \in T_2 \cup T_4$.

<u>Sampling Experiments</u>. The introduction of Sargan-distributed errors in the disequilibrium model raises at least three questions: (1) Are estimates with Sargan-distributed errors routinely computable? (2) What practical difference does it make whether the error distribution is assumed to be normal or Sargan? (3) Does the new error pdf yield a computational advantage? In order to provide some tentative answers to these questions, we report the results of some limited sampling experiments.

For all cases considered the model was

$$
\begin{aligned}
D_t &= \alpha_1 + \alpha_2 x_{1t} + \alpha_3 x_{2t} + u_{1t} \\
S_t &= \alpha_4 + \alpha_5 x_{1t} + \alpha_6 x_{3t} + u_{2t} \qquad (4\text{-}7) \\
Q_t &= \min(D_t, S_t)
\end{aligned}
$$

with $\alpha_1 = 200.0$, $\alpha_2 = -8.0$, $\alpha_3 = 1.0$, $\alpha_4 = 100.0$, $\alpha_5 = 10.0$, $\alpha_6 = 2.0$ and with x_{1t} , x_{2t} , x_{3t} being exogenous, having been generated once-and-for-all from uniform densities over the ranges (2.5, 12.5), (25.0, 85.0), (0.0, 20.0) respectively. The substance of the experiments was (1) to generate repeated samples of Q_t from (4-7) with the specification that u_{1t} , u_{2t} are independently normal and then estimate the parameters of the equations both under the (correct) assumption that the u's are normal and the (incorrect) assumption that they are Sargan-distributed; (2) to generate samples of Q_t with the specification that the u's are Sargan-distributed and estimate the parameters under the (incorrect) assumption of normality and the (correct) assumption of the Sargan pdf. The parameters of the Sargan pdf used for generating errors were $\alpha_7 = .2$ and $\alpha_8 = .2$ for the pdf's of u_{1t} and u_{2t} respectively. When the normal distributions were used for generating errors they were assumed to have the same variances as the Sargan distributions $(4/\alpha_7^2 = 4/\alpha_8^2 = 100)$ so that u_{1t} and u_{2t} have the same mean (0) and variance (100) irrespective of whether they were normally or Sargan-distributed. Sample sizes N were 30, 60 and 100. Optimization was by the quadratic hill climbing algorithm GRADX (Goldfeld and Quandt (1972)) and analytic first and second partial derivatives were used in optimization. The experiments were replicated 50 times.[4]

---

4. For N = 30 we also performed 10 replications each for normal and Sargan-distributed u's with optimization based on numerically evaluated first and second derivatives.

Only two computational failures occurred in the total of 600 optimization problems solved (2 states of the true error distribution $\times$ 2 estimating methods $\times$ 3 sample sizes $\times$ 50 replications per case). We conclude that the computation of estimates is straightforward.

Table 3 displays the root mean square errors (RMSE's) for all cases. Table 4 contains the ratios of the asymptotic standard errors of the coefficients to the RMSE's, where the former are estimated from the negative inverse Hessian of the log likelihood and averaged over the replications. For consistent estimators these ratios are asymptotically unity. Table 5 contains the ratios of the RMSE's in each case when estimation is based on the assumption of Sargan errors to the RMSE's in the corresponding case when estimation is based on normal errors. Table 6 contains the fractions of times that the correctly specified likelihood exceeds the incorrectly specified one. We note the following: (1) When in truth the errors are normal, the RMSE's based on the assumption of Sargan errors are always higher. With the exception of two coefficients in the case of $N = 30$, the increase due to misspecification is 2 to 18 percent, with an overall median of 12 percent. (2) When in truth the errors are Sargan-distributed, the RMSE's based on normality exceed those of the correctly specified estimation procedure in 10 out of 18 instances. In the cases in which the incorrectly specified estimation procedure does have an apparent advantage in terms of RMSE, this advantage is negligible in size. The median disadvantage of the incorrect estimating method is 2 percent. (3) The RMSE's for both correctly and incorrectly specified estimating procedures decline with the sample size. What is remarkable is that the percentage declines are essentially the same, irrespective of whether the estimating method is correctly specified or not with respect to the distribution of the error terms. By just looking at the improvements in RMSE's as sample size increases there is no way of telling which is the correctly specified estimating procedure. (4) The pattern of ratios of the mean asymptotic standard errors to the RMSE's (Table 4) suggests that the misspecified normal is somewhat better than the misspecified Sargan but all ratios are reasonably close to unity. (5) The correctly specified likelihood function tends to exceed in value the incorrectly specified one, but not overwhelmingly except perhaps for the largest of sample sizes considered (Table 6).

The computational times displayed in Table 7 indicate somewhat slower estimation for the Sargan specification, with the advantage of the normal specification increasing with sample size.[5] Since the operations required to evaluate the normal and the Sargan likelihood functions are comparable, the advantage of the normal likelihood function is due principally to yielding somewhat faster convergence to the optimum. Because of the very significant increase in computer time that occurs with the normal likelihood as one goes from an m-market to an (m+1)-market model (Goldfeld and Quandt (forthcoming)), and since the computer time necessary to evaluate the Sargan likelihood increases only modestly as the number of interrelated markets goes up, one may expect that estimation in much larger disequilibrium models will be feasible if Sargan densities are employed. The disadvantage of Sargan densities is, of course, that a substantial once-and-for-all effort must be incurred to calculate the required integrals in closed form.

---

5. Computations were performed on an IBM 370/3033.

5. CONCLUSIONS

The Sargan distribution appears to be a viable alternative to the normal distribution in a number of contexts in which maximum likelihood estimation requires the repeated evaluation of integrals of the underlying density function. In particular, in the single-market disequilibrium model, employment of the Sargan likelihood function yields estimates that are difficult to distinguish from those obtained by maximizing the normal likelihood, whether the underlying errors are Sargan-distributed or normal. In this sense, the family of pdf's given by (2-1) represents a robust alternative to the normal density with distinct computational advantages.

Several aspects of employing (2-1) remain to be investigated. The principal ones, to be investigated in future work, are as follows. (1) Will the computational savings in larger disequilibrium models become as substantial as predicted? (2) Will the assumption of the error distribution (2-1) yield tractable procedures and sensible answers in other types of models such as discrete choice models? (3) Which members of the class (2-1) will be most useful overall? (4) Will the tractability of the Sargan likelihood hold up if errors are not assumed to be independent? (5) Will the application of the Sargan likelihood yield sensible economic parameter estimates in concrete economic models previously estimated by maximizing normal likelihoods? We shall attempt to answer some of these questions in future work.

TABLE 2

Mean Bias and Mean Square Error

Mean Bias

N	Sargan Estimates		OLS Estimates	
	β_1	β_2	β_1	β_2
50	-.04076	.00617	.01063	-.00310
100	.00007	.00184	.02564	-.00224
200	.03232	-.00286	.02316	-.00083
400	.03273	-.00112	.03455	-.00143

Mean Square Error

N	Sargan β_1	Sargan β_2	OLS β_1	OLS β_2
50	.28956	.00750	.38104	.01115
100	.17352	.00478	.19638	.00538
200	.04933	.00146	.06644	.00187
400	.02538	.00100	.03582	.00128

TABLE 3

Root Mean Square Error

True Error Distribution	Normal						Sargan					
Sample Size	30		60		100		30		60		100	
Assumed Error Distribution	Normal	Sargan	Normal	Sargan	Normal	Sargan	Normal	Sargan	Normal	Sargan	Normal	Sargan
Coefficients												
α_1	21.242	30.624	14.675	15.017	10.265	11.444	26.453	26.811	13.088	12.504	10.429	9.624
α_2	2.307	3.237	1.555	1.634	1.028	1.148	2.678	2.825	1.578	1.477	1.020	.978
α_3	.202	.226	.146	.164	.087	.098	.197	.198	.130	.135	.077	.081
α_4	10.203	11.655	9.538	10.307	7.182	7.705	12.738	13.188	11.308	9.998	6.508	6.282
α_5	1.611	1.897	1.587	1.714	1.048	1.193	1.891	1.999	1.617	1.405	1.049	1.113
α_6	.498	.510	.315	.357	.275	.285	.658	.630	.414	.371	.245	.237

TABLE 4

Ratio of Mean Asymptotic Standard Errors to RMSE's

True Error Distribution	Normal						Sargan					
Sample Size	30		60		100		30		60		100	
Assumed Error Distribution	Normal	Sargan	Normal	Sargan	Normal	Sargan	Normal	Sargan	Normal	Sargan	Normal	Sargan
Coefficients												
α_1	.84	.57	.80	.75	.92	.82	.61	.60	.87	.90	.90	.89
α_2	.82	.59	.80	.75	.90	.81	.64	.61	.77	.81	.90	.86
α_3	.83	.75	.84	.74	1.04	.93	.82	.77	.94	.88	1.19	1.04
α_4	1.04	.94	.93	.84	.95	.90	.84	.77	.71	.78	.99	.98
α_5	.94	.82	.84	.76	.96	.87	.81	.72	.74	.82	.90	.83
α_6	.90	.92	1.06	.92	.93	.91	.69	.68	.75	.80	.98	.95

TABLE 5

Ratio of RMSE's from Assumed Sargan Errors to RMSE's from Assumed Normal Errors

True Error Distribution	Normal			Sargan		
Sample Size N	30	60	100	30	60	100
Coefficients						
α_1	1.44	1.02	1.11	.99	1.05	1.08
α_2	1.39	1.05	1.12	.95	1.07	1.04
α_3	1.12	1.12	1.13	.99	.96	.95
α_4	1.14	1.08	1.07	.97	1.13	1.04
α_5	1.18	1.08	1.14	.95	1.15	.94
α_6	1.02	1.13	1.04	1.04	1.12	1.04

TABLE 6

Fraction of Times That the Correctly Specified Likelihood Exceeds the Incorrectly Specified Likelihood

	Truth	
	Normal Errors	Sargan Errors
N		
30	.64	.54
60	.68	.56
100	.82	.76

TABLE 7

Average Time in Seconds for Obtaining Estimates

	Truth			
	Normal Errors		Sargan Errors	
Estimation	Normal	Sargan	Normal	Sargan
N				
30	.59	.65	.57	.69
60	.83	.99	.78	1.04
100	1.08	1.42	1.04	1.48
30 (numerical derivatives)	5.17	5.55	5.13	5.65

REFERENCES

Clark, C. E. (1961), "The Greatest of a Finite Set of Random Variables," Operations Research, 9, 145-162.

Daganzo, C. F., F. Bouthelier and Y. Sheffi (1977), "Multinominal Probit and Qualitative Choice: A Computationally Efficient Algorithm," Transportation Science, II, 339-358.

Domencich, T. A., and D. McFadden (1975), Urban Travel Demand, Amsterdam: North Holland.

Dutt, J. E. (1976), "Numerical Aspects of Multivariate Normal Probabilities in Econometric Models," Annals of Economic and Social Measurement, 5 547-561.

Goldfeld, S. M. and R. E. Quandt (1972), Nonlinear Methods in Econometrics, Amsterdam: North Holland.

________________ (1978), "Some Properties of the Simple Disequilibrium Model with Covariance," Economics Letters, 1, 343-346.

________________ (forthcoming), "Estimation in Multimarket Disequilibrium Models," Economics Letters.

Gourieroux, C., J.-J. Laffont, A. Monfort (1980), "Disequilibrium Econometrics in Simultaneous Equations Systems," Econometrica, 48, 75-96.

Hausman, J. A. and D. A. Wise (1978), "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," Econometrica, 46, 403-426.

Ito, T., (1980), "Methods of Estimation for Multimarket Disequilibrium Models," Econometrica, 48, 97-126.

Maddala, G. S. and F. D. Nelson (1974), "Maximum Likelihood Methods for Models of Markets in Disequilibrium," Econometrica, 42, 1013-1030.

Manski, C. F. and S. R. Lerman (1978), "On the Use of Simulated Frequencies to Approximate Choice Probabilities," forthcoming in Econometric Analysis of Discrete Data, ed. by C. F. Manski and D. McFadden, Cambridge, MA: MIT Press.

Owen, D. B. (1956), "Tables for Computing Bivariate Normal Probabilities," Annals of Mathematical Statistics, 27, 1075-1090.

Quandt, R. E. (1980), "Computational Methods and Problems," forthcoming in Handbook of Econometrics, Amsterdam: North Holland.