ESTIMATING CONTAMINATED LIMITED DEPENDENT VARIABLE MODELS*

Manuel Arellano and Olympia Bover

Institute of Economics and Statistics, Oxford

Nuffield College, Oxford

April 1986

*We are very grateful to Richard Blundell, Andrew Chesher, Chris Gilbert, David Hendry, Tony Lancaster, Costas Meghir, John Muellbauer, Steve Nickell, Denis Sargan and Richard Smith for helpful comments and to Richard Blundell for kindly making his data available.

ABSTRACT

In limited dependent variable models the estimators obtained by maximising the normal likelihood function are generally inconsistent when the assumption of normality is false. Arabmazar and Schmidt (1982) have shown that the bias from non-normality can be substantial when the degree of censoring (or truncation) in the population is relatively large. The issue is relevant because usually we do not have any a priori reason to believe that our disturbances are normally distributed.

While extensive research has recently been carried out on non-normality tests, little has been done on ways of relaxing the normality assumption itself. The view is usually taken that evidence of non-normality is a signal of structural misspecification. Although this will often be the case, it may also be indicating genuine non-normal features in the distribution of the errors. In the latter situation more general distributional assumptions are called for.

In cross section or panel data a common source of non-normality is due to the existence of extreme values in the samples which determine thick tails. We propose to use mixtures of normal distributions when the objective is to allow for leptokurtosis. Essentially, the advantages of this approach are that the normal model is nested as a special case and also the computational simplicity of the normal model is maintained. This track also allows us to construct simple tests for non-normality. Both estimation and testing issues are discussed in the paper and illustrations are provided.

1. INTRODUCTION

In limited dependent variable models the estimators obtained by maximising the normal likelihood function are generally inconsistent when the assumption of normality is false. Arabmazar and Schmidt (1982) have shown that the bias from non-normality can be substantial when the degree of censoring (or truncation) in the population is relatively large. The issue is relevant because usually we do not have any a priori reason to believe that our disturbances are normally distributed. However, the same is true with respect to any other specific distributional assumption. Therefore it is potentially appealing to investigate ways of relaxing the normality assumption in order to be able to obtain consistent estimates under a broader family of distributions.

Under the assumption of normality, the third order moment vanishes and also the fourth order moment is forced to be three times the square of the variance, thus constraining the amount of probability mass which is allowed in the tails of the distribution. In cross section or panel data a common source of non-normality is due to the existence of extreme values in the samples which determine thick tails. Thus, it is of interest to find ways of relaxing the constraints on fourth order moments. While there are available distributions which produce thicker tails than the normal (eg a t_5 produces a fourth order moment which is nine times the squared of the variance and a Laplace distribution six times), we propose to use mixtures of normal distributions when the objective is to allow for leptokurtosis. Essentially, the advantages of this approach are that the normal model is nested as a special case and also the computational simplicity of the normal model is maintained.

This track also allows us to construct simple tests for non-normality.

The consequences of distributional misspecification in the Tobit model have been considered in a number of papers. Goldberger (1983) and Arabmazar and Schmidt (1982) have investigated the asymptotic biases that result in estimating the mean of a population under sample selection. Robinson (1982) showed that in general normal-ML estimators are inconsistent under non-normality. Bera, Jarque and Lee (1984) have proposed a Lagrange Multiplier test of normality within the Pearson family of distributions. Several methods for detecting the failure of distributional assumptions are also examined in Chesher, Lancaster and Irish (1985).

Section 2 sets up the model and discusses its properties. In Section 3 the contaminated normal-ML estimation of censored, truncated and binary models is considered; estimation is performed conditional on a fixed variance ratio. Non-normality tests are discussed in Section 4. Section 5 presents numerical calculations of asymptotic biases of the normal-MLE when the errors are contaminated normal in a model containing only a constant term. In Section 6 an empirical application illustrates the performance of our suggested methods. Finally, Section 7 states the conclusions.

2. A TOBIT MODEL WITH CONTAMINATED NORMAL ERRORS

A probability distribution which is the mixture of two or more normal distributions is designed as a compound or contaminated normal distribution. In the simplest case, a random variable X is said to be distributed as a contaminated normal if its cdf is given by

$$(1) \qquad F(x) = (1-\rho)\, \Phi\left(\frac{x-\mu_1}{\sigma_1}\right) + \rho\Phi\left(\frac{x-\mu_2}{\sigma_2}\right)$$

where Φ is the N(0,1) cdf. F depends on five parameters: μ_1, μ_2, σ_1, σ_2 and ρ.[1]

The mean and variance of X are given by

$$E(X) = (1-\rho)\,\mu_1 + \rho\mu_2,$$
$$Var(X) = (1-\rho)\,\sigma_1^2 + \rho\sigma_2^2 + \rho(1-\rho)\,(\mu_1 - \mu_2)^2.$$

In econometrics, this distribution has been applied to the switching regression problem (see Kiefer(1978), Quandt and Ramsey (1978) and Schmidt (1982)) but given the robustness of OLS estimators little attention has been paid to regression models with contaminated normal errors. However, this is of interest in sample selection contexts as a way of relaxing the normality assumption allowing for leptokurtosis.

We assume the model

$$(2) \qquad y_i^* = \beta' x_i + u_i \qquad (i = 1, \dots, T)$$

where β is a vector of unknown coefficients and x_i is a vector of known constants. The u_i are random i.i.d.$(0,\sigma^2)$ variables with cdf

$$(3) \qquad F(z) = (1-\rho)\,\Phi(z/\sigma_1) + \rho\Phi(z/\sigma_2) \qquad 0 < \rho \leq 1.$$

x_i is always observed but y_i^* is not, instead we observe y_i such that

$$(4) \qquad y_i = y_i^* \quad \text{if} \quad \beta' x_i + u_i > 0$$

$$y_i = 0 \quad \text{otherwise,}$$

and we assume that the observations have been ordered so that the first n y's are the positive observations. If $\rho = 1$ the model reduces to the normal Tobit. But in general, defining $k = \frac{1}{3} E(u^4)/\sigma^4$ as our kurtosis measure, we have

$$(5) \qquad k = (1-\rho)\, r^2 + \rho q^2$$

where $r = \sigma_1^2/\sigma^2$ and $q = \sigma_2^2/\sigma^2$, and since $(1-\rho)\, r + \rho q = 1$ we can write k as

$$(5a) \qquad k = (q + r) - qr = 1 + (q-1)(1-r).$$

Direct inspection reveals that $k \geq 1$ (the normal value is $k = 1$) in view of $r > 0$, $q > 0$ and $0 < \rho \leq 1$ (what requires $r < 1$ if $q > 1$ and vice-versa), thus defining a wider family of distributions which is able to accommodate thicker tails than the normal ones.

Now we turn to consider the truncated first and second order moments of the (zero-mean) contaminated normal distribution. Let the density of u_i be

$$(6) \qquad f(z) = (1-\rho)\, \frac{1}{\sigma_1}\, \phi\left(\frac{z}{\sigma_1}\right) + \rho \frac{1}{\sigma_2}\, \phi\left(\frac{z}{\sigma_2}\right)$$

where ϕ is the N(0,1) density. The density of u_i given our selection rule is

$$(7) \qquad \frac{f(z)}{\Pr(\beta' x_i + u_i > 0)} = \frac{f(z)}{F(\beta' x_i)}$$

which is the required density in order to calculate the truncated moments. These are given by[2]

$$(8) \quad E(u_i | u_i > -\beta' x_i) = \frac{(1-\rho)\sigma_1 \ \phi(\beta' x_i/\sigma_1) + \rho\sigma_2 \ \phi(\beta' x_i/\sigma_2)}{(1-\rho) \ \Phi(\beta' x_i/\sigma_1) + \rho\Phi(\beta' x_i/\sigma_2)}$$

$$(9) \quad E(u_i^2 | u_i > -\beta' x_i) =$$

$$\frac{(1-\rho) \ \sigma_1\left[\sigma_1 \ \Phi(\frac{\beta' x_i}{\sigma_1}) - \beta' x_i \ \phi(\frac{\beta' x_i}{\sigma_1})\right] + \rho\sigma_2\left[\sigma_2\Phi(\frac{\beta' x_i}{\sigma_2}) - \beta' x_i \ \phi(\frac{\beta' x_i}{\sigma_2})\right]}{(1-\rho) \ \Phi(\frac{\beta' x_i}{\sigma_1}) + \rho\Phi(\frac{\beta' x_i}{\sigma_2})}$$

Note that if $\rho = 1$, (8) and (9) reduce to $\sigma_2 m(\beta' x_i/\sigma_2)$ and $\sigma_2^2 - \beta' x_i \sigma_2 m(\beta' x_i/\sigma_2)$, respectively, as it should be under normality. Here $m(.) = \phi(.)/\Phi(.)$.

3. MAXIMUM LIKELIHOOD ESTIMATION

The likelihood function of a mixture of normal distributions is well-known to be unbounded if the ratio of the mixing variances is left unrestricted. However, as shown by Kiefer (1978), the likelihood equations for the switching regression model have a consistent root. This irregularity has led to the investigation of alternative estimators, usually based on the method of moments (eg Cohen (1967)). Particularly relevant is the Quandt and Ramsey method based on the moment generating function for the switching regression case.

However, given that our purpose is to relax the constraint k = 1, fixing the ratio of variances does not seem to be an unreasonable restriction. We are not interested in the components of the mixture per se but in the significant departures from normality that can be achieved using mixtures. In this context, as we have a common mean, it is of little help to be able to estimate a mixture for any value of the mixing parameter ρ if $\sigma_1^2 \simeq \sigma_2^2$ (or $\sigma_1^2 \simeq Var(u)$). Letting σ_1/σ_2 to be unconstrained would allow us to leave unrestricted just one more higher order moment. At a second stage, a grid search can be conducted over σ_1/σ_2. We distinguish three cases:

a) Censored (Tobit) model. The log likelihood function apart from a constant term is

$$(10) \qquad L = \sum_{i=1}^{n} \log f(y_i - \beta' x_i) + \sum_{j=n+1}^{T} \log F(-\beta' x_j)$$

where f(.) and F(.) are given in (6) and (3), respectively. Enforcing the constraint $\sigma_1^2 = s\,\sigma_2^2$, L can be maximised as a function of β, σ_2 and ρ with the restriction $0 < \rho \leq 1$.[3] Let

$$\Phi_{ki} = \Phi\left(\frac{y_i - \beta' x_i}{\sigma_k}\right) \qquad (k = 1,2;\ i = 1, \ldots, T)$$

$$f_i = f(y_i - \beta' x_i) \qquad (i = 1, \ldots, T)$$

$$\Phi_{ki} = \Phi(-\beta' x_i/\sigma_k) \qquad (k = 1,2;\ i = n+1, \ldots, T)$$

$$F_i = F(-\beta' x_i) \qquad (i = n+1, \ldots, T)$$

$$g_{ki} = (1-\rho)\ s^{-k/2}\ \phi_{1i} + \rho\phi_{2i} \qquad (i = 1, \ldots, T)$$

Then the first derivatives are[4]:

(11) $$\frac{\partial L}{\partial \beta} = \frac{1}{\sigma_2^2} \sum_{i=1}^{n} (g_{3i}/g_{1i})\ (y_i - \beta' x_i)\ x_i - \sum_{j=n+1}^{T} (f_j/F_j)\ x_j$$

(12) $$\frac{\partial L}{\partial \sigma_2} = \frac{1}{\sigma_2} \left[- n + \frac{1}{\sigma_2^2} \sum_{i=1}^{n} (g_{3i}/g_{1i})\ (y_i - \beta' x_i)^2 + \sum_{j=n+1}^{T} (f_j/F_j)\ \beta' x_j\right]$$

(13) $$\frac{\partial L}{\partial \rho} = \frac{1}{\sigma_2} \sum_{i=1}^{n} \left(\frac{\phi_{2i} - \phi_{1i}/s^{\frac{1}{2}}}{f_i}\right) + \sum_{j=n+1}^{T} \left(\frac{\Phi_{2j} - \Phi_{1j}}{F_j}\right)$$

Var(u) and k can be obtained from σ_2 and ρ for a given value of s as follows

(14) $$\sigma^2 = \mathrm{Var}(u) = \sigma_2^2\left[s + (1-s)\rho\right]$$

and from (5) and (14)

(15) $$k = \frac{s^2 + (1-s^2)\rho}{[s + (1-s)\rho]^2} .$$

Let us consider the matrix

(16) $$D = \begin{bmatrix} \frac{\partial\sigma^2}{\partial\sigma_2} & \frac{\partial\sigma^2}{\partial\rho} \\ \frac{\partial k}{\partial\sigma_2} & \frac{\partial k}{\partial\rho} \end{bmatrix} = \begin{bmatrix} 2\sigma_2[s + (1-s)\rho] & \sigma_2^2\ (1-s) \\ 0 & \frac{(1-s)^2[s-(1+s)\rho]}{[s + (1-s)\rho]^3} \end{bmatrix}$$

then

(17) $$\text{Asy Var} \begin{bmatrix} \hat{\sigma}^2 \\ \hat{k} \end{bmatrix} = \bar{D}\ \text{Asy Var} \begin{bmatrix} \hat{\sigma}_2 \\ \hat{\rho} \end{bmatrix} \bar{D}'$$

where $\bar{D}$ indicates D evaluated at the true values of the parameters.

Thus in particular Asy $\text{var}(\hat{k}) = \left[\frac{\partial k}{\partial p}\right]^2$ Asy $\text{var}(\hat{\rho})$

b) Truncated model. The log likelihood function is

(18) $$L = \sum_{i=1}^{n} \log f(y_i-\beta' x_i) - \sum_{i=1}^{n} \log F(\beta' x_i)$$

The first derivatives are

(19) $$\frac{\partial L}{\partial\beta} = \frac{1}{\sigma_2^2} \sum_{i=1}^{n} (g_{3i}/g_{1i})\ (y_i-\beta' x_i)\ x_i - \sum_{i=1}^{n} \left[f_i/(1-F_i)\right] x_i$$

(20) $$\frac{\partial L}{\partial \sigma_2} = \frac{1}{\sigma_2}\left\{ -n + \frac{1}{\sigma_2^2}\sum_{i=1}^{n} (g_{3i}/g_{1i})\ (y_i-\beta' x_i)^2 + \sum_{i=1}^{n}\left[f_i/(1-F_i)\right]\beta' x_i\right\}$$

(21) $$\frac{\partial L}{\partial \rho} = \frac{1}{\sigma_2}\sum_{i=1}^{n}\left(\frac{\Phi_{2i}-\Phi_{1i}/s^{\frac{1}{2}}}{f_i}\right) + \sum_{i=1}^{n}\left(\frac{\Phi_{2i}-\Phi_{1i}}{1-F_i}\right)$$

c) Binary model. The log likelihood function is

(22) $$L = \sum_{i=1}^{n} \log F(\beta' x_i) + \sum_{j=n+1}^{T} \log F(-\beta' x_j)$$

A normalisation is required in this case since the β's are only identified up to a scalar factor. Taking $\sigma^2 = 1$ amounts to set $\sigma_2^2 = [s + (1-s)\rho]^{-1}$, but in this context it is simpler to take $\sigma_2^2 = 1$, thus considering the estimation of $\beta/\sigma_2 = \beta^*$, say, ie

$$F(-\beta' x_j) = (1-\rho)\ \Phi(-\beta^{*\prime} x_j/c) + \rho\Phi(-\beta^{*\prime} x_j) = F_j$$

where $c = s^{\frac{1}{2}}$ (if β/σ is required, it can be calculated as $\beta^*/[s + (1-s)\rho]^{\frac{1}{2}}$). L is maximised as a function of β^* and ρ. The first derivatives are

(23) $$\frac{\partial L}{\partial \beta^*} = \sum_{i=1}^{n} [f_i/(1-F_i)]x_i - \sum_{j=n+1}^{T} (f_j/F_j)\ x_j$$

where now

$$f_i = (1-\rho)\,\frac{1}{c}\,\phi(-\beta^{*\prime}x_i/c) + \rho\phi(-\beta^{*\prime}x_i)$$

(24) $$\frac{\partial L}{\partial \rho} = -\sum_{i=1}^{n}\left(\frac{\Phi_{2i}-\Phi_{1i}}{1-F_i}\right) + \sum_{j=n+1}^{T}\left(\frac{\Phi_{2j}-\Phi_{1j}}{F_j}\right)$$

where $\Phi_{1i} = \Phi(-\beta^{*\prime}x_i/c)$ and $\Phi_{2i} = \Phi(-\beta^{*\prime}x_i)$.

Note that if the only regressor is a constant term, i.e. $\beta^{*\prime}x_i = \mu$, the contaminated binary model is not identified. In this case,

$$\text{Prob}\ (y_i = 1) = (1-\rho)\ \Phi(\mu/c) + \rho\Phi(\mu)$$

A particular value of Prob $(y_i=1)$ defines an observable structure. For any given value of μ, we can find a corresponding value of ρ given by

$$\rho = \frac{\Phi(\mu/c) - \text{Prob}\ (y_i=1)}{\Phi(\mu/c) - \Phi(\mu)}$$

which produces an observationally equivalent structure. On the other hand, non-existence of the maximum likelihood estimator only seems to occur in the case of complete sample separation, which is common to probit and logit specifications.

If σ_1/σ_2 is not constrained, the censored and truncated likelihood functions are unbounded. The censored likelihood is

$$L(\beta,\ \sigma_1,\ \sigma_2,\ \rho)\ =\ \prod_{i=1}^{n} f_i \prod_{j=n+1}^{T} F_j$$

considering the partitions $\beta' = (\beta_1'\ \beta_2)$ and $x_i' = (x_{1i}'\ x_{2i})$ where β_2 and x_{2i} are scalars, if for example we set $\beta = \beta^*$ with $\beta_2^* = (y_1 - \beta_1^{*}{}'x_1)/x_{21}$ so that y_1 equals $\beta^{*}{}'x_1$, then

$$f_1 = (1-\rho)\frac{1}{\sqrt{2\pi}\sigma_1} + \rho\frac{1}{\sqrt{2\pi}\sigma_2}$$

Now, as $\sigma_1 \rightarrow 0$, $f_1 \rightarrow \infty$, and since for $i \neq 1$

$$f_i \rightarrow \rho\frac{1}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{1}{2\sigma_2^2}(y_i - \beta^{*}{}'x_i)^2\right] \neq 0$$

and

$$F_j \rightarrow (1-\rho) + \rho\Phi\left[\frac{-\beta^{*}{}'x_j}{\sigma_2}\right] \neq 0$$

we have that $L(\beta^*,\ 0,\ \sigma_2,\ \rho) = \infty$. The same is true for the unrestricted truncated likelihood function. However, the unrestricted binary likelihood which is given by

$$L(\beta,\rho,c)\ =\ \prod_{i=1}^{n}(1-F_i)\prod_{j=n+1}^{T} F_j$$

is not unbounded since the F_i's are bounded themselves. This suggests the possibility of obtaining consistent estimates of c from binary analysis, which can be used as conditioning values at the Tobit stage or as initial values if unrestricted Tobit estimation is attempted. Notice that we would have an observationally equivalent model by assuming that β changes over regimes by a proportional factor and the variances are the same over regimes.

Thus, it is of some interest to investigate the amount of information about c contained in the binary likelihood function. So we turn to consider the information matrix for this problem.

The log-likelihood for one observation is

$$L_i = y_i \log F(\beta' x_i) + (1-y_i) \log F(-\beta' x_i)$$

Now, letting $\theta' = (\beta' \;\rho\; c)$ we have

$$\frac{\partial L_i}{\partial \theta} = \xi_i \begin{bmatrix} f(\beta' x_i) x_i \\ \Phi(\beta' x_i) - \Phi(\beta' x_i / c) \\ -(1-\rho)\; \phi(\beta' x_i / c) \frac{1}{c^2} \beta' x_i \end{bmatrix} = \xi_i b_i \tag{25}$$

where

$$\xi_i = \frac{y_i}{F(\beta' x_i)} - \frac{(1-y_i)}{1-F(\beta' x_i)}$$

with $E(\xi_i) = 0$ and $E(\xi_i^2) = \dfrac{1}{F(x_i'\beta)\; F(-x_i'\beta)}$

Hence $E(\partial L_i / \partial \theta) = 0$ and

$$\lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{T} E\left[\frac{\partial L_i}{\partial \theta} \frac{\partial L_i}{\partial \theta'}\right] \lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{T} E(\xi_i^2)\; b_i\; b_i' = \Theta \text{ say} \tag{26}$$

where

$$b_i b_i' = \begin{bmatrix} f_i^2 \underset{\sim}{x}_i \underset{\sim}{x}_i' & (\Phi_{2i}-\Phi_{1i}) f_i \underset{\sim}{x}_i & \frac{-(1-\rho)}{c^2} \phi_{1i} f_i \underset{\sim}{x}_i' \underset{\sim}{x}_i' \beta \\ & (\Phi_{2i}-\Phi_{1i})^2 & \frac{-(1-\rho)}{c^2} \phi_{1i} \underset{\sim}{x}_i' \beta \\ & & \frac{(1-\rho)^2}{c^4} \phi_{1i}^2 \beta' \underset{\sim}{x}_i \underset{\sim}{x}_i' \beta \end{bmatrix}$$

Assuming that x_i is bounded and that its empirical distribution function converges to a distribution function, Jenrich (1969)'s theorems can be applied to establish the convergence of (26) (see also Amemiya (1973)).

Furthermore, under mild assumptions it can be shown that

$$(27) \qquad \sqrt{T}\,(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Theta^{-1})$$

Thus, it was felt convenient to compute Θ^{-1} for various single models in order to obtain some qualitative information about the unconditional M.L. estimator. In the specification we consider β is a scalar parameter which is specified to be the same for all cases, $\beta = .8$. The series of $\{x_i\}$ was generated as normal with mean zero and unit variance and the infinite sum in (26) was truncated at T = 10,000. Two alternative values of c were used: c = .05 or .15 and five alternative values of ρ were used for each c:ρ = .1, .25, .5, .75, .9. Table IV reports the results. In view of (27) we can use the approximation AVAR($\hat{\theta}$) $\simeq$ $(1/T)\Theta^{-1}$ in order to compute the asymptotic standard errors (ASE) in samples of a given size T. In this way, we may for example determine which T is required for the ASE of a particular coefficient to be one half of the corresponding true value. Thus, in relation to c we have that for $\rho = .5$, the required sample sizes are 4,450 and 1,680, respectively for $c = .05$ and $.15$, and when $\rho = .75$ these are 17,800 and 6,700. Not surprisingly, as ρ increases so does the AVAR of $\hat{c}$, and the less well determined is c, since when $\rho=1$, c is not identified. Of course, these results are no more than suggestive since they are model specific.

Notice that Heckman's two stage procedure can be extended to this case to provide consistent initial estimates for the maximum likelihood estimation of the Tobit model. Arguing as Heckman (1979), the regression function for the subsample of selected data is

$$(28) \qquad E(y_i | u_i > -\beta' x_i) = \beta' x_i + E(u_i | u_i > -\beta' x_i)$$

but in view of (8), the second term of the right hand side equals

$$(29) \qquad E(u_i | u_i > -\beta' x_i) = \sigma_2 \left[\frac{(1-\rho) \; c\phi(\beta^{*\prime} x_i / c) + \rho\phi(\beta^{*\prime} x_i)}{F_i} \right]$$

$$= \sigma_2 \lambda_i, \text{ say.}$$

λ_i can be estimated from binary analysis using the full sample, and then the estimated series can be used as a regressor in equation (28) to provide consistent estimators of β and σ_2.

4. TESTING NON-NORMALITY

A test of the constraint $\rho = 1$ is a test of non-normality. However, in the unconstrained model this is not an admissible value since when $\rho = 1$ c is not identified. Therefore, a standard test based on the more general model is not possible. On the other hand, under the null hypothesis of normality, maximum likelihood estimators conditional to an arbitrary value of c such that $0 < c < 1$, are consistent and asymptotically normal. This suggests the possibility of constructing specification tests for all three cases censored, truncated and binary. In fact, what we have is an example of the situation discussed by Davies (1977) where a parameter is identified only under the alternative.[5]

It is perhaps useful to develop our results in a slightly more general case in which the symmetry assumption is relaxed by introducing a further parameter a. Namely, we replace (3) by

$$F(z) = (1-\rho)\,\Phi\left(\frac{z-a}{\sigma_1}\right) + \rho\Phi\left(\frac{z - (1-\rho^{-1})a}{\sigma_2}\right) \tag{30}$$

By allowing a to be non-zero, we may have a non-vanishing third order moment while maintaining the zero mean property. The first fourth moments are given by

$$E(u_i) = 0$$

$$E(u_i^2) = \sigma^2 = \sigma_2^2[s+ (1-s)\rho] -a(1-\rho^{-1})$$

$$E(u_i^3) = a(1-\rho)[(a/\rho)^2\ (2\rho-1) - 3(1-s)\ \sigma_2^2] \tag{31}$$

$$\frac{1}{3}\,E(u_i^4) = \sigma_2^4[s^2 + (1-s^2)\rho]$$

$$- 2a^2\sigma_2^2\ (1-\rho^{-1})\ [1 - (1-s)\rho]$$

$$- a^4\ (1-\rho^{-1})\ (\rho^2 - \rho + \tfrac{1}{3})/\rho^2.$$

Note that in general when $s = 1$ and $a \neq 0$, $k \neq 1$ (actually, $k = 1$ only for the values of ρ that solve $6\rho^2 - 6\rho + 1 = 0$). It does not seem to be much point in testing for normal kurtosis if symmetry has been previously rejected. A test of skewness seems to be a more useful specification diagnostic, which in our framework amounts to a test of the constraint $a = 0$, provided $\rho < 1$. It could also be of interest to test for heterokurtosis if it is found that for a symmetric distribution $E(u_i^4) \neq 3\sigma^4$, as a further diagnostic which in some cases can aid to distinguish between structural misspecification and genuine non-normal tails.

In any event, both c and a are irrelevant under the null hypothesis $\rho = 1$, and in principle a generalisation of Davies test to two parameters present only under the alternative can be use. Davies solution consists in obtaining the LM statistic for each pair (a,c) and then base the test on the supremum of these. Any arbitrary value of a and c will produce asymptotically a chi-square with one degree of freedom, but this is not generally the case for their maximum since these chi-squares are not independent. Davies suggested a method to calculate upper bounds on the asymptotic size of the test for any critical value. To be more specific, let $s(\bar{a},\bar{c})$ be the normalised score for a = $\bar{a}$ and c = $\bar{c}$. That is, letting $\theta' = (\beta' \ \sigma_2)$,

$$\begin{bmatrix} V_{\theta\theta} & v_{\theta\rho} \\ v'_{\theta\rho} & v_{\rho\rho} \end{bmatrix} = -\begin{bmatrix} \dfrac{\partial^2 L}{\partial\theta \ \partial\theta'} & \dfrac{\partial^2 L}{\partial\theta \ \partial\rho} \\ \dfrac{\partial^2 L}{\partial\rho\partial\theta'} & \dfrac{\partial^2 L}{\partial\rho^2} \end{bmatrix}\Bigg|_{\theta=\hat{\theta},\ \rho=1;\ a=\bar{a},\ c=\bar{c}}$$

and

$$q = \frac{\partial L}{\partial\rho}\Bigg|_{\theta=\hat{\theta},\ \rho=1;\ a=\bar{a},\ c=\bar{c}}$$

where $\hat{\theta}$ is the gaussian maximum likelihood estimator of θ and L is the relevant log likelihood function conditional on a and c, we have

(32) $$s(\bar{a},\bar{c}) = \frac{q}{(v_{\rho\rho} - v'_{\theta\rho} V^{-1}_{\theta\theta} \ v_{\theta\rho})^{1/2}} .$$

Then under standard assumptions $s(\bar{a},\bar{c})$ is asymptotically a N(0,1) variable under the null. Davies statistic is

(33) $$D = \sup_{0<c<1,\ a} s(a,c)$$

As discussed by Watson (1982), D can be approximated by the maximum of s(a,c) over an arbitrary number of values of a and c, $\psi_1, \ldots, \psi_n$, say, where $\psi' = (a \ c)$. Now since $[s(\psi_1), \ldots, s(\psi_n)]$ is asymptotically distributed as a multivariate normal, one can calculate rejection probabilities associated to this approximation if n is not too large, and in any event several upper bounds are available (see Watson (1982) for the details).

In the remainder of this section we consider explicit expressions of q for different models to gain some insight into the interpretation of this procedure. First, we consider a standard regression model; in this case, the log likelihood function is simply

$$L(\beta,\ \sigma_2,\rho|c,a) = \Sigma_{i=1}^{T} \log f_i$$

with

$$f_i = (1-\rho)\ \frac{1}{\sigma_1}\ \phi\left(\frac{u_i - a}{\sigma_1}\right) + \rho\ \frac{1}{\sigma_2}\ \phi\left(\frac{u_i-(1-\rho^{-1})a}{\sigma_2}\right)$$

where $u_i = y_i - x_i'\beta$ and $\sigma_1 = c\sigma_2$. Let $\phi_{1i} = \phi\left(\frac{u_i-a}{\sigma_1}\right)$ and

$$\phi_{2i} = \phi\left(\frac{u_i-(1-\rho^{-1})a}{\sigma_2}\right).$$

Under the null, $\sigma_2^{-1}\ \phi_{2i}$ is the true pdf and in this case $\hat{\beta}$ and $\hat{\sigma}_2$ are the OLS estimates of β and σ_2. We have

$$\frac{\partial \log f_i}{\partial \rho} = \frac{1}{\sigma_2 f_i}\left[-\frac{1}{c}\ \phi_{1i} + \phi_{2i}\ \{1 + \left(\frac{u_i-\ (1-\rho^{-1})a}{\sigma_2}\right)\ \frac{a}{\rho\sigma_2}\}\right]$$

Thus

$$q = \sum_{i=1}^{T} \left(1- \frac{c^{-1}\hat{\phi}_{1i}}{\hat{\phi}_{2i}}\right) + a\hat{\sigma}_2^{-2} \sum_{i=1}^{T} \hat{u}_i = \sum_{i=1}^{T} \left(1- \frac{c^{-1}\hat{\phi}_{1i}}{\hat{\phi}_{2i}}\right) \tag{34}$$

where $\hat{u}_i = y_i - x_i'\hat{\beta}$ and $\hat{\phi}_{1i}$ and $\hat{\phi}_{2i}$ are ϕ_{1i} and ϕ_{2i} evaluated at $\hat{u}_i$, $\hat{\sigma}_2$ and $\rho=1$. Clearly

(35) $$E(1-\frac{c^{-1}\phi_{1i}}{\phi_{2i}}) = 1 - \int \frac{c^{-1}\phi_{1i}}{\phi_{2i}} f_i \, d\eta$$

which vanishes if $f_i = \sigma_2^{-1} \phi_{2i}$. Our test statistic is simply replacing this expectation by its sample counterpart. Under this interpretation, $\sigma_1^{-1} \phi_{1i}$ is playing the role of a pivotal density and, since in principle it could be replaced by any density g_i with mean a, say, other than $\sigma_2^{-1} \phi_{2i}$, this suggests considering a wider class of distributions of the form

(36) $$f_i = (1-\rho)g_i + \rho\frac{1}{\sigma_2} \phi_{2i}.$$

Next, we consider a binary model. In this case, the log likelihood function is

$$\begin{aligned} L(\beta,\rho|c, a) &= \Sigma_{i=1}^{T}[y_i \log(1-F_i) + (1-y_i)\log F_i] \\ &= \Sigma_{i=1}^{T} L_i \end{aligned}$$

(Note that since f_i is not necessarily symmetric, we cannot replace $1-F_i$ by $F(x_i'\beta)$.)

We have

$$\frac{\partial L_i}{\partial \rho} = \left[\frac{(1-y_i)}{F_i} - \frac{y_i}{(1-F_i)}\right]\frac{\partial F_i}{\partial \rho}$$

with

$$\frac{\partial F_i}{\partial \rho} = \Phi(z_i - (1-\rho^{-1})a) - \Phi(\frac{z_i-a}{c}) - (a/\rho)\ \phi(z_i - (1-\rho^{-1})a)$$

where $z_i = -x_i'\beta$ and we have set $\sigma_2 = 1$. Thus

$$q = \Sigma_{i=1}^{T}\ \hat{w}_i(a,c)\ (y_i-\Phi(x_i'\hat{\beta})) \tag{37}$$

with

$$\hat{w}_i(a,c) = \Phi(\frac{\hat{z}_i-a}{c}) + \Phi(\hat{z}_i) + a\phi(\hat{z}_i) \tag{38}$$

where $\hat{z}_i = -x_i'\hat{\beta}$ and $\hat{\beta}$ is the probit estimate of β. Again,

$$E[w_i(a,c)\ (y_i - \Phi(x_i'\beta))] = w_i(a,c)\ [Pr(y=1) - \Phi(x_i'\beta)] \tag{39}$$

which vanishes in the absence of misspecification, thus providing a simple interpretation for our statistic.

Finally, let us consider the Tobit model which combines the two previous cases. Since the results are essentially the same, we present the a=0 case to simplify the presentation. In view of (13) we have

$$\frac{\partial L}{\partial \rho} = \sum_{i=1}^{T}\left[\left[\frac{\sigma_2^{-1}\phi_{2i}-\sigma_1^{-1}\phi_{1i}}{f_i}\right] w_i + \left[\frac{\Phi_{2i}-\Phi_{1i}}{F_i}\right](1-w_i)\right]$$

where w_i is observable and takes the values $w_i = 1$ if $y_i^* > 0$. $w_i = 0$

otherwise, so that $E(w_i) = 1 - F_i$. Thus

$$q = \sum_{i=1}^{T} \left(1 - \frac{c^{-1}\hat{\Phi}_{1i}}{\hat{\Phi}_{2i}}\right) w_i - \sum_{i=1}^{T} \left(\frac{(1-\hat{\Phi}_{2i}) - (1-\hat{\Phi}_{1i})}{\hat{\Phi}_{2i}}\right)(1-w_i) \tag{40}$$

Again, note that

$$E\left(1 - \frac{c^{-1}\Phi_{1i}}{\Phi_{2i}} \;\middle|\; y_i^* > 0\right) = 1 - \int_{-x_i'\beta}^{\infty} \frac{c^{-1}\Phi_{1i}}{\Phi_{2i}} \frac{f_i}{(1-F_i)} d\eta$$

which equals

$$1 - \frac{1}{(1-\Phi_{2i})} \int_{-x_i'\beta}^{\infty} \sigma_1^{-1}\Phi_{1i} d\eta = \frac{(1-\Phi_{2i}) - (1-\Phi_{1i})}{(1-\Phi_{2i})}$$

under null.

5. CALCULATIONS OF ASYMPTOTIC BIASES

In this Section we follow Arabmazar and Schmidt (1982) to calculate the inconsistency of the normal MLE when the errors are actually contaminated normal. As they do, we concentrate on the special case in which the model contains only a constant term (i.e. $\beta' x_i = \mu$) but the error variance is unknown. We solve numerically for the probability limits of the normal MLE's $\hat{\mu}$ and $\hat{\sigma}$ the equations obtained as the probability limits of the normal first order conditions. These are as follows:

Censored case

$$\tilde{\mu} + A\tilde{\sigma}m(-\tilde{\mu}/\tilde{\sigma}) - B = 0$$

$$\tilde{\mu}^2 - \tilde{\sigma}^2 + A\ \tilde{\mu}\ \sigma\ m(-\tilde{\mu}/\tilde{\sigma}) - 2B\tilde{\mu} + C = 0$$

Truncated case

$$\tilde{\mu} + \tilde{\sigma}\ m(\tilde{\mu}/\tilde{\sigma}) - B = 0$$

$$\tilde{\mu}^2 - \tilde{\sigma}^2 + \tilde{\mu}\tilde{\sigma}m(\tilde{\mu}/\tilde{\sigma}) - 2B\tilde{\mu} + C = 0$$

where $\tilde{\mu}$ = plim $\hat{\mu}$, $\tilde{\sigma}$ = plim $\hat{\sigma}$. For comparability with their results we set $Var(u_i) = \sigma^2 = 1$. Then A, B and C are given by (see Arabmazar and Schmidt for the details):

$$A = F(-\mu)/F(\mu),$$

$$B = \mu + E(u_i|u_i > -\mu),$$

$$C = \mu^2 + E(u_i^2|u_i>-\mu) + 2\mu E(u_i>-\mu).$$

The expressions for the first and second order truncated moments under the assumption that the cdf of u_i is contaminated normal are given in (8) and (9).

We are interested in computing asymptotic biases for mixtures of distributions giving rise to specified values of the kurtosis measure k. In this regard, it turns out to be more convenient to set $r = \sigma_1^2/\sigma^2$ to some fixed value and parameterise ρ and σ_2^2 in terms of k and σ^2. Note that solving for q in (5a) we have $q = (k-r)/(1-r)$ and that since $(1-\rho)r + \rho q = 1$, $\rho = (1-r)/(q-r)$. Therefore, making replacements we obtain

$$\sigma_1^2 = r\sigma^2,$$

$$\sigma_2^2 = (\frac{k-r}{1-r})\sigma^2,$$

$$\rho = \frac{(1-r)^2}{k - 2r + r^2}.$$

For all the distributions considered $r = 1/3$ with k equal to 2, 3 and 4. In all cases $\sigma = 1$, while μ varies from -2.8 to 2.8, thus determining the degree of censoring or truncation in the population.

The results are given in Tables I-III, each table corresponding to a different value of k. For comparability with Arabmazar and Schmidt's results it is convenient to remark that for the Laplace distribution $k = 2$ and for a t_5 this is $k = 3$. Our results are qualitatively very similar to those found by Arabmazar and Schmidt, thus stressing their main conclusions. If anything, we may notice that the biases found for the contaminated normals tend to be more persistent for lower degrees of censoring or truncation than those corresponding to the t distributions. As it would be expected, the bias is worse for distributions with a larger value of k(i.e. in which the kurtosis measure tends to be more non-normal).

6. EMPIRICAL ILLUSTRATION

In order to illustrate the procedures developed above, it was decided to estimate a female labour supply equation. This equation is identical to the one presented by Blundell and Smith (1985) in their study of simultaneous equations Tobit models. Female labour supply, measured by weekly hours in paid work is described by a reduced form equation which includes as explanatory variables other household income, three child dummy variables and linear and quadratic age and education effects. The exogeneity of other income can be questioned and actually the test performed by Blundell and Smith rejects this hypothesis. Therefore we may expect simultaneity bias from the application of standard methods.

The data consists of 2539 married women of working age which are not self employed, from the 1981 Family Expenditure Survey for the U.K. Column 2 in Table V presents contaminated Tobit estimates conditional on an optimal choice for the variance ratio obtained over a grid search. Column 1 provides the normal Tobit estimates for comparisons. Table VI gives $(-\hat{L})$, $\hat{\rho}$, $\hat{k}$ and $(1-\hat{\rho})/SE(\hat{\rho})$ for different values of s. In all cases the differences with respect to the normal estimates remained and there were no noticeable departures in the estimated regression coefficients from the results reported in Column 2, Table V. The normal values were used as initials and in no occasion the likelihood function failed to converge. In fact, convergence was attained very quickly, typically after four or five iterations.

The estimated value of ρ in Column 2 (Table V) is consistent with a kurtosis measure ($\gamma_2 = 3.7$) not very far from the gaussian specification and in fact the contaminated estimates are typically only between 10 and

20 percent larger in absolute value than the normal ones. However, a completely different picture emerges once asymmetry is allowed into the distribution of the errors by introducing a further parameter as described in Section 4. The results, again conditional on an optimal value of s, are reported in Column 3, Table V. The new parameter is highly significant and so is our estimated skewness measure calculated as $E(u_i^3)/\sigma^3$ using the formulae given in (31). Accordingly $\hat{L}$ is considerably improved. Our estimates imply a distribution skewed to the right with a long left tail. Actually, the mode is 12.25! On the other hand, with the exception of the child dummy variables, the slope coefficients are much altered.

Although further investigation on more elaborate models would be required, these results suggest that the observed non-normality is due to structural misspecification and not to genuine non-normality in the errors. In any event, the proposed estimators seem to be a useful generalisation towards robustness in limited dependent variable models.

7. CONCLUSIONS

While extensive research has recently been carried out on non-normality tests, little has been done on ways of relaxing the normality assumption itself. The view is usually taken that evidence of non-normality is a signal of structural misspecification. Although this will often be the case (particularly if the residuals are found to be skewed as it happens in our illustration) it may also be indicating genuine non-normal features in the distribution of the errors. When this is the case, more general distributional assumptions are called for.

We have proposed a method of relaxing the normality assumption in sample selection models when the suspected cause of non-normality is in the form of longer tails leading to leptokurtosis. Normality is nested as a particular case, thus providing the basis for simple tests of non-normality. These issues are of practical importance because the asymptotic biases due to non-normal kurtosis can be substantial.

REFERENCES

Amemiya, T. (1973): "Regression Analysis when the Dependent Variable is Truncated Normal", Econometrica, 41, 997-1016.

Arabmazar, A. and P. Schmidt (1982): "An Investigation of the Robustness of the Tobit Estimator to Non-normality", Econometrica 50, 1055-1063.

Bera, A., C. Jarque and L.-F. Lee (1984): "Testing the Normality Assumption in Limited Dependent Variable Models", International Economic Review 25, 563-578.

Blundell, R.W. and R.J. Smith (1985): "Estimation in the Simultaneous Equation Tobit Model", Mimeo.

Chesher, A., T. Lancaster and M. Irish (1985): "On Detecting the Failure of Distributional Assumptions", Annales de l'INSEE, 59/60, 7-45.

Cohen, A.C. (1967): "Estimation of Mixtures of Two Normal Distributions", Technometrics 9, 15-28.

Davies, R.B. (1977): "Hypothesis Testing when a Nuisance Parameter is Present Only Under the Alternative", Biometrika, 64, 247-254.

Engle, R.F. (1984): Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics", in Z Griliches and M.D. Intriligator (eds.): Handbook of Econometrics, vol II.

Goldberger, A.S. (1983): "Abnormal Selection Bias", in S Karlin, T.Amemiya and L.A. Goodman (eds.): Studies in Econometrics, Time Series, and Multivariate Statistics, Academic Press, London.

Heckman, J. (1979): :Sample Bias as a Specification Error", Econometrica 47, 153-162.

Jennrich, R.I. (1969): "Asymptotic Properties of Non-linear Least Squares Estimators", The Annals of Mathematical Statistics, 40, 633-643.

Kiefer, N.M. (1978):"Discrete Parameter Variation: Efficient Estimation of a Switching Regression Model", Econometrica 46, 427-434.

Quandt, R.E. and J.B. Ramsey (1978): "Estimating Mixtures of Normal Distributions and Switching Regressions", JASA, 73, 730-738.

Robinson, P.M. (1982): "On the Asymptotic Properties of Estimators of Models Containing Limited Dependent Variables", Econometrica 50, 27-41.

Schmidt, P. (1982): "An Improved Version of the Quandt-Ramsey MGF Estimator for Mixtures of Normal Distributions and Switching Regressions", Econometrica 50, 501-516.

Watson, M.W. (1982): "A test for Regression Coefficient Stability when a Parameter is Identified Only Under the Alternative", Harvard Discussion Paper 906.

FOOTNOTES

1. Note that

$$F(x|\bar{\mu}_1, \bar{\mu}_2, \bar{\sigma}_1, \bar{\sigma}_2, \bar{\rho}) = F(x|\bar{\mu}_2, \bar{\mu}_1, \bar{\sigma}_2, \bar{\sigma}_1, 1-\bar{\rho})$$

thus some convention must be adopted in order to have a well defined true parameter vector.

2. In the derivation of (8) and (9) we have used the results

$$\int_\theta^\infty \frac{v}{k}\, \phi(v/k)\, dv = k\phi(\theta/k)$$

and

$$\int_\theta^\infty \frac{v^2}{k^2}\, \phi(v/k)\, dv = \theta k\phi(\theta/k) + k^2[1 - \Phi(\theta/k)].$$

3. A very useful parameterisation to restrict ρ is $\rho = 4\theta^2/(1+\theta^2)^2$.

4. In calculating the derivatives we make use of the result

$$\frac{\partial}{\partial x}\, \phi[f(x)] = -\, f(x)\, \phi[f(x)]\, \frac{\partial f(x)}{\partial x}\ .$$

5. Davies method has been applied to the problem of testing coefficient stability in a regression model by Watson (1982). See also Engle (1984).

TABLE I

Asymptotic Biases (k = 2)

		μ		σ	
μ	$P(y^*>0)$	Censored	Truncated	Censored	Truncated
-2.8	.01	-1.35	0.00	0.83	0.58
-2.4	.02	-1.44	-0.02	0.87	0.58
-2.0	.03	-1.54	-0.13	0.91	0.61
-1.6	.05	-1.55	-0.85	0.91	0.76
-1.2	.08	-1.29	-5.78	0.78	1.60
-0.8	.15	-0.75	-11.70	0.49	2.19
-0.4	.29	-0.30	-12.19	0.20	2.07
0.0	.50	-0.07	-10.99	0.02	1.96
0.4	.71	0.02	-2.34	-0.06	0.69
0.8	.85	0.03	-0.41	-0.08	0.15
1.2	.92	0.03	-0.04	-0.08	-0.03
1.6	.95	0.02	0.05	-0.06	-0.09
2.0	.97	0.01	0.06	-0.04	-0.10
2.4	.98	0.01	0.05	-0.02	-0.08
2.8	.99	0.00	0.04	-0.01	-0.06

TABLE II

Asymptotic Biases (k = 3)

μ	$P(y^*>0)$	μ Censored	μ Truncated	σ Censored	σ Truncated
-2.8	.01	-2.75	0.00	1.55	1.00
-2.4	.02	-2.92	-0.02	1.62	1.00
-2.0	.03	-3.08	-0.21	1.68	1.05
-1.6	.04	-3.03	-1.74	1.66	1.38
-1.2	.07	-2.34	-13.87	1.33	3.11
-0.8	.13	-1.18	-16.18	0.74	2.83
-0.4	.28	-0.41	-14.60	0.27	2.29
0.0	.50	-0.10	-13.28	0.03	2.14
0.4	.72	0.01	-6.28	-0.07	1.43
0.8	.87	0.04	-0.65	-0.11	0.23
1.2	.93	0.04	-0.08	-0.11	-0.02
1.6	.96	0.03	0.05	-0.09	-0.10
2.0	.97	0.02	0.07	-0.07	-0.13
2.4	.98	0.01	0.07	-0.05	-0.12
2.8	.99	0.01	0.05	-0.04	-0.10

TABLE III

Asymptotic Biases (k = 4)

		μ		σ	
μ	$P(y^*>0)$	Censored	Truncated	Censored	Truncated
-2.8	.02	-4.15	0.00	2.20	1.35
-2.4	.02	-4.37	-0.02	2.29	1.35
-2.0	.03	-4.57	-0.29	2.37	1.41
-1.6	.03	-4.41	-2.84	2.30	1.98
-1.2	.06	-3.18	-18.85	1.74	4.00
-0.8	.12	-1.45	-18.61	0.88	3.15
-0.4	.27	-0.48	-15.59	0.31	2.35
0.0	.50	-0.11	-14.00	0.04	2.16
0.4	.73	0.01	-10.09	-0.08	1.95
0.8	.88	0.04	-0.82	-0.12	0.28
1.2	.94	0.04	-0.11	-0.12	-0.02
1.6	.97	0.03	0.04	-0.11	-0.11
2.0	.97	0.02	0.07	-0.09	-0.14
2.4	.98	0.02	0.07	-0.07	-0.14
2.8	.98	0.01	0.06	-0.05	-0.13

TABLE IV

Asymptotic Variance Matrices for Contaminated

Binary Models (c unrestricted)

	β	ρ	c
$\rho = .5,\ c = .05$			
β	6.131		
ρ	0.224	1.668	
c	0.395	-0.293	2.784
$\rho = .5\quad c = .15$			
β	5.380		
ρ	0.214	2.289	
c	1.310	-1.006	9.466
$\rho = .75\quad c = .05$			
β	4.053		
ρ	0.298	2.245	
c	0.249	-0.805	11.126
$\rho = .75\quad c = .15$			
β	3.915		
ρ	0.315	3.098	
c	1.315	-2.745	37.714
$\rho = .9\quad c = .05$			
β	3.305		
ρ	0.324	2.508	
c	0.166	-2.277	69.669
$\rho = .9\quad c = .15$			
ρ	3.282		
ρ	0.356	3.492	
c	1.986	-7.796	236.201

TABLE IV (continued)

$\rho = .1,\ c = .05$

β	15.579		
ρ	0.027	0.400	
c	0.986	-0.039	0.900

$\rho = .1,\ c = .15$

β	7.136		
ρ	0.015	0.591	
c	1.393	-0.149	2.983

$\rho = .25,\ c = .05$

β	10.735		
ρ	0.109	0.926	
c	0.689	-0.107	1.259

$\rho = .25,\ c = .15$

β	7.156		
ρ	0.081	1.295	
c	1.494	-0.384	4.283

TABLE V

CONTAMINATED TOBIT ESTIMATES AND NORMAL TOBIT ESTIMATES

Dependent Variable: Female Weekly hours in paid work[f]

	Normal Tobit	Contaminated Tobit[a1]	Asymmetrical[a2] Contaminated Tob
Constant	29.5830 (1.2200)	33.5103 (1.1689)[b]	23.1534 (2.8754)
Other Income[c]	-.1935 (.0113)	-.2131 (.0132)	-.1302 (.0181)
(Age-40)/10	-3.9967 (.4484)	-4.5596 (.4525)	-2.1306 (.3592)
$(Age-40)^2/100$	-1.9806 (.4239)	-2.3195 (.3888)	-1.1233 (.3287)
(Educ-8)	.7690 (.1808)	.7329 (.2344)	.3136 (.1425)
$(Educ-8)^2$	-.0118 (.0175)	-.0415 (.0223)	.0006 (.0140)
D1[d]	-33.0579 (1.2784)	-35.9424 (1.2577)	-33.8706 (1.1373)
D2	-11.8723 (1.2745)	-13.5010 (1.1452)	-11.7707 (1.0513)
D3	-2.2573 (1.3681)	-2.9802 (1.3166)	-2.1849 (1.1708)
σ[e]	18.1020	18.7995	23.3949
ρ	- -	.7813 (.0321)	.4816 (.0792)
a	-	-	12.5801 (2.2569)
Skewness measure	-	-	-1.4773 (.6324)
$s = (\sigma_1/\sigma_2)^2$	-	.0800	.0700
$-\hat{L}$	7003.212	6978.778	6909.716

Table 5 notes

a1 Estimation conditional to variance ratio s = .08.

a2 Estimation conditional to s = .07

b Standard errors in parentheses.

c Other household income contains husbands income, unearned income and dissaving.

d D_1, D_2 and D_3 are dummy variables representing the presence of preschool children, children of age 5-10 and children of age 11+ respectively.

e Estimated standard deviation of the errors.

f T = 2539.
No. of positive observations = 1460
Percentage of censoring in the sample = 42.5

TABLE VI

s	$-\hat{L}$	$\hat{\rho}$	$\hat{k}$	$(1-\hat{\rho})/SE(\hat{\rho})$
.05	6980.16	.828	1.184	6.68
.08	6978.78	.781	1.227	6.81
.10	6979.11	.757	1.245	6.69
.15	6981.07	.701	1.272	6.22
.20	6983.34	.650	1.281	5.75
.25	6985.58	.604	1.272	5.36
.30	6987.78	.567	1.248	5.06
.40	6992.00	.519	1.177	4.49
.50	6995.71	.497	1.111	3.80
.60	6998.68	.488	1.063	3.02
.70	7000.84	.487	1.031	2.21
.80	7002.24	.490	1.012	1.41
.90	7002.99	.495	1.003	0.77
Normal	7003.21	–	1.	–