A Test for Misspecification in the Censored Normal Model

Forrest D. Nelson

Econometrica, Vol. 49, No. 5. (Sep., 1981), pp. 1317-1329.

Econometrica, Vol. 49, No. 5 (September, 1981)

A TEST FOR MISSPECIFICATION IN THE CENSORED NORMAL MODEL

BY FORREST D. NELSON[1]

Estimates of parameters in Tobit and other models for limited, truncated and censored dependent variables are not robust against misspecification. A test of the standard assumptions against a general misspecified alternative in the univariate censored normal model is derived and extended to the Tobit regression case. Computational ease and freedom from specification of a specific alternative hypothesis are primary attractions of the test.

IT IS WELL KNOWN that ordinary least squares will produce inconsistent estimates of the regression parameters if the dependent variable is censored or truncated.[2] Maximum likelihood estimation on Tobit and other limited dependent variable models is being employed with increasing frequency to avoid this inconsistency. But the assumptions required of these models are quite strong and any violation, such as heteroscedasticity or nonnormality, may result in an asymptotic bias as severe as in the naive OLS formulation.[3]

The purpose of this paper is to suggest a general test for misspecification in these models. Section 1 introduces the simple nonregression case of a censored variable. Likelihood equations for the location and scale parameters are obtained and the method of moments estimator is discussed. A specification test following Hausman [5] is then derived in Section 2 for the general alternative hypothesis of misspecification. Section 3 contains a generalization of the model and the specification test to the case of a regression formulation. The results are summarized in Section 4.

1. THE MODEL AND MOM AND ML ESTIMATORS

We consider in this section the case of a censored normal variate y defined by the distribution function

$$F(y) = \Phi\left(\frac{y-\mu}{\sigma}\right) \qquad \text{for } y \geqslant 0,$$

$$= 0 \qquad \text{for } y < 0,$$

[1] Richard Rosett prompted the author's interest in misspecification of Tobit models and the work presented here benefited from earlier collaboration with G. S. Maddala. The paper was written while the author was on the faculty at California Institute of Technology.

[2] In the statistical literature, the term censored applies to a sample in which some observations are recorded only as above (or below) some threshold, the exact value in such a case having been censored. The term truncated is applied to samples in which such observations are excluded entirely. In the econometrics literature, on the other hand, the term truncated is often applied to the censored sample case, apparently in reference to the variable rather than the sample. This note will follow the statistical usage.

[3] Hausman and Wise [6] have noted inconsistencies arising from misspecification in probit-logit models. The effect of heteroscedasticity has been examined by Maddala and Nelson [11] and by Maddala [10] in the case of the Tobit model and by Hurd [7] in a truncated variable model. Nelson [12] provides a more detailed analysis of the effect of heteroscedasticity, and Goldberger [3] examines the effect of nonnormality.

where $\Phi(a)$ is the unit normal cumulative density function (c.d.f.)

$$\Phi(a)=\int_{-\infty}^{a}\phi(u)\,du=\int_{-\infty}^{a}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^{2}\right)du.$$

The c.d.f. allows for a mass point at zero so that a sufficiently large sample will contain some observations on y at the threshold of zero. While this distribution has received considerable attention in the statistical literature,[4] it will prove useful to review here the method of moments and maximum likelihood estimators.

Let $(y_1, \ldots, y_N)$ be a random sample of observations on y and define the random variable v_i as

$$v_i=\begin{cases}1 & \text{if } y_i>0,\\ 0 & \text{if } y_i=0.\end{cases}$$

Then the sample fraction of noncensored observations and the first two sample moments (about zero) are given by

$$P=\sum v_i/N,$$

$$M_1=\sum y_i/N=\sum v_i y_i/N,$$

and

$$M_2=\sum y_i^2/N=\sum v_i y_i^2/N,$$

respectively. Since the third and fourth population moments exist, these three sample moments will converge to their population counterparts[5] as N tends to infinity:

$$P\xrightarrow{p}\operatorname{pr}(y>0)=1-\Phi\left(\frac{-\mu}{\sigma}\right)=\Phi\left(\frac{\mu}{\sigma}\right)\equiv\Phi, \tag{1.1}$$

$$M_1\xrightarrow{p}E(y;\,\mu,\sigma)=\mu\Phi\left(\frac{\mu}{\sigma}\right)+\sigma\phi\left(\frac{\mu}{\sigma}\right)\equiv E_1, \tag{1.2}$$

$$M_2\xrightarrow{p}E(y^2;\,\mu,\sigma)=\mu^2\Phi\left(\frac{\mu}{\sigma}\right)+\sigma^2\Phi\left(\frac{\mu}{\sigma}\right)+\mu\sigma\phi\left(\frac{\mu}{\sigma}\right)\equiv E_2. \tag{1.3}$$

MOM estimates $\tilde{\mu}$ and $\tilde{\sigma}$ are found by replacing E_1 and E_2 on the right-hand side of (1.2) and (1.3) by M_1 and M_2 respectively, substituting $\tilde{\mu}$ and $\tilde{\sigma}$ for μ and σ on the left, and solving for $\tilde{\mu}$ and $\tilde{\sigma}$. Those equations are nonlinear so that iterative solution procedures are typically used.[6] These MOM estimators are consistent but not asymptotically efficient.

[4] See, for example, Cohen [**2**], Hald [**4**], and Pearson and Lee [**13**].

[5] See Johnson and Kotz [**8**] for a derivation of the moments. Note that the three relations, Φ, E_1 and E_2, depend on only two parameters, μ and σ, and are thus not independent, a detail implicitly recognized in maximum likelihood estimation but not in the method of moments.

[6] Alternative linear estimators are found by solving in turn the three equations:

The log likelihood is given by

$$\text{(1.4)}\qquad \log L[\mu,\sigma;(y_1,\ldots,y_N)] = c + \sum_{i=1}^{N}\left\{(1-v_i)\log\left[1-\Phi\left(\frac{\mu}{\sigma}\right)\right] - v_i\log\sigma - \frac{1}{2}v_i(y_i-\mu)^2/\sigma^2\right\}.$$

The likelihood equations can be written as

$$\text{(1.5)}\qquad \hat{\mu}P + \hat{\sigma}\frac{\hat{\phi}}{1-\hat{\Phi}}(1-P) = M_1,$$

$$\text{(1.6)}\qquad \hat{\mu}^2P + \hat{\sigma}^2P + \hat{\mu}\hat{\sigma}\frac{\hat{\phi}}{1-\hat{\Phi}}(1-P) = M_2,$$

where $\hat{\phi} \equiv \phi(\hat{\mu}/\hat{\sigma})$ and $\hat{\Phi} \equiv \Phi(\hat{\mu}/\hat{\sigma})$. The maximum likelihood estimators are the nonlinear solutions of (1.5) and (1.6) for $\hat{\mu}$ and $\hat{\sigma}$.

Second derivatives of the log likelihood divided by N are given by

$$H(\mu,\sigma;M_1,M_2,P) \equiv \frac{\partial^2 1/N\log L}{\partial(\mu,\sigma)'\partial(\mu,\sigma)} = \frac{-1}{\sigma^2}\begin{bmatrix} a & b \\ b & c\end{bmatrix}$$

where

$$a = \frac{1}{\sigma}\frac{\phi}{(1-\Phi)^2}(1-P)[\mu - E_1] - P,$$

$$b = \frac{1}{\sigma^2}\frac{\phi}{(1-\Phi)^2}(1-P)[\mu^2+\sigma^2-E_2] - 2\frac{1}{\sigma}(M_1 - P\mu),$$

$$c = \frac{1}{\sigma^2}\frac{\mu}{\sigma}\frac{\phi}{(1-\Phi)^2}(1-P)[\mu^2+\sigma^2-E_2] + P - 3\frac{1}{\sigma^2}[M_2 - 2\mu M_1 + P\mu^2],$$

$$\tilde{\gamma} = \left(\frac{\tilde{\mu}}{\sigma}\right) = \Phi^{-1}(P),$$

$$\tilde{\mu} = M_1\Big/\left(P + \frac{1}{\tilde{\gamma}}\phi(\tilde{\gamma})\right),$$

$$\tilde{\sigma} = \tilde{\mu}/\tilde{\gamma},$$

or the three equations:

$$\tilde{\gamma} = \left(\frac{\tilde{\mu}}{\sigma}\right) = \Phi^{-1}(P),$$

$$\tilde{\sigma} = [M_2/(\tilde{\gamma}^2P + P + \tilde{\gamma}\phi(\tilde{\gamma}))]^{1/2},$$

$$\tilde{\mu} = \tilde{\gamma}\tilde{\sigma}.$$

and E_1 and E_2 are the first and second moments defined by equations (1.2) and (1.3) and ϕ and Φ are both evaluated at μ/σ. The information matrix is the negative of the expectation of H and can be written as

$$(1.7) \qquad \mathcal{I}(\mu,\sigma) = \frac{1}{\sigma^2} \times \begin{bmatrix} d & e \\ e & f \end{bmatrix}$$

where

$$(1.8) \qquad d = \Phi - \frac{1}{\sigma}\frac{\phi}{1-\Phi}(\mu - E_1),$$

$$(1.9) \qquad e = \frac{1}{\sigma^2}\frac{\phi}{1-\Phi}\left[\sigma^2 + \mu^2 - E_2\right],$$

and

$$(1.10) \qquad f = 2\Phi - \frac{1}{\sigma^2}\frac{\mu}{\sigma}\frac{\phi}{1-\Phi}\left[\sigma^2 + \mu^2 - E_2\right].$$

The inverse of $\mathcal{I}$ is the covariance matrix for $(\hat{\mu},\hat{\sigma})'$. It may be estimated by $-H(\hat{\mu},\hat{\sigma};M_1,M_2,P)^{-1}$ or, perhaps better, by $\mathcal{I}(\hat{\mu},\hat{\sigma})^{-1}$.

A proof of the consistency, asymptotic normality, and asymptotic efficiency (i.e., that $AC(\hat{\mu},\hat{\sigma}) = \mathcal{I}^{-1}$) is provided by Amemiya [**1**] for the more general case of a regression model formulation. Inspection of the likelihood equations (1.5) and (1.6) reveals the nature of the consistency. Since P, M_1, and M_2, converge to $\Phi(\mu/\sigma)$, E_1, and E_2 respectively, solution of equations (1.5) and (1.6) requires $\hat{\mu} = \mu$ and $\hat{\sigma} = \sigma$ in the limit. The efficiency gain of $\hat{\mu}$ and $\hat{\sigma}$ over the MOM estimators $\tilde{\mu}$ and $\tilde{\sigma}$ arises from the use of additional sample information, namely P, in the MLE's.

It may be the case that parameters of interest are not the location and scale parameters μ and σ but rather some sample moment(s) or the probability of a noncensored observation. In this case the sample moments themselves are the consistent MOM estimates of the population moments and P is consistent for $\Phi(\mu/\sigma)$. But they lack asymptotic efficiency relative to the maximum likelihood estimates $\hat{E}_1 = E(y;\,\hat{\mu},\hat{\sigma})$, $\hat{E}_2 = E(y^2;\,\hat{\mu},\hat{\sigma})$, and $\hat{\Phi} = \Phi(\hat{\mu}/\hat{\sigma})$. Again the gain in efficiency arises from use of more sample information and implicit recognition of the dependency among those three parameters.

The above noted asymptotic properties of the ML and MOM estimators depend crucially on the assumption of an i.i.d. censored normal sample. If any of these assumptions are violated, for example if the distribution is not normal or if all observations do not come from the same distribution, then the sample statistics, P, M_1, and M_2 will not, in general, converge to the relations given by equation (1.1), (1.2), and (1.3) respectively. Solution of any two of these three equations for the MOM estimates $\tilde{\mu}$ and $\tilde{\sigma}$ or the likelihood equations (1.5) and (1.6) for the MLE's $\hat{\mu}$ and $\hat{\sigma}$ will therefore yield inconsistent estimates.[7]

[7] Apparently the MLE's are reasonably robust against failure of the independence assumption. Lee [**9**] has established strong consistency in the regression case with serially correlated disturbances.

To measure the extent of the bias, Nelson [12] examined the asymptotic solution for the MLE's from the misspecified i.i.d. censored normal model when the sample was generated from a mixture of censored normal distributions with common location parameter μ but different scale parameters σ_1 and σ_2. As an example of those results, true parameter values of $\mu = -.5$, $\sigma_1/\sigma_2 = 2.0$, $(\sigma_1^2 + \sigma_2^2)/2 = 1$ led to asymptotic biases of $-.164$ for $\hat{\mu}$, $-.0042$ for $\hat{\Phi}$, .0050 for $\hat{E}_1$, and $-.0085$ for $\hat{E}_2$. Such results suggest that estimates of the moments Φ, E_1, and E_2 may not be too sensitive to misspecification but that the nonrobustness of the location and scale parameter estimates may be quite severe.

2. AN ASYMPTOTIC TEST AGAINST MISSPECIFICATION

The sensitivity of MLE's to specification error motivates a search for some reasonably general test. We suggest in this section an asymptotic specification test derived from the work of Hausman [5]. Hausman's procedure may be outlined as follows. Let $\hat{\theta}_0$ and $\hat{\theta}_1$ be two estimators of the parameter vector θ such that under the null hypothesis, H_0, they are both consistent and asymptotically normal with asymptotic variances V_0 and V_1. Further, let $\hat{\theta}_0$ be asymptotically efficient so that $V_0 = \mathcal{I}^{-1}$ and $V_1 - V_0$ is nonnegative definite. Then, as Hausman shows, $\sqrt{N}\, q = \sqrt{N}(\hat{\theta}_1 - \hat{\theta}_0)$ is asymptotically normal with variance $V_1 - V_0$. Letting $\hat{V}_1$ and $\hat{V}_0$ be consistent for V_1 and V_0 respectively, he constructs the statistic $m = Nq'(\hat{V}_1 - \hat{V}_0)^{-1}q$ which, he argues, is asymptotically $\chi^2_{(K)}$ under H_0, where K is the dimension of θ.

Consider now an alternative hypothesis, H_a, such that, under H_a, $\text{plim}\,\hat{\theta}_0 \neq \theta$ but $\text{plim}\,\hat{\theta}_1 = \theta$. Under these conditions q does not converge to zero and the distribution of m is skewed to the right relative to the chi-square. Thus m serves as a test statistic, with the hypothesis of no misspecification to be rejected with large values of m. Power considerations require knowledge of the distribution of m under the alternative hypothesis of specification error. Often this distribution will be asymptotically noncentral chi-square; Hausman gives some fairly general conditions under which this will be the case.

The apparent attractions of Hausman's asymptotic test are the ease with which the variance of q may be obtained and the generality of the procedure. As regards the latter, the test is, simultaneously, against all alternatives under which $\hat{\theta}_1$ is consistent but $\hat{\theta}_0$ inconsistent, though of course the power of the test will vary with H_a. Thus a particular alternative hypothesis need not be fully specified —all that is needed is an asymptotically efficient estimator and a second consistent but inefficient estimator which exhibits a fair degree of robustness.

The test appears particularly apt for the censored normal problem of Section 1 above. There we have a maximum likelihood estimator with all the desired asymptotic properties under the maintained assumptions but which may exhibit severe bias under a variety of seemingly innocuous misspecifications. We will, in what follows, adapt Hausman's test to this case.

Regardless of the parameters of interest, maximum likelihood estimation yields the estimators $\hat{\mu}$ and $\hat{\sigma}$ as either an intermediate or a final step. This vector $(\hat{\mu}, \hat{\sigma})'$

exhibits the necessary properties of the efficient estimator $\hat{\theta}_0$, but there does not exist a robust estimator of $\theta = (\mu, \sigma)'$ to serve the role of $\hat{\theta}_1$. For example, the MOM estimator, $(\tilde{\mu}, \tilde{\sigma})'$, noted in Section 1 is subject to the same sensitivity to misspecification as is the MLE.[8] The first two sample moments, M_1 and M_2, on the other hand are, under very general conditions, consistent for the first two population moments of whatever population is being sampled. And the MLE for these moments, $\hat{E}_1 = E(y; \hat{\mu}, \hat{\sigma})$ and $\hat{E}_2 = E(y^2; \hat{\mu}, \hat{\sigma})$, as obtained from the invariance property, serves as the efficient counterpart. Furthermore we should note that the sample proportion of noncensored observations, P, is, again under general conditions and random sampling, consistent for the corresponding population fraction. It's efficient counterpart is $\hat{\Phi} = \Phi(\hat{\mu}/\hat{\sigma})$. We thus have three population parameters for which both efficient and robust estimators are readily available.

One might reason intuitively that since the censored normal distribution has only two parameters, the test statistic can and should be constructed from only two of the three available estimator pairs—use of all three would surely result in a singularity in the variance-covariance matrix for q. As will be shown directly, the problem is even more severe. Define q_1, q_2, and q_3 as

$$q_1 \equiv P - \Phi(\hat{\mu}/\hat{\sigma}), \tag{2.1}$$

$$q_2 \equiv M_1 - E(y; \hat{\mu}, \hat{\sigma}) = \frac{\hat{\mu} - \hat{E}_1}{1 - \hat{\Phi}}(P - \hat{\Phi}), \tag{2.2}$$

$$q_3 \equiv M_2 - E(y^2; \hat{\mu}, \hat{\sigma}) = \frac{\hat{\mu}^2 + \hat{\sigma}^2 - \hat{E}_2}{1 - \hat{\Phi}}(P - \hat{\Phi}), \tag{2.3}$$

where the equalities in (2.2) and (2.3) are obtained after substitution from equations (1.2) and (1.5), and (1.3) and (1.6) respectively. Consider the expansions of q_2 and q_3 about μ and σ. We obtain for q_2:

$$q_2 = \frac{\mu - E_1}{1 - \Phi}(P - \Phi) - \frac{P - \Phi}{1 - \Phi}(\hat{E}_1 - E_1) + \frac{\mu - E_1}{1 - \Phi}\,\frac{P - \Phi}{1 - \Phi}(\hat{\Phi} - \Phi)$$
$$- \frac{\mu - E_1}{1 - \Phi}(\hat{\Phi} - \Phi) + \frac{P - \Phi}{1 - \Phi}(\hat{\Phi} - \Phi) + \frac{P - \Phi}{1 - \Phi}(\hat{\mu} - \mu) + R_2$$

where R_2 includes all second and higher order terms. Note that R_2 and all terms like $(P - \Phi) \cdot (\hat{E}_1 - E_1)$ are of smaller order than $N^{-1/2}$ since P, $\hat{E}_1$, $\hat{\mu}$, and $\hat{\Phi}$ are all consistent under H_0. Thus q_2 may be simplified to

$$q_2 = \frac{\mu - E_1}{1 - \Phi}(P - \hat{\Phi}) + o(N^{-1/2}). \tag{2.4}$$

[8] Hausman's condition that $\hat{\theta}_1$ be consistent under H_a may be stronger than necessary—his test might serve well, so long as plim $\hat{\theta}_1 \neq$ plim $\hat{\theta}_0$ under H_a. In the present case, that would mean the test could be based on $(\hat{\mu}, \hat{\sigma})$ and $(\tilde{\mu}, \tilde{\sigma})$. We have not investigated that possibility since $(\tilde{\mu}, \tilde{\sigma})$ are computationally more difficult than other statistics we can use.

Similarly for q_3 we obtain

$$q_3 = \frac{\mu^2+\sigma^2-E_2}{1-\Phi}(P-\Phi) - \frac{\mu^2+\sigma^2-E_2}{1-\Phi}(\hat{E}_2-E_2)(P-\Phi)$$

$$-\frac{\mu^2+\sigma^2-E_2}{1-\Phi}(\hat{\Phi}-\Phi) + \frac{\mu^2+\sigma^2-E_2}{(1-\Phi)^2}(\hat{\Phi}-\Phi)(P-\Phi)$$

$$+\frac{2\mu}{1-\Phi}(\hat{\mu}-\mu)(P-\Phi) + \frac{2\sigma}{1-\Phi}(\hat{\sigma}-\sigma)(P-\Phi) + R_2.$$

Again, consistency of $\hat{E}_2$, P, $\hat{\Phi}$, $\hat{\mu}$, and $\hat{\sigma}$ allows simplification to

$$(2.5)\qquad q_3 = \frac{\mu^2+\sigma^2-E_2}{1-\Phi}(P-\hat{\Phi}) + o(N^{-1/2}).$$

In the limit, then, q_2 and q_3 are constant multiples of $q_1 = (P-\hat{\Phi})$ so that the asymptotic covariance matrix $V(q)$, where $q = (q_1, q_2, q_3)'$, must have rank one.

The Hausman article failed to acknowledge the possibility that $V(\hat{\theta}_1) - V(\hat{\theta}_0)$ might sometimes or always be singular in a particular application. But the resolution of such a difficulty is obvious—base the test on some subset of the estimator pairs which is not perfectly colinear. In the case at hand we will choose the estimator pair $(\hat{\Phi}, P)$ on computational grounds, but in fact it makes little difference which of the three we choose.

The next step is to obtain the asymptotic variance of $P-\hat{\Phi}$. Rather than compute it directly, we will obtain it, as did Hausman, from $V(P) - V(\hat{\Phi})$. P is of course binomial and $\sqrt{N}(P-\Phi) \sim \mathrm{AN}(0, \Phi(1-\Phi))$, so that the asymptotic variance of P is $V(P) = \Phi(1-\Phi)$.

The asymptotic distribution of $\hat{\Phi}$ and, for completeness, $\hat{E}_1$ and $\hat{E}_2$, are obtained as follows. Expand each of the three terms in a first-order Taylor series about (μ, σ). (Consistency guarantees that higher order terms are $o(N^{-1/2})$ so they may be neglected). We obtain

$$(2.6)\qquad \hat{\Phi} - \Phi \simeq (\hat{\mu}-\mu)\phi\frac{1}{\sigma} - (\hat{\sigma}-\sigma)\phi\frac{\mu}{\sigma^2},$$

$$(2.7)\qquad \hat{E}_1 - E_1 \simeq (\hat{\mu}-\mu)\Phi + (\hat{\sigma}-\sigma)\phi,$$

$$(2.8)\qquad \hat{E}_2 - E_2 \simeq (\hat{\mu}-\mu)\cdot 2\cdot(\mu\Phi+\sigma\phi) + (\hat{\sigma}-\sigma)\cdot 2\cdot \sigma\Phi.$$

Each of the three statistics times $\sqrt{N}$ will, in the limit, follow the same asymptotic normal distribution as the respective linear combination of $\sqrt{N}(\hat{\mu}-\mu)$ and $\sqrt{N}(\hat{\sigma}-\sigma)$. That is,

$$\sqrt{N}\begin{bmatrix}\hat{\Phi}-\Phi\\ \hat{E}_1-E_1\\ \hat{E}_2-E_2\end{bmatrix} \sim \mathrm{AN}(0, A'\mathscr{I}^{-1}A)$$

where $\mathcal{I}$ is the information matrix defined in (1.7) and A is given by

$$(2.9)\qquad A=\begin{bmatrix} \frac{1}{\sigma}\phi & \Phi & 2\cdot E_1 \\ -\frac{1}{\sigma}\frac{\mu}{\sigma}\phi & \phi & 2\cdot\sigma\Phi \end{bmatrix}.$$

In particular the asymptotic variance of $\hat{\Phi}$ is given by

$$(2.10)\qquad V(\hat{\Phi})=\left(\frac{1}{\sigma}\phi\right)^2 V(\hat{\mu})+\left(\frac{1}{\sigma}\frac{\mu}{\sigma}\phi\right)^2 V(\hat{\sigma})-2\frac{1}{\sigma^2}\frac{\mu}{\sigma}\phi^2\operatorname{cov}(\hat{\mu},\hat{\sigma}).$$

In principle, any consistent estimators of $V(P)$ and $V(\hat{\Phi})$ may be employed in construction of the test statistic. The following variance estimator is guaranteed to be positive, and experimentation suggests that it serves the purpose well:[9]

$$(2.11)\qquad \hat{V}(P-\hat{\Phi})=\hat{\Phi}\cdot(1-\hat{\Phi})-\left[\frac{1}{\hat{\sigma}}\hat{\phi}\;\; -\frac{1}{\hat{\sigma}}\frac{\hat{\mu}}{\hat{\sigma}}\hat{\phi}\right]\left[\mathcal{I}(\hat{\mu},\hat{\sigma})\right]^{-1}\begin{bmatrix}\frac{1}{\hat{\sigma}}\hat{\phi}\\ -\frac{1}{\hat{\sigma}}\frac{\hat{\mu}}{\hat{\sigma}}\hat{\phi}\end{bmatrix}$$

where $\mathcal{I}$ is defined by equations (1.7)–(1.10).

We have, then, the following result which defines the asymptotic specification test. Under the maintained hypothesis of a sample from an i.i.d. censored normal population with location and scale parameters μ and σ, the statistic

$$(2.12)\qquad m=N\cdot(P-\hat{\Phi})^2/\hat{V}(P-\hat{\Phi})$$

follows, asymptotically, a χ^2 distribution with one degree of freedom.

The power characteristics of the test under various alternative hypotheses are not derived here. But we do offer, as evidence on the efficacy of the test, the following results from some simple simulation experiments. Six experiments were run under varying conditions with respect to sample size, location parameters, and degree of misspecification. In the first of the experiments the model was correctly specified, while the next five involved a heteroscedastic misspecification. In each experiment, two samples of size $N/2$ were drawn randomly from $N(\mu,\sigma_1^2)$ distributions, the two subsamples were combined and censored at zero, ML estimates $\hat{\mu}$ and $\hat{\sigma}$ were obtained under the i.i.d. censored normal assumption, and the statistic m was computed. This process was repeated fifty times (100 in the correctly specified experiment) to obtain fifty (100) observations on the statistic m under the prespecified structure. The six experiments differed in sample size N (100, 250, 500 or 1000), and the location parameter μ ($-.5$ or $+.5$). In all five misspecified experiments, the two population scale parameters

[9] Use of $P(1-P)$ in place of $\hat{\Phi}\cdot(1-\hat{\Phi})$ and/or $-H^{-1}$ in place of $\hat{\mathcal{I}}^{-1}$ will yield the same asymptotic results but produce the unesthetic small sample result of occasional negative variance estimates.

were fixed at $\sigma_1 = .6325$ and $\sigma_2 = 1.2649$, corresponding to $\lambda = \sigma_1/\sigma_2 = 2$ and $(\sigma_1^2 + \sigma_2^2)/2 = 1$.

Table I summarizes the results of those six experiments. For each experiment the table contains the nine decile values for the statistic; its mean and variance; the proportion of the sample exceeding critical χ^2_α values for tests with $\alpha = .01$, .05, .10, and .25; and computed asymptotic values for $\bar{\hat{\Phi}} - \bar{P}$ and $\bar{\hat{\mu}} - \mu$. A column containing relevant parameters for the $\chi^2_{(1)}$ distribution is included as a benchmark.

The results from experiment "H_0" suggest that with no misspecification the statistic m fits the $\chi^2_{(1)}$ distribution reasonably well even for the moderate sample size of 100. With large samples the test seems quite effective at detecting the employed degree of misspecification—the null hypothesis is rejected at $\alpha = .05$ in 48 of the 50 samples in experiment "H_5" with $N = 1000$ and 23 of the 50 samples in experiment "H_3" with $N = 500$. For smaller sample sizes the results are less encouraging—rejection rates at $\alpha = .05$ are 6/50 and 3/50 in the two misspecified experiments with $N = 100$ and 12/50 in the one with $N = 250$.

TABLE I

PERFORMANCE OF TEST STATISTIC $m = N(P - \hat{\Phi})^2/\hat{V}(P - \hat{\Phi})$ IN SIX SAMPLING EXPERIMENTS

	Experiment						
	H_0	H_1	H_2	H_3	H_4	H_5	$\chi^2_{(1)}$
Experiment Structure							
$\lambda = \sigma_1/\sigma_2$	1	2	2	2	2	2	
μ	−.5	−.5	−.5	−.5	+.5	+.5	
N	100	100	250	500	100	1000	
Number of Samples	100	50	50	50	50	50	
Sampling Statistics							
Mean of m	1.38	1.92	2.46	5.32	1.56	18.34	1
Variance of m	7.46	13.65	8.01	22.65	7.41	95.26	2
mean of $(\hat{\mu} - \mu)$	−.007	−.150	−.148	−.167	.013	.018	
mean of $(P - \hat{\Phi})$	−.0013	.0028	.0031	.0042	.0101	.0181	
Decile Values for m							
.9	3.268	8.463	5.962	11.262	3.656	33.33	2.705
.8	2.043	2.436	3.8994	8.853	2.040	28.52	1.641
.7	1.360	1.181	3.016	7.241	1.561	20.88	1.074
.6	.963	.871	2.385	4.709	1.201	19.27	.708
.5	.419	.503	1.417	3.306	.740	16.24	.458
.4	.287	.387	.888	2.856	.455	14.97	.276
.3	.089	.176	.460	2.165	.090	10.96	.148
.2	.037	.033	.128	1.467	.040	9.52	.065
.1	.007	.006	.055	.409	.015	6.76	.016
Rejection Rates ($\%m > \chi^2_{1,\alpha}$)							
% > 1.32	33	24	52	82	36	100	25
% > 2.71	13	18	38	64	18	98	10
% > 3.84	7	12	24	46	6	96	5
% > 6.63	2	10	10	32	4	92	1

3. THE EXTENSION TO A REGRESSION MODEL

Section 2 introduced a specification test for the case of an i.i.d. censored normal random variate. We sketch here the extension to a regression model.

Let X_i be a k-element vector of exogenous variables, β be a k-element vector of unknown regression parameters, and specify

$$F(y_i) = \Phi\left(\frac{y_i - \beta' X_i}{\sigma}\right) \quad \text{for } y_i \geqslant 0, \qquad (3.1)$$

$$= 0 \qquad \text{for } y_1 < 0.$$

This is of course the Tobit model more commonly described by

$$y_i = \beta' X_i + u_i \quad \text{if RHS} > 0,$$

$$= 0 \qquad \text{otherwise;}$$

$$u_i \sim \text{IN}(0, \sigma^2).$$

The likelihood for a random sample of size N is given by equation (1.4) with μ replaced by $\beta' X_i$.

Define X as the $N \times K$ matrix containing X_i' in the ith row; Y as the $N \times 1$ vector with typical element y_i; W as the $N \times N$ diagonal matrix containing the indicator variable, $w_{ii} = 1$ if $y_i > 0$, 0 otherwise, along the diagonal; $\underline{\phi}$ as the $N \times 1$ vector with $\phi(\beta' X_i/\sigma)$ at element i; and $\overline{\underline{\Phi}}$ as the $N \times N$ diagonal matrix with $\Phi(\beta' X_i/\sigma)$ at position ii. When $\underline{\phi}$ and $\overline{\underline{\Phi}}$ are evaluated at the MLE's $\hat{\beta}$ and $\hat{\sigma}$, they will be indicated as $\hat{\underline{\phi}}$ and $\hat{\overline{\underline{\Phi}}}$ respectively. Otherwise they will be evaluated at the true values, β_0 and σ_0.

Now the likelihood equations may be written, after simplification, as

$$X' W X \hat{\beta} + \hat{\sigma} X' [I - W][I - \hat{\overline{\underline{\Phi}}}]^{-1} \hat{\underline{\phi}} = X' Y \qquad (3.2)$$

and

$$\tilde{\beta}' X' W X \hat{\beta} + \hat{\sigma}^2 \text{tr}[W] + \hat{\sigma} \hat{\beta}' X' [I - W][I - \hat{\overline{\underline{\Phi}}}]^{-1} \hat{\underline{\phi}} = Y' Y. \qquad (3.3)$$

So long as the y_i's are random with distribution as specified in (3.1) and the sequence X_i is such that

$$\lim_{N \to \infty} \frac{1}{N} X' X = Q \text{ pos. def.},$$

solution of (3.2) and (3.3) will yield estimates which are consistent, asymptotically normal, and asymptotically efficient. That is,

$$\sqrt{N}\left[\begin{pmatrix} \hat{\beta} \\ \hat{\sigma} \end{pmatrix} - \begin{pmatrix} \beta \\ \sigma \end{pmatrix}\right] \sim \text{AN}\left[0, \lim_{N \to \infty} \mathcal{I}(\beta, \sigma)^{-1}\right]$$

with $\mathcal{I}$ defined as

$$\mathcal{I}(\beta,\sigma) = \frac{1}{\sigma^2}\frac{1}{N}\begin{bmatrix} X'[C+\underline{\Phi}]X & X'[\underline{\phi}-CB] \\ [\underline{\phi}-CB]'X & B'CB - B'\underline{\phi} + 2\operatorname{tr}[\underline{\Phi}] \end{bmatrix} \tag{3.4}$$

where C is an $N \times N$ diagonal matrix with typical diagonal element

$$c_{ii} = \frac{\phi(\beta'X_i/\sigma)^2}{1-\Phi(\beta'X_i/\sigma)} - \frac{\beta'X_i}{\sigma}\phi\left(\frac{\beta'X_i}{\sigma}\right)$$

and B is an N element vector with typical element $b_i = (\beta'X_i)/\sigma$.

Violation of any of the distributional assumptions will in general lead to an inconsistent estimator. We seek then a general test for those assumptions. The test we use is again that proposed by Hausman, based this time on estimates of $E((1/N)X'Y)$.[10] Under fairly general conditions on X_i and the distribution of y_i, $(1/N)X'Y$ will be consistent for its expectation. Under the maintained assumptions for the censored normal regression model, it will be consistent and asymptotically normal though inefficient. Taking X_i as fixed, the first two moments of y_i are given by

$$E(y_i;\ \beta,\sigma) = \beta'X_i\Phi\left(\frac{\beta'X_i}{\sigma}\right) + \sigma\phi\left(\frac{\beta'X_i}{\sigma}\right) \tag{3.5}$$

and

$$E(y_i^2;\ \beta,\sigma) = (\beta'X_i)^2\Phi\left(\frac{\beta'X_i}{\sigma}\right) + \sigma^2\Phi\left(\frac{\beta'X_i}{\sigma}\right) + \beta'X_i\sigma\phi\left(\frac{\beta'X_i}{\sigma}\right). \tag{3.6}$$

Thus

$$E_{xy} \equiv E\left(\frac{1}{N}X'Y;\ \beta,\sigma\right) = \frac{1}{N}\left[X'\underline{\Phi}X\beta + \sigma X'\underline{\phi}\right] \tag{3.7}$$

and the variance of $(1/N)X'Y$ is

$$V_1 \equiv V\left(\frac{1}{N}X'Y;\ \beta,\sigma\right) = \frac{1}{N}X'V_YX \tag{3.8}$$

where V_y is an $N \times N$ diagonal matrix with diagonal elements $E(y_i^2) - E(y_i)^2$ as defined in (3.5) and (3.6). Thus,

$$\sqrt{N}\left(\frac{1}{N}X'Y - E_{XY}\right) \sim \mathrm{AN}\left(0, \lim_{n\to\infty} V_i\right).$$

$(1/N)X'Y$ is the consistent but inefficient estimator we require for the test statistic and its variance is given by expression (3.8).

[10] As before, statistics for $E(Y'Y)$ and $\operatorname{tr}(\underline{\Phi})$ might be included as well but would involve a singularity in the asymptotic var-cov matrix for the difference vector. Of the $K+2$ possible statistic pairs, we must choose only k.

The corresponding efficient estimator is the maximum likelihood estimator for E_{XY}. Define the statistic $\hat{E}_{XY}$ as expression (3.7) evaluated at the MLE's $\hat{\beta}$ and $\hat{\sigma}$. Its variance is obtained by expanding it about β and σ,

$$\hat{E}_{XY} - E_{XY} = \frac{1}{N}\left[X'\underline{\Phi}X(\hat{\beta}-\beta) + X'\underline{\phi}(\hat{\sigma}-\sigma)\right] + o(N^{-1/2}). \tag{3.9}$$

The left side of (3.9) will thus have the same asymptotic distribution as the indicated linear combination of $(\hat{\beta}-\beta)$ and $(\hat{\sigma}-\sigma)$. That is,

$$\sqrt{N}\,(\hat{E}_{XY} - E_{XY}) \sim \mathrm{AN}\Big(0, \lim_{N\to\infty} V_0\Big)$$

where V_0 is defined by

$$V_0 = \frac{1}{N^2}\left[X'\underline{\Phi}X\;X'\underline{\phi}\right]\mathcal{I}(\beta,\sigma)^{-1}\begin{bmatrix} X'\underline{\Phi}X \\ \underline{\phi}'X \end{bmatrix}. \tag{3.10}$$

Combining these results, we obtain the desired test statistic,

$$m = N\Big(\frac{1}{N}X'Y - \hat{E}_{XY}\Big)'(\hat{V}_1 - \hat{V}_0)^{-1}\Big(\frac{1}{N}X'Y - \hat{E}_{XY}\Big) \tag{3.11}$$

where $\hat{V}_1$ and $\hat{V}_0$ are obtained by evaluation of (3.8) and (3.10) respectively at the MLE's $\hat{\beta}$ and $\hat{\sigma}$.[11] Under the maintained assumptions, this statistic will follow, asymptotically, a $\chi^2_{(k)}$ distribution.

4. SUMMARY

The Tobit model and maximum likelihood estimation of it are being employed with increasing frequency in economics and other areas. The assumptions of that model are quite strong, and more attention must be paid to the effect of violating those assumptions to avoid erroneous inferences. We have argued above that MLE's for this model lack robustness against misspecification.

Given this sensitivity, some general test against misspecification would be most helpful. Such a test was developed along the lines of the asymptotic test proposed by Hausman. That test requires two estimators: One exhibiting consistency and asymptotic efficiency under the null hypothesis and inconsistency under misspecification, and the other exhibiting consistency under the alternative as well as the null hypothesis. The estimators proposed for this test are, respectively, the maximum likelihood and the method of moments estimators for certain population moments.

The suggested test statistics are given by expressions (2.12) and (3.11) for the nonregression and regression cases respectively. Consistent estimators of the required asymptotic covariance matrices are suggested which will be positive definite even with finite samples. The performance of the test statistic in the

[11] Again there exist other consistent estimators for V_1 and V_0, use of $-H^{-1}$ in (3.9) for example, but they will not guarantee a positive definite variance estimate for the difference.

nonregression case was examined by Monte Carlo methods at the end of Section 2. The results suggested that the test statistic fits its asymptotic χ^2 distribution reasonably well even for moderate sample sizes and was quite effective in detecting a heteroscedastic misspecification in samples greater than 500. The test appears to exhibit rather weak power, however, with smaller sample sizes.

The question of what to do if the test detects a significant misspecification has not been addressed for two reasons. First it would appear to be impossible to distinguish between sources of bias without a priori speculation about particular sources. And, second, with well formulated hypotheses, estimation is in principal straightforward and standard tests for distinguishing between alternatives are available.

University of Iowa

Manuscript received November, 1979; final revision received September, 1980.

REFERENCES

[1] AMEMIYA, TAKESHI: "Regression Analysis when the Dependent Variable is Truncated Normal," *Econometrica*, 41(1973), 997–1016.

[2] COHEN, A. C.: "Estimating the Mean and Variance of Normal Populations from Singly Truncated and Doubly Truncated Samples," *Annals of Mathematical Statistics*, 21(1950), 557–569.

[3] GOLDBERGER, ARTHUR S.: "Abnormal Selection Bias," SSRI Workshop Paper No. 8006, University of Wisconsin, May, 1980.

[4] HALD, A.: "Maximum Likelihood Estimation of the Parameters of a Normal Distribution which is Truncated at a Known Point," *Skandinavisk Aktuarietidskrift*, 32(1949), 119–134.

[5] HAUSMAN, J. A.: "Specification Tests in Econometrics," *Econometrica*, 46(1978), 1251–1272.

[6] HAUSMAN, J. A., AND D. A. WISE: "A Conditional Probit Model for Qualitative Choice," *Econometrica*, 46(1978), 403–426.

[7] HURD, M.: "Estimation in Truncated Samples when there is Heteroscedasticity," *Journal of Econometrics*, 11(1979), 247–258.

[8] JOHNSON, N. L., AND S. KOTZ: *Distributions in Statistics, Continuous Univariate Distributions-1*. New York: Wiley, 1972.

[9] LEE, LUNG-FEI: "Strong Consistency of the Tobit Estimator in the Presence of Serially Correlated Disturbances," mimeograph, University of Minnesota, 1980.

[10] MADDALA, G. S.: "Specification Errors in Limited Dependent Variable Models," Discussion Paper, University of Florida, May, 1979.

[11] MADDALA, G. S., AND F. D. NELSON: "Specification Errors in Limited Dependent Variable Models," NBER Working Paper No. 96, July, 1975.

[12] NELSON, F. D.: "The Effect of and a Test for Misspecification in the Censored-Normal Model," Caltech Social Science Working Paper No. 291, September, 1979.

[13] PEARSON, K., AND A. LEE: "On the Generalized Probable Error in Multiple Normal Correlation," *Biometrika*, 6(1908), 59–68.

LINKED CITATIONS

[Footnotes]

[3] **A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences**
Jerry A. Hausman; David A. Wise
Econometrica, Vol. 46, No. 2. (Mar., 1978), pp. 403-426.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28197803%2946%3A2%3C403%3AACPMFQ%3E2.0.CO%3B2-8

[4] **Estimating the Mean and Variance of Normal Populations from Singly Truncated and Doubly Truncated Samples**
A. C. Cohen, Jr.
The Annals of Mathematical Statistics, Vol. 21, No. 4. (Dec., 1950), pp. 557-569.
Stable URL:
http://links.jstor.org/sici?sici=0003-4851%28195012%2921%3A4%3C557%3AETMAVO%3E2.0.CO%3B2-K

[4] **On the Generalised Probable Error in Multiple Normal Correlation**
Karl Pearson; Alice Lee
Biometrika, Vol. 6, No. 1. (Mar., 1908), pp. 59-68.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28190803%296%3A1%3C59%3AOTGPEI%3E2.0.CO%3B2-9

References

LINKED CITATIONS

[1] **Regression Analysis when the Dependent Variable Is Truncated Normal**
Takeshi Amemiya
Econometrica, Vol. 41, No. 6. (Nov., 1973), pp. 997-1016.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28197311%2941%3A6%3C997%3ARAWTDV%3E2.0.CO%3B2-F

[2] **Estimating the Mean and Variance of Normal Populations from Singly Truncated and Doubly Truncated Samples**
A. C. Cohen, Jr.
The Annals of Mathematical Statistics, Vol. 21, No. 4. (Dec., 1950), pp. 557-569.
Stable URL:
http://links.jstor.org/sici?sici=0003-4851%28195012%2921%3A4%3C557%3AETMAVO%3E2.0.CO%3B2-K

[5] **Specification Tests in Econometrics**
J. A. Hausman
Econometrica, Vol. 46, No. 6. (Nov., 1978), pp. 1251-1271.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28197811%2946%3A6%3C1251%3ASTIE%3E2.0.CO%3B2-X

[6] **A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences**
Jerry A. Hausman; David A. Wise
Econometrica, Vol. 46, No. 2. (Mar., 1978), pp. 403-426.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28197803%2946%3A2%3C403%3AACPMFQ%3E2.0.CO%3B2-8

[13] **On the Generalised Probable Error in Multiple Normal Correlation**
Karl Pearson; Alice Lee
Biometrika, Vol. 6, No. 1. (Mar., 1908), pp. 59-68.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28190803%296%3A1%3C59%3AOTGPEI%3E2.0.CO%3B2-9

NOTE: *The reference numbering from the original has been maintained in this citation list.*