DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

THE EFFECT OF AND A TEST FOR MISSPECIFICATION IN THE CENSORED-NORMAL MODEL

Forrest D. Nelson

SOCIAL SCIENCE WORKING PAPER 291

September 1979

THE EFFECT OF AND A TEST FOR MISSPECIFICATION IN THE CENSORED-NORMAL MODEL

Forrest D. Nelson

It is well-known that ordinary least-squares will produce inconsistent estimates of the regression parameters if the dependent variable is censored or truncated. Maximum likelihood estimation with a normality assumption on Tobit and other limited dependent variable models is being employed with increasing frequency to avoid this inconsistency. It is not so commonly acknowledged, however, that such estimates lack robustness.[1] 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 formulations. But to recognize the potential inconsistency in the face of misspecification without a test for and solution to such misspecification is of little use.

The purpose of this paper is to examine the nature of the inconsistency and to suggest a general test for misspecification. Section I considers the simple nonregression case of a censored variable. Likelihood equations for the location and scale parameters are obtained and simplified to show that they involve three sample statistics. In section II the general problem of inconsistency resulting from misspecification is then made clear, with the example of heteroscedasticity used to demonstrate the problem. A specification test following Hausman [1978] is then derived in section III for the general alternative hypothesis of no misspecification. These first three sections treat the nonregression case for ease of exposition, but the results are readily generalized to a regression model. Section IV contains a derivation of the specification test for the regression formulation. The results are summarized in section V.

I. THE MODEL AND MOM AND ML ESTIMATORS

We consider the case of a censored-normal variate y defined by the distribution function

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

where $\Phi(a)$ is the unit normal c.d.f.,

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

The first four moments of y can be written as:[1]

$$E_1 \equiv E(y;\mu,\sigma) = \mu\,\Phi\left(\frac{\mu}{\sigma}\right) + \sigma\,\phi\left(\frac{\mu}{\sigma}\right) \tag{1.1}$$

$$E_2 \equiv E(y^2;\mu,\sigma) = \mu^2\,\Phi\left(\frac{\mu}{\sigma}\right) + \sigma^2\,\Phi\left(\frac{\mu}{\sigma}\right) + \mu\sigma\,\phi\left(\frac{\mu}{\sigma}\right) \tag{1.2}$$

$$E(y^3;\mu,\sigma) = \mu E_2 + 2\sigma^2 E_1 \tag{1.3}$$

$$E(y^4;\mu,\sigma) = (\mu^2 + 3\sigma^2)E_2 + 2\mu\sigma^2 E_1 \tag{1.4}$$

The likelihood function for a random sample $(y_1,\ldots,y_N) = y$ is

$$L(\mu,\sigma;y) \propto \prod_{i\in\psi_1}\left[1-\Phi\left(\frac{\mu}{\sigma}\right)\right] \cdot \prod_{i\in\psi_2} \frac{1}{\sigma}\phi\left(\frac{y_i-\mu}{\sigma}\right) dy_i \tag{1.5}$$

where $\psi_1 = \{i|y_i = 0\}$ and $\psi_2 = \{i|y_i > 0\}$. Define the variable v_i as $v_i = 1$ if $y_i > 0$, $v_i = 0$ if $y_i = 0$. Then the log likelihood may be written as

$$\log L = C + \sum_{i=1}^{N}\{(1-v_i)\log[1-\Phi(\tfrac{\mu}{\sigma})] - v_i\log\sigma - \tfrac{1}{2}v_i(y_i-\mu)^2/\sigma^2\}. \tag{1.6}$$

Differentiation of (1.6) w.r.t. μ and σ yields

$$\frac{\partial \log L}{\partial\mu} = \sum_{i=1}^{N}\left\{\frac{-(1-v_i)}{\sigma}\,\frac{\phi\left(\frac{\mu}{\sigma}\right)}{1-\Phi\left(\frac{\mu}{\sigma}\right)} + \frac{v_i}{\sigma}\left(\frac{y_i-\mu}{\sigma}\right)\right\} \tag{1.7}$$

$$\frac{\partial \log L}{\partial\sigma} = \sum_{i=1}^{N}\left\{\frac{(1-v_i)}{\sigma}\,\frac{\mu}{\sigma}\,\frac{\phi\left(\frac{\mu}{\sigma}\right)}{1-\Phi\left(\frac{\mu}{\sigma}\right)} - \frac{v_i}{\sigma} + \frac{v_i}{\sigma}\left(\frac{y_i-\mu}{\sigma}\right)^2\right\} \tag{1.8}$$

The likelihood equations are obtained by setting (1.7) and (1.8) to zero and replacing μ and σ by $\hat{\mu}$ and $\hat{\sigma}$. Divide those likelihood equations by N and rearrange terms to obtain:

$$\hat{\mu}P + \hat{\sigma}\frac{\phi(\hat{\mu}/\hat{\sigma})}{1-\Phi(\hat{\mu}/\hat{\sigma})}(1-P) = M_1 \tag{1.9}$$

$$-\hat{\mu}^2 P + \hat{\sigma}^2 P - \hat{\sigma}\hat{\mu}\frac{\phi(\hat{\mu}/\hat{\sigma})}{1-\Phi(\hat{\mu}/\hat{\sigma})}(1-P) = M_2 - 2\mu M_1, \tag{1.10}$$

where $P = (\Sigma v_i)/N$ is the proportion of non-censored observations, $M_1 = (\Sigma y_i)/N = (\Sigma v_i y_i)/N$ is the sample mean, and $M_2 = (\Sigma y_i^2)/N = (\Sigma v_i y_i^2)/N$ is the second sample moment. Finally, define $\hat{\phi} = \phi(\hat{\mu}/\hat{\sigma})$ and $\hat{\Phi} = \Phi(\hat{\mu}/\hat{\sigma})$, and subtract twice μ times equation (1.9) from (1.10) to obtain the estimating equations:[2]

$$\hat{\mu}P + \hat{\sigma}\frac{\hat{\phi}}{1-\hat{\Phi}}(1-P) = M_1 \tag{1.11}$$

$$\hat{\mu}^2 P + \hat{\sigma}^2 P + \hat{\mu}\hat{\sigma}\frac{\hat{\phi}}{1-\hat{\Phi}}(1-P) = M_2 \qquad (1.12)$$

These equations are of course nonlinear in $\hat{\mu}$ and $\hat{\sigma}$ and require numerical procedures for solution.

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

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

where

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

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

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

where E_1 and E_2 are the first and second moments defined by equations (1.1) and (1.2) 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

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

where

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

$$e = \frac{1}{\sigma^2}\frac{\phi}{1-\Phi}[\sigma^2+\mu^2-E_2]$$

$$f = 2\Phi - \frac{1}{\sigma^2}\frac{\mu}{\sigma}\frac{\phi}{1-\Phi}[\sigma^2+\mu^2-E_2]$$

The inverse of 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 $I(\hat{\mu},\hat{\sigma})^{-1}$.

A proof of the consistency, asymptotic normality and asymptotic efficiency (i.e. that $AC(\hat{\mu},\hat{\sigma}) = I^{-1}$) is provided by Amemiya [1973] for the more general case of a regression model formulation. Inspection of the solution equations (1.11) and (1.12) reveals the nature of the consistency. The i.i.d. censored normal assumption implies that P, M_1 and M_2 will converge to $\Phi(\mu/\sigma)$, E_1 and E_2 respectively, so that in the limit solution of equations (1.11) and (1.12) requires $\hat{\mu}=\mu$ and $\hat{\sigma}=\sigma$.

An alternative estimator is provided by the method of moments. After replacing E(y) and $E(y^2)$ on the left side of equations (1.1) and (1.2) by the first and second sample moments, M_1 and M_2, respectively and substituting $\tilde{\mu}$ and $\tilde{\sigma}$ for μ and σ on the right, numerical procedures can be used to obtain the nonlinear solutions. Existence of the second and third moments guarantees strong convergence of the first two sample moments, so that the MOM estimators $\tilde{\mu}$ and $\tilde{\sigma}$ will be consistent. They lack asymptotic efficiency, however, and have no computational advantage over the preferred maximum likelihood estimates. Comparison of equations (1.11) and

(1.12) with (1.1) and (1.2) reveal the source of the efficiency gain of the MLE estimators $\hat{\mu}$ and $\hat{\sigma}$ over the MOM estimators $\tilde{\mu}$ and $\tilde{\sigma}$. The former employ one additional piece of information from the sample, namely the proportion of noncensored observations, P.

Note that the normality assumption imposes a nonlinear dependency among $\Phi(\mu/\sigma)$, E_1 and E_2; knowledge of any two allows solution for the third. This restriction suggests a modification to the MOM estimator which has distinct computational advantages. Solution of

$$P = \Phi(\tilde{\tilde{\gamma}}) \tag{1.15}$$

for $\tilde{\tilde{\gamma}}$ provides a consistent estimate of $\gamma = \mu/\sigma$. Substitution of $\tilde{\tilde{\gamma}}$ into either

$$M_1 = \tilde{\tilde{\sigma}}\tilde{\tilde{\gamma}}P + \tilde{\tilde{\sigma}}\phi(\tilde{\tilde{\gamma}}) \tag{1.16}$$

or

$$M_2 = \tilde{\tilde{\sigma}}^2\tilde{\tilde{\gamma}}^2 P + \tilde{\tilde{\sigma}}^2 P + \tilde{\tilde{\sigma}}^2\tilde{\tilde{\gamma}}\phi(\tilde{\tilde{\gamma}}), \tag{1.17}$$

as obtained from equations (1.1) and (1.2) respectively yields $\tilde{\tilde{\sigma}}$ and, in turn, $\tilde{\tilde{\mu}} = \tilde{\tilde{\gamma}} \cdot \tilde{\tilde{\sigma}}$ without the need for iterative solution procedures. Finally, the two alternative estimators obtainable in this fashion might both be computed and combined, say as a weighted average, to achieve some gain in asymptotic efficiency.

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 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.

II. ML ESTIMATES UNDER MISSPECIFICATION

The i.i.d. censored normal assumptions are sufficient for the asymptotic properties of the estimators discussed above. If any of those assumptions are violated, that is if the model is misspecified, the properties are no longer guaranteed. As an example, we examine in this section the consequences of a violation of the identically distributed assumption.

Suppose that the random sample is drawn from two distinct censored-normal populations with common location parameters μ but different scale parameters σ_1 and σ_2. We might ask whether ignorance of such sampling affects the consistency of the maximum likelihood estimates $\hat{\mu}$, $\hat{E}_1$, $\hat{E}_2$ and $\hat{\Phi}$. (Consistency of $\hat{\sigma}$ is of course a meaningless question in this case.) As will be shown below, the answer is yes.

Let r_1 and r_2 be the (fixed) proportions of observations in the sample from each of the two populations ($r_1 + r_2 = 1$). Estimates $\hat{\mu}$ or $\hat{\sigma}$ from the misspecified homoscedastic model will be solutions, in the limit, to the two equations

$$\hat{\mu}(r_1\Phi_1 + r_2\Phi_2) + \hat{\sigma}\frac{\hat{\phi}}{1-\hat{\Phi}}(1 - r_1\Phi_1 - r_2\Phi_2)$$

$$= \mu(r_1\Phi_1 + r_2\Phi_2) + r_1\sigma_1\phi_1 + r_2\sigma_2\phi_2 \qquad (2.1)$$

$$(\hat{\mu}^2 + \hat{\sigma}^2)(r_1\Phi_1 + r_2\Phi_2) + \hat{\mu}\hat{\sigma}\frac{\hat{\phi}}{1-\hat{\Phi}}(1 - r_1\Phi_1 - r_2\Phi_2)$$

$$= r_1[(\mu^2 + \sigma_1^2)\Phi_1 + \mu\sigma_1\phi_1] + r_2[(\mu^2 + \sigma_2^2)\Phi_2 + \mu\sigma_2\phi_2], \qquad (2.2)$$

where $\Phi_i = \Phi(\mu/\sigma_i)$ and $\phi_i = \phi(\mu/\sigma_i)$. These equations are obtained from equations (1.11) and (1.12) by taking the sample statistics to their probability limits. In general, $\hat{\mu} = \mu$ is not a solution so that $\hat{\mu}$ is not consistent.

No closed form expression for the bias is obtainable but it can be computed numerically for specified values of μ, σ_1, σ_2, r_1 and r_2. Table 1 below contains the results of such computations for a number of values of μ and varying degress of heteroscedasticity. For purposes of this illustration, the sampling ratio was fixed at $r_1 = r_2 = 1/2$ and σ_1 and σ_2 were chosen so that $r_1\sigma_1^2 + r_2\sigma_2^2 = 1$. Probability limits for $\hat{\mu}$ and $\hat{\sigma}$ were computed for values of μ ranging from -2 to +4 and for $\lambda = \sigma_1/\sigma_2$ ranging 1.5 to 4. These limits were then employed to compute $\hat{E}_1$, $\hat{E}_2$, and $\hat{\Phi}$ as defined in the previous section. The table contains the asymptotic bias for the maximum likelihood estimators of each of the four parameters.

[Table 1 about here]

The pattern of the bias is not particularly easy to summarize. $|\hat{\mu} - \mu|$ gets small as μ grows large since the degree of censoring diminishes. It seems to increase monotonically with λ, the degree of heteroscedasticity, be negative for large negative μ and positive for large positive μ. The most serious bias is at negative μ and there are values for μ which yield zero bias.

The asymptotic bias in $\hat{\mu}$ is translated into biased estimates of Φ, E_1 and E_2, but the pattern is quite different. The error in these statistics appears quite small, relative to $(\hat{\mu} - \mu)$,

<u>Table 1</u>

ASYMPTOTIC BIAS IN ML ESTIMATORS UNDER HETEROSCEDASTIC MISSPECIFICATION*

σ_1/σ_2	μ = -2.0	-1.0	-.5	0.0	.5	1.0	1.5	2.0	4.0
					Plim $(\hat{\mu}-\mu)$				
4.0	-.746	-.911	-.657	-.084	.056	.040	.020	.009	.0002
3.5	-.737	-.867	-.504	-.074	.047	.037	.019	.009	.0002
3.0	-.723	-.778	-.415	-.062	.037	.034	.018	.008	.0001
2.5	-.698	-.620	-.288	-.048	.026	.027	.015	.007	.0001
2.0	-.630	-.392	-.164	-.031	.014	.018	.011	.005	.0001
1.5	-.382	-.143	-.056	-.012	.005	.007	.005	.002	.0000
					Plim $(\hat{\Phi}-\Phi)$				
4.0	.0000	.0002	-.0032	-.0324	-.0511	-.0102	.0139	.0167	.0009
3,5	.0000	.0002	-.0031	-.0288	-.0448	-.0115	.0121	.0156	.0008
3.0	.0000	.0000	-.0041	-.0244	-.0368	-.0122	.0098	.0140	.0007
2.5	.0000	-.0003	-.0046	-.0188	-.0273	-.0117	.0066	.0115	.0005
2.0	.0000	-.0007	-.0042	-.0121	-.0176	-.0088	.0030	.0078	.0004
1.5	-.0000	-.0006	-.0021	-.0046	-.0061	-.0037	.0006	.0030	.0001
					Plim $(\hat{E}_1 - E_1)$				
4.0	-.0001	-.0006	.0036	.0276	.0277	.0046	-.0056	-.0060	-.0002
3.5	-.0001	-.0004	.0049	.0244	.0248	.0052	-.0049	-.0056	-.0002
3.0	-.0001	.0001	.0059	.0204	.0210	.0057	-.0040	-.0050	-.0002
2.5	-.0001	.0007	.0061	.0156	.0161	.0056	-.0027	-.0042	-.0001
2.0	-.0000	.0012	.0050	.0099	.0101	.0043	-.0013	-.0029	-.0001
1.5	.0000	.0009	.0033	.0037	.0038	.0019	-.0002	-.0011	-.0000
					Plim $(\hat{E}_2 - E_2)$				
4.0	.0002	.0017	-.0087	-.0366	-.0251	-.0037	.0040	.0039	.0001
3.5	.0002	.0010	-.0210	-.0321	-.0230	-.0042	.0036	.0036	.0001
3.0	.0002	-.0002	-.0122	-.0268	-.0200	-.0047	.0029	.0032	.0001
2.5	.0002	-.0018	-.0115	-.0203	-.0158	-.0046	.0020	.0028	.0000
2.0	.0001	-.0028	-.0085	-.0126	-.0103	-.0038	.0010	.0019	.0000
1.5	-.0001	-.0017	-.0036	-.0047	-.0040	-.0017	.0002	.0008	.0000

*The numerical calculation of the asymptotic bias assumed equal sampling from two populations with the same location parameter μ, distinct scale parameters σ_1 and σ_2, and censoring at $y=0$. σ_1 and σ_2 were chosen so that $\sigma_1/\sigma_2 = \lambda$ and $(\sigma_1^2 + \sigma_2^2)/2 = 1$.

as one might expect -- ML produces estimates $\hat{\mu}$ and $\hat{\sigma}$ which in some sense best fit $\{\hat{\Phi}, \hat{E}_1$ and $\hat{E}_2\}$ to $\{P, M_1, M_2\}$ while constraining the former to satisfy the implicit dependency. Curiously enough, the maximum error in, for example, $(\hat{\Phi}-\Phi)$ appears to occur very near the point (value of μ) where $\hat{\mu} = \mu$ at each value of λ.

As noted above, heteroscedasticity is only one example of a misspecification which will lead to inconsistent maximum likelihood estimates -- it is used here only to illustrate the problem. Departure from normality, contamination, sampling from heterogeneous populations and perhaps even some nonrandom sampling,[1] may all lead to similar failures. This is, by and large, in contrast with estimation in noncensored normal samples. In that case, ML estimates of the location and scale parameters are independent and any misspecification which, for example, leaves the expectation of the sample, or more correctly the probability limit of the sample mean, unaffected will not cause inconsistency of the ML estimate of the location parameter or, identically the first population moment. Such robustness does not hold for the censored normal case of concern here. Any misspecification which effects the probability limit of any of these sample statistics, P, M_1 or M_2, will generate inconsistent estimates. If those probability limits coincide with some member of the family of censored normals, then $\hat{\Phi}$, $\hat{E}_1$ and $\hat{E}_2$ will be consistent even though $\hat{\mu}$ and/or $\hat{\sigma}$ are not. If no censored normal, i.e. no pair of values, $\sigma > 0$ and μ, would generate those probability limits as values of $\Phi(\mu/\sigma)$, $E(y;\mu,\sigma)$ and $E(y^2;\mu,\sigma)$ respectively, then, in general, none of the five MLEs considered above will be consistent. The latter was exemplified by the heteroscedastic example.

Finally it should be noted that the MOM estimators may in some limited sense be more robust. That is, so long as P, M_1, and M_2 converge, those statistics are consistent for the population parameters to which they converge. MOM estimators $\hat{\mu}$ and $\hat{\sigma}$ will not generally be consistent for anything of interest, on the other hand, since the functional relationships employed for their derivation will not, in general, be correct under misspecification.

III. AN ASYMPTOTIC TEST AGAINST MISSPECIFICATION

The sensitivity of MLEs 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 [1978]. 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 $V_0 = I^{-1}$ and $V_1 - V_0$ is nonnegative definite. Then, as Hausman shows, $q = \hat{\theta}_1 - \hat{\theta}_0$ 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 m is not asymptotically χ^2, so that m serves as a test statistic. Hausman proceeds to outline conditons under which m will follow a noncentral χ^2 so that the power of the test may be examined.

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 sections I and II 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 I is subject to the same sensitivity to misspecification as is the MLE.[1] 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

$$q_1 \equiv P - \Phi(\hat{\mu}/\hat{\sigma}) \tag{3.1}$$

$$q_2 \equiv M_1 - E(y; \hat{\mu}, \hat{\sigma}) = \frac{\hat{\mu} - \hat{E}_1}{1 - \hat{\Phi}} (P - \hat{\Phi}) \tag{3.2}$$

$$q_3 \equiv M_2 - E(y^2; \hat{\mu}, \hat{\sigma}) = \frac{\hat{\mu}^2 + \hat{\sigma}^2 - \hat{E}_2}{1 - \hat{\Phi}} (P - \hat{\Phi}) \tag{3.3}$$

where the equalities in 3.2 and 3.3 are obtained after substitution from equations (1.1) and (1.11), and (1.2) and (1.12) 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

$$q_2 = \frac{\mu - E_1}{1 - \Phi}(P - \hat{\Phi}) + o(N^{-1/2}). \tag{3.4}$$

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

$$q_3 = \frac{\mu^2 + \sigma^2 - E_2}{1-\Phi}(P - \hat{\Phi}) + o(N^{-1/2}). \tag{3.5}$$

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 $\sqrt{N}(P - \Phi) \sim$

$AN(0, \Phi(1-\Phi))$, so that the asymptotic variance of P is $V(P) = \Phi(1-\Phi)$.

The asmyptotic distribution of $\hat{\Phi}$ and, for completeness, $\hat{E}_1$ and $\hat{E}_2$ is 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

$$\hat{\Phi} - \Phi \simeq (\hat{\mu} - \mu)\phi\frac{1}{\sigma} - (\hat{\sigma} - \sigma)\phi\frac{\mu}{\sigma^2} \tag{3.6}$$

$$\hat{E}_1 - E_1 \simeq (\hat{\mu} - \mu)\Phi + (\hat{\sigma} - \sigma)\phi \tag{3.7}$$

$$\hat{E}_2 - E_2 \simeq (\hat{\mu} - \mu)\cdot 2 \cdot (\mu\Phi + \sigma\phi) + (\hat{\sigma} - \sigma)\cdot 2 \cdot \sigma\Phi \tag{3.8}$$

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,

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

where I is the information matrix defined in (1.14) and A is given by

$$A = \begin{bmatrix} \frac{1}{\sigma}\phi & \Phi & 2\cdot E_1 \\ -\frac{1}{\sigma}\frac{\mu}{\sigma}\phi & \phi & 2\cdot\sigma\Phi \end{bmatrix}. \tag{3.9}$$

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

$$V(\hat{\Phi}) = (\frac{1}{\sigma}\phi)^2 V(\hat{\mu}) + (\frac{1}{\sigma}\frac{\mu}{\sigma}\phi)^2 V(\hat{\sigma}) - 2\frac{1}{\sigma^2}\frac{\mu}{\sigma}\phi^2 \operatorname{Cov}(\hat{\mu}, \hat{\sigma}). \tag{3.10}$$

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:[2]

$$\hat{V}(P-\hat{\Phi}) = \hat{\Phi}\cdot(1-\hat{\Phi}) - \left[\frac{1}{\hat{\sigma}}\hat{\phi} \quad -\frac{1}{\hat{\sigma}}\frac{\hat{\mu}}{\hat{\sigma}}\hat{\phi}\right][I(\hat{\mu},\hat{\sigma})]^{-1}\begin{bmatrix}\frac{1}{\hat{\sigma}}\hat{\phi} \\ -\frac{1}{\hat{\sigma}}\frac{\hat{\mu}}{\hat{\sigma}}\hat{\phi}\end{bmatrix} \tag{3.11}$$

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 propulation with location and scale parameters μ and σ, the statistic

$$m = N\cdot(P-\hat{\Phi})^2 \ / \ \hat{V}(P-\hat{\Phi}) \tag{3.12}$$

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 as examined in section II. In each experiment, two samples of size N/2 were drawn randomly from a $N(\mu, \sigma_i^2)$ 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 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 2 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, for comparison with table 1, computed values for $\overline{\hat{\Phi}-P}$ and $\overline{\hat{\mu}}-\mu$. A column containing relevant parameters for the $\chi^2_{(1)}$ distribution is included as a benchmark.

[Table 2 about here]

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 2

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

IV. THE EXTENSION TO A REGRESSION MODEL

Section III 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

$$F(y_i) = \Phi\left(\frac{y_i - \beta' X_i}{\sigma}\right) \qquad \text{for } y_i \geq 0$$
$$= 0 \qquad \text{for } y_1 < 0 \qquad (4.1)$$

This is of course the tobit model more commonly described by

$$y_i = \beta' X_i + u_i \qquad \text{if RHS} > 0$$
$$= 0 \qquad \text{otherwise}$$
$$u_i \sim IN(0, \sigma^2) .$$

The likelihood for a random sample of size N is given by equation 1.5 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; $\underline{\phi}$ be the $N \times 1$ vector with $\phi(\beta' X_i/\sigma)$ at element i; and $\underline{\overline{\Phi}}$ be the $N \times N$ diagonal matrix with $\Phi(\beta' X_i/\sigma)$ at position ii. When $\underline{\phi}$ and $\underline{\overline{\Phi}}$ are evaluated at the MLEs $\hat{\beta}$ and $\hat{\sigma}$, they will be indicated as $\hat{\underline{\phi}}$ and $\hat{\underline{\overline{\Phi}}}$ 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

$$X'WX\hat{\beta} + \hat{\sigma}X'[I-W][I-\hat{\underline{\Phi}}]^{-1}\hat{\underline{\phi}} = X'Y \tag{4.2}$$

and

$$\hat{\beta}'X'WX\hat{\beta} + \hat{\sigma}^2\mathrm{Tr}[W] + \hat{\sigma}\hat{\beta}'X'[I-W][I-\hat{\underline{\Phi}}]^{-1}\hat{\underline{\phi}} = Y'Y \tag{4.3}$$

So long as the y_is are random with distribution as specified in 4.1 and the sequence X_i is such that

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

solution of 4.2 and 4.3 will yield estimates which are consistent, asymptotically normal, and asymptotically efficient. That is,

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

with I defined as

$$I(\beta,\sigma) = \frac{1}{\sigma^2}\frac{1}{N}\begin{bmatrix} X'[C+\underline{\Phi}]X & X'[\underline{\phi} - CB] \\ [\underline{\phi} - CB]'X & B'CB - B'\underline{\phi} + 2\mathrm{Tr}[\underline{\Phi}] \end{bmatrix} \tag{4.4}$$

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

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

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 propose is again the Hausman test, based this time on estimates of $E(\frac{1}{N}X'Y)$. Under fairly general conditions on X_i and the distribution of y_i, $\frac{1}{N}X'Y$ 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

$$E(y_i;\beta,\sigma) = \beta'X_i\Phi\left(\frac{\beta'X_i}{\sigma}\right) + \sigma\phi\left(\frac{\beta'X_i}{\sigma}\right) \tag{4.5}$$

and

$$E(y_i^2;\beta,\sigma) = (\beta'X_i)^2\Phi\left(\frac{\beta'X_i}{\sigma}\right) + \sigma^2\Phi\left(\frac{\beta'X_i}{\sigma}\right) + \beta'X_i\sigma\phi\left(\frac{\beta'X_i}{\sigma}\right) \tag{4.6}$$

Thus

$$E_{xy} \equiv E(\tfrac{1}{N}X'Y;\beta,\sigma) = \tfrac{1}{N}[X'\underline{\Phi}X\beta + \sigma X'\underline{\phi}] \tag{4.7}$$

and the variance of $\frac{1}{N}X'Y$ is

$$V_1 \equiv V(\tfrac{1}{N}X'Y;\beta\sigma) = \tfrac{1}{N}X'V_YX \tag{4.8}$$

where V_y is an $N\times N$ diagonal matrix with diagonal elements $E(y_i^2) - E(y_i)^2$ as defined in (4.5) and (4.6). Thus,

$$\sqrt{N}(\tfrac{1}{N}X'Y - E_{XY}) \sim AN(0, \lim_{n\to\infty} V_1).$$

$\frac{1}{N}X'Y$ is the consistent but inefficient estimator we require for the test statistic and its variance is given by expression (4.8).

The corresponding efficient estimator is the maximum likelihood estimator for E_{XY}. Define the statistic $\hat{E}_{XY}$ as expression 4.7 evaluated at the MLEs $\hat{\beta}$ and $\hat{\sigma}$. Its variance is obtained by expanding it about β and σ,

$$\hat{E}_{XY} - E_{XY} = \tfrac{1}{N}[X'\underline{\Phi}X(\hat{\beta}-\beta) + X'\underline{\phi}(\hat{\sigma}-\sigma)] + o(N^{-1/2}) \tag{4.9}$$

The left side of (4.9) will thus have the same asymptotic distribution as the indicated linear combination of $(\hat{\beta} - \beta)$ and $(\hat{\sigma} - \sigma)$. That is,

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

where V_0 is defined by

$$V_0 = \frac{1}{N^2}[X'\underline{\bar{\Phi}}X \quad X'\underline{\phi}]\,I(\beta,\sigma)^{-1}\begin{bmatrix} X'\underline{\bar{\Phi}}X \\ \underline{\phi}'X \end{bmatrix} \tag{4.10}$$

Combining these results, we obtain the desired test statistic,

$$m = N(\tfrac{1}{N}X'Y - \hat{E}_{XY})'(\hat{V}_1 - \hat{V}_0)^{-1}(\tfrac{1}{N}X'Y - \hat{E}_{XY}) \tag{4.11}$$

where $\hat{V}_1$ and $\hat{V}_0$ are obtained by evaluation of (4.8) and (4.10) respectively at the MLEs $\hat{\beta}$ and $\hat{\sigma}$.[2] Under the maintained assumptions, this statistic will follow, asymptotically, a $\chi^2_{(k)}$ distribution.

V. 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 MLEs for this model lack robustness against misspecification. This was illustrated in section II for the nonregression case with numerical results on the asymptotic bias arising from heteroscedasticity. Similar results will hold for other violations of the assumptions and extend to the regression case as well.

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 natural estimators to employ for this test would be those for the location and scale parameters. But, for the types of misspecification of concern here, those parameters are not necessarily the same under the maintained and alternative models. Thus we suggest using estimators for population moments. We further demonstrate a singularity in the asymptotic covariance matrix when the test is applied to a pair of estimators whose dimension equals the total number of unknown parameters. The test must therefore be based on some reduced set of estimators.

The suggested test statistics are given by expressions (3.12) and (4.11) 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 nonregression case was examined by Monte-Carlo methods at the end of section III. 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.

FOOTNOTES

Introduction

1. Hausman and Wise [1978] have noted inconsistencies arising from misspecification in probit-logit models. The effect of heteroscedasticity has been examined by Maddala and Nelson [1975] and by Maddala [1979] in the case of the tobit model and by Hurd [1977] in a truncated variable model.

Section I

1. Amemiya [19] presents the moments from a truncated normal from which these are readily derived.
2. Cohen [1950] presented similar equations for a variety of censoring and truncation schemes. He proposed estimation of $\xi = (\mu - \tau)/\sigma$ where τ is the (known) censoring threshold.

Section II

1. The independence assumption is perhaps the least crucial. Under first-order serial correlation, for example, the three relevant sample statistics will converge to the corresponding population parameters, guaranteeing consistency.

Section III

1. Hauseman's condition that $\hat{\theta}_1$ be consistent under H_a may be stronger than necessary -- his test might serve well,

so long as $\text{Plim}\,\hat{\theta}_1 \neq \text{Plim}\,\hat{\theta}_0$ 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.

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

SECTION IV

1. As before, statistics for $E(Y'Y)$ and $Tr(\underline{\Phi})$ 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.

2. Again there exist other consistent estimators for V_1 and V_0, use of $-H^{-1}$ in (4.9) for example, but they will not guarantee a positive definite variance estimate for the difference.

REFERENCES

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

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.

HAUSMAN, J.A., "Specification Tests in Econometrics," Econometrica 46 (1978): 1251-1272.

HAUSMAN, J.A. and D. A. Wise, "A Conditional Probit Model for Qualitative Choice," Econometrica 46 (1978): 403-426.

HURD, M., "Estimation in Truncated Samples when there is Heteroscedasticity," Economics Series Working Paper No. 65, Stanford University, October 1977.

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

MADDALA, G.S. and F. D. Nelson, "Specification Errors in Limited Dependent Variable Models," NBER Working Paper No. 96, July 1975.