The Stata Journal (2002)
2, Number 2, pp. 125–139

Bootstrapping a conditional moments test for normality after tobit estimation

David M. Drukker
Stata Corporation
ddrukker@stata.com

Abstract. Categorical and limited dependent variable models are routinely estimated via maximum likelihood. It is well-known that the ML estimates of the parameters are inconsistent if the distribution or the skedastic component is misspecified. When conditional moment tests were first developed by Newey (1985) and Tauchen (1985), they appeared to offer a wide range of easy-to-compute specification tests for categorical and limited dependent variable models estimated by maximum likelihood. However, subsequent studies found that using the asymptotic critical values produced severe size distortions. This paper presents simulation evidence that the standard conditional moment test for normality after tobit estimation has essentially no size distortion and reasonable power when the critical values are obtained via a parametric bootstrap.

Keywords: st0011, conditional moment tests, bootstrap, tobit, normality

1 Introduction

Building on the work of White (1982), Newey (1985), and Tauchen (1985) independently developed conditional moment tests. At first, these conditional moment tests appeared to offer a gold mine of easy-to-compute specification tests for models estimated by maximum likelihood. These tests promised to be especially valuable in checking for evidence against homoskedasticity and assumed distributions after estimating categorical and limited dependent variable models. A high water mark in this tide of tests was reached by Pagan and Vella (1989), who derived several conditional moment specification tests for probit, logit and tobit models and linked them to previously derived specification tests. However, subsequent simulation studies (Skeels and Vella 1997, 1999; Ericson and Hansen 1999, *inter alia*) found that obtaining critical values from the asymptotic distributions of these tests causes large size distortions.

This paper evaluates the performance of one of these tests, when the critical values are obtained using a parametric bootstrap. This paper presents evidence that obtaining the critical values from a parametric bootstrap with 500 repetitions essentially removes the size distortions and still yields reasonable power.

2 Conditional moment test for normality after tobit estimation

2.1 Intuition and conditional moment restrictions

Building on the work of White (1982), the theory of conditional moment specification tests after maximum likelihood estimation was formulated by Newey (1985) and Tauchen (1985).[1] Pagan and Vella (1989) derived a number of conditional moment tests to test the specification of binary and censored regression models. Skeels and Vella (1997, 1999) provided simulation evidence that the test against the null of normality derived by Pagan and Vella and an alternative version of the test that they proposed were all oversized.

Following the notation of Skeels and Vella (1999), suppose that one has N observations on (y_i, x_i) and that the maintained hypothesis is that these data follow a simple normal tobit model. Specifically,

$$y_i^* = x_i'\beta + u_i, \qquad i = 1, \dots, N \tag{1}$$

where y_i^* is the unobserved latent variable, x_i is a $(K \times 1)$ vector of covariates, β is a $(K \times 1)$ vector of parameters, and u_i is the disturbance term which is $N(0, \sigma^2)$ under the null hypothesis. Since this is a tobit model, only

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

is observed over the N observations.

The importance of testing that the u_i are normally distributed comes from the well-known fact that the standard tobit estimator is not consistent if the u_i are not normally distributed[2].

The intuition behind the conditional moments test after maximum likelihood is straightforward. Since the model was estimated by maximum likelihood, the assumed data-generating process specifies all moments of disturbances conditional on the covariates. These conditional moments can be used to write down conditional moment restrictions that have conditional expected values of zero.[3] This produces a vector of conditional moment restrictions that are zero under the null. Under the null, you would expect the sample averages to be close to zero. The trick is to find the appropriate weighting matrix that accounts for any covariance in the moment restrictions and scales the sample averages so that the resulting statistic converges to a known distribution. Newey (1985) and Tauchen (1985) found such a matrix, here denoted by $\widehat{Q}^{-1}$, so that

[1] See Wooldridge (2001), Section 13.7 for a textbook introduction to these tests.

[2] See Wooldridge (2001), Chapter 16.6.3 for a discussion of this issue.

[3] For instance, in the classical normal linear model, the fourth moment of the residuals is $E[u_i^4|x_i] = 3\sigma^4$. This conditional moment can be used to form the restriction that $E[u_i^4 - 3\sigma^4|x_i] = 0$.

$$\tau = \iota' \widehat{M} \widehat{Q}^{-1} \widehat{M}' \iota \rightarrow_d \chi^2(r) \tag{2}$$

where ι is an $(N \times 1)$ vector of ones, $\widehat{M}$ is the $(N \times r)$ matrix of sample realizations of the r moment restrictions, and $\widehat{Q}^{-1}$ is a feasible weighting matrix that properly scales the inner product of sample averages $\iota'\widehat{M}$. Before dealing with the details of the scaling matrix $\widehat{Q}^{-1}$, consider the moment conditions themselves.

Essential to these tests is a vector of functions whose expected value is zero. These are the conditional moment restrictions. Since the moment-based methods that test for normality use the third and fourth moment of the normal distribution, the vector of population moment conditions will have dimension (2×1); i.e., $r = 2$ in this case. To build intuition, temporarily ignore the fact that the u_i are disturbances in a tobit model. If the u_i were disturbances from a simple linear model, then $u_i = y_i - x_i'\beta$. If the u_i are normally distributed, then

$$E[u_i^3|x_i] = 0$$

and

$$E[u_i^4 - 3 * \sigma^4|x_i] = 0$$

In this case, the ith realization of the restriction on the third moment of u_i is

$$m_{i,1}(\theta) = u_i^3$$

where $\theta' = (\beta', \sigma^2)$. Similarly, the ith realization of the condition on the fourth moment of u_i is

$$m_{i,1}(\theta) = u_i^4 - 3\sigma^2$$

As a vector,

$$m_i(\theta) = \begin{pmatrix} u_i^3 \\ u_i^4 - 3\sigma^2 \end{pmatrix}$$

Note that $E[m_i(\theta)|x_i] = (0,0)'$. Since $m_i(\theta)$ depends on the true θ, it is not feasible. If we were interested in testing for the normality of the disturbances after estimating a classical normal linear model by maximum likelihood, we would use $\widehat{\theta}' = \widehat{\beta}', \widehat{\sigma}^2$ in place of θ. The scaling matrix $\widehat{Q}^{-1}$ is explicitly constructed to account for the fact that θ was estimated via maximum likelihood.

Now, returning to the problem at hand, the u_i are not simple linear disturbances, but rather they are tobit disturbances. Lee and Maddala (1985) showed that the third moment for tobit residuals is

$$m_{i,1}(\theta) = I_i u_i^3 - (1 - I_i)(z_i^2 + 2)\sigma^3 \lambda_i \tag{3}$$

where

$$I_i = \begin{cases} 1 & \text{if } y_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

$z_i = x_i'\beta/\sigma$, $\lambda_i = \phi(z_i)/(1-\Phi(z_i))$, $\phi()$ is the standard normal density, and $\Phi()$ is the standard normal CDF. The intuition for equation (3) is that when $I_i = 1$, no censoring occurs, and the third-moment restriction is the same as for the classical normal linear model; i.e., $E[u_i|I_i = 1] = u_i^3$. When $I_i = 0$ the expected value of u_i conditional on $I_i = 0$ must account for the censoring; i.e., the $E[u_i|I_i = 0] = (z_i^2 + 2)\sigma^3\lambda_i$, as shown by Lee and Maddala (1985).

The second-moment restriction, which is derived from the fourth moment of a normal random variable, has an analogous intuition. The moment restriction is

$$m_{i,2}(\theta) = I_i(u_i^4 - 3\sigma^4) + (1 - I_i)(z_i^2 + 3)\sigma^4\lambda_i z_i \tag{4}$$

As above, when $I_i = 1$ (i.e., there is no censoring), the u_i have the same restriction as in the classical normal linear model. When $I_i = 0$, the fourth moment of the u_i is adjusted for the censoring.

As described in Section 6, the $(N \times 2)$ matrix $\widehat{M}$ in the test statistic in equation (2) is obtained by plugging the tobit estimates $\widehat{\beta}$ and $\widehat{\sigma}$ in for the true values.

2.2 The scaling matrix

Let $f(y_i;\theta)$ be the contribution of observation i to the tobit log-likelihood. Letting θ be $(p \times 1)$ implies that the score matrix, $S(\theta)$, will be $(N \times p)$ and that the average information matrix, $\Im(\theta)$, will be $(p \times p)$; i.e.,

$$S_{ik}(\theta) = \frac{\partial f(y_i;\theta)}{\partial \theta_k}, \qquad i = 1,\ldots,N \qquad k = 1,\ldots,p$$

and

$$\Im_{k,l}(\theta) = \frac{1}{N}\sum_{i=1}^{N} E\left\{\frac{\partial^2 f(y_i;\theta)}{\partial\theta_k\partial\theta_l}\right\}, \qquad k,l = 1,\ldots,p$$

In subsection 2.1, $m_i(\theta)$ was defined to be $m_i(\theta) = (m_{i,1}(\theta), m_{i,2}(\theta))'$, and the $m_{i,r}(\theta)$, $r = 1,2$ were explicitly given. The scaling matrix requires a $(p \times r)$ matrix $W(\theta)$, where $W(\theta)$ is an average of derivatives of $m_i(\theta)$ with respect to the p parameters in θ. Specifically, let

$$W_{k,l}(\theta) = \frac{1}{N}\sum_{i=1}^{N} E\left\{\frac{\partial m_{i,l}(\theta)}{\partial\theta_k}\right\}, \qquad k = 1,\ldots,p, \qquad l = 1,\ldots,r$$

Finally, let $\widehat{\Im} = \Im(\widehat{\theta})$, $\widehat{S} = S(\widehat{\theta})$, $\widehat{W} = W(\widehat{\theta})$, and require that

$$\widehat{\Im} \rightarrow_p^{H_0} \lim_{N\to\infty} \Im(\theta)$$

and

$$\widehat{W} \rightarrow_p^{H_0} \lim_{N\to\infty} W(\theta)$$

Then

$$\widehat{Q} = \left(\widehat{M} - \widehat{S}\widehat{\Im}^{-1}\widehat{W}\right)' \left(\widehat{M} - \widehat{S}\widehat{\Im}^{-1}\widehat{W}\right)'$$

Note that while $\widehat{S}$ is pinned down, there is some freedom in choosing $\widehat{\Im}$ and $\widehat{W}$. Skeels and Vella (1999) discuss the restrictions required in choosing $\widehat{\Im}$ and $\widehat{W}$ and two resulting estimators of $\widehat{Q}$. In this paper, only one of the two is implemented.[4] Specifically, choosing the consistent estimators

$$\widehat{W} = N^{-1}\widehat{S}'\widehat{M}$$

and

$$\widehat{\Im} = N^{-1}\widehat{S}'\widehat{S}$$

yields an estimator $\widehat{Q}$ that is consistent for Q.

This estimator of Q was chosen for two reasons. First, Skeels and Vella (1999) show that it always has the larger size of the two tests and that it has higher power in their tobit simulations. Second, the primary goal of this paper is to investigate the extent that a parametric bootstrap can resolve the small sample problems with the τ that uses the conventional estimator of Q. In contrast, the second estimator of Q considered by Skeels and Vella (1999) can be seen as a new estimator of Q that was postulated to have better finite sample properties.

Finally, it should be mentioned that τ can be calculated using an artificial regression. This point is discussed in Section 6.

2.3 Parametric bootstrap critical values

Applying the results of Newey (1985) and Tauchen (1985), Pagan and Vella (1989) show that τ has an asymptotic chi-squared distribution with 2 degrees of freedom. However, as noted above, several studies have found that the sample sizes frequently encountered in practice are not large enough for the distribution of τ to be reasonably approximated by its asymptotic distribution. This paper investigates whether obtaining critical values from a parametric bootstrap with 500 replications yields reasonable size and power properties.

Following Efron and Tibshirani (1993), the parametric bootstrap is quite straightforward, but some additional notation will make it even clearer. First, write the statistic τ as $\tau(\theta)$. This makes the dependence of τ on θ explicit and is useful in describing the parametric bootstrap algorithm.

For the test of interest, we can reject H_0 at the level α if

$$\tau(\theta) > F_{1-\alpha}$$

where F is the distribution of $\tau(\theta)$ and $F_{1-\alpha}$ is the $(1-\alpha)$ centile of the distribution F.

[4] Skeels and Vella (1999) call it the OPG estimator. OPG stands for "outer product of the gradient".

The parametric bootstrap uses the following algorithm to estimate F:

1. Estimate the parameters θ on the sample $(y_i, x_i), i = 1, \dots, N$. Call these estimates $\widehat{\theta}_0$.
2. For R replications, repeat steps (a)–(f).
   (a) Generate N observations of a pseudonormal variate $\tilde{u}_i$ with zero mean and standard deviation given in $\widehat{\theta}_0$.
   (b) Use $x_i, \widehat{\theta}_0$, and $\tilde{u}_i$ to generate $\tilde{y}_i^*$.
   (c) Use $\tilde{y}_i^*$ to generate
   $$\tilde{y}_i = \begin{cases} \tilde{y}_i^* & \text{if } \tilde{y}_i^* > 0 \\ 0 & \text{Otherwise} \end{cases}$$
   (d) Estimate a tobit model on $(\tilde{y}_i, x_i)$ to obtain estimates $\widehat{\theta}_r$.
   (e) Using $\widehat{\theta}_r$, calculate $\tau_r(\widehat{\theta}_r)$.
   (f) Save $\tau_r(\widehat{\theta}_r)$ in a file.
3. The collection of $\tau_r(\widehat{\theta}_r)$, $r = 1, \dots, R$ can be used to estimate the empirical distribution of F. In particular, the $(1 - \alpha)$ centile of F is estimated by the $(1 - \alpha)$ centile of $\tau_r(\widehat{\theta}_r)$, $r = 1, \dots, R$. This estimated centile, $\widehat{F}_{1-\alpha}$, is the parametric bootstrap critical value.

Hence, we can reject H_0 if $\tau(\widehat{\theta}_0) > \widehat{F}_{1-\alpha}$.

3 Design of the simulations

Simulations were run to estimate the size and power of this conditional moment test when the critical values were obtained via a parametric bootstrap. Two basic designs were used in the simulations. Following Skeels and Vella (1999), one design was formulated around the data used by Moffit (1984). A second set of simulations was performed around a simple specification at three different sample sizes. As noted by Skeels and Vella (1999), running the simulations based on the Moffit (1984) data provides some idea of how the test may perform with the type of data actually encountered. The second set of simulations reveals what happens to the performance of the test when the sample size is changed, holding the specification constant.

In the first set of simulations, data were generated according to

$$\begin{aligned} H_i^* &= \beta_{\text{nwi}} nwi_i + \beta_{\text{ms}} ms_i + \beta_{\text{age}} age_i + \beta_{\text{race}} race_i \\ &\quad + \beta_{\text{clt6}} clt6 + \beta_{\text{educ}} educ + constant + u_i \end{aligned} \tag{5}$$

where H_i^* is the latent number of hours per week, *nwi* is the annual level of nonwage income, *ms* is an indicator of marital status, *age* gives the age, *race* is a binary variable (white=1, 0 otherwise), *clt*6 is the number of children under 6 in the household, and *educ* gives the number of years of education.[5] The parameter estimates used are reported in the output below.

```
. use moffit, clear

. describe

Contains data from moffit.dta
  obs:           610
 vars:            21                          26 Feb 2002 14:19
 size:        53,680 (99.0% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
hours           float  %9.0g
x2              float  %9.0g
nwi             float  %9.0g
const           float  %9.0g
ms              float  %9.0g
age             float  %9.0g
race            float  %9.0g
x8              float  %9.0g
clt6            float  %9.0g
cgt6            float  %9.0g
x11             float  %9.0g
educ            float  %9.0g
x13             float  %9.0g
x14             float  %9.0g
x15             float  %9.0g
x16             float  %9.0g
x17             float  %9.0g
x18             float  %9.0g
x19             float  %9.0g
x20             float  %9.0g
x21             float  %9.0g
-------------------------------------------------------------------------------
Sorted by:
```

(*Continued on next page*)

[5] See Moffit (1984) for more on the data and the model involved. I would like to thank Chris Skeels for making the data available to me.

```
. tobit hours nwi ms age race clt6 educ, ll(0)
Tobit estimates                                   Number of obs   =        610
                                                  LR chi2(6)      =      24.18
                                                  Prob > chi2     =     0.0005
Log likelihood = -1694.4825                       Pseudo R2       =     0.0071

------------------------------------------------------------------------------
       hours |      Coef.   Std. Err.       t     P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         nwi |  -.0108405    .1754488    -0.06    0.951    -.3554044    .3337234
          ms |  -9.308479    3.902876    -2.39    0.017    -16.97333   -1.643622
         age |  -1.114833    .4674299    -2.39    0.017    -2.032818   -.1968478
        race |   1.778366    3.726882     0.48    0.633    -5.540855    9.097587
        clt6 |  -6.346814    4.266223    -1.49    0.137    -14.72525    2.031617
        educ |   2.017293    .5898988     3.42    0.001     .8587914    3.175795
       _cons |   39.92167    22.07323     1.81    0.071    -3.427918    83.27127
-------------+----------------------------------------------------------------
         _se |   34.09747    1.611069            (Ancillary parameter)
------------------------------------------------------------------------------

  Obs. summary:        313  left-censored observations at hours<=0
                       297     uncensored observations
```

For the size simulations, u_i was generated from a normal$(0, _se^2)$ and

$$H_i = \begin{cases} H_i^* & \text{if } H_i^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

Similar to Skeels and Vella (1999), the power of the test was examined by five sets of simulations in which the u_i were generated from the distributions in Figure 1.[6] For each run of each simulation, the pseudorandom variates were scaled by the estimated *_se* from the tobit fit to the Moffit data. The details of how the data were generated are presented in Section 6.

Figure 1: Distributions used in power simulations.

t_1
t_5
$\chi^2_1 - 1$
$\chi^2_5 - 5$
$.4(\chi^2_1 - 1) + .6(\chi^2_{25} - 25)$

In the second set of simulations, the functional form and the average number of censored observations were arbitrarily fixed, while the sample size was allowed to vary over 100, 500, and 1,000. The data were generated according to

$$y_i^* = 1 + x_1 + x_2 + x_3 + u_i$$

[6]Skeels and Vella (1999) considered all of these distributions except for the $\chi^2_5 - 5$.

where

$$x_1 \sim N(0,1)$$

$$x_2 = .3 * x_1 + \epsilon_2 \text{ and } \epsilon_2 \sim N(0,1)$$

and

$$x_3 = .3 * x_1 + \epsilon_3 \text{ and } \epsilon_3 \sim N(0,1)$$

The censored y_i was then computed as

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

The same size and power simulations were run with this data generating process as with the Moffit data simulations.

4 Results

The results are presented in Tables 1–4. Each simulation was repeated 2,000 times. This implies that each of the estimates in the tables below has a standard error of $\sqrt{p * (1-p)/2000}$, where p is the estimate.

Table 1 contains the nominal sizes for the different simulations using the asymptotic critical values. As in previous studies, using the asymptotic critical values results in severely oversized tests. The fact that the test is severely oversized even at the sample size of 1,000 is noteworthy. The 95% confidence interval for the Moffit case at the 5% level is [.1248,.1552]. Since this interval contains the Skeels and Vella (1999) result of 15.34, this is a replication of their result.

Table 1: Asymptotic sizes.

Sample	%10	%5	%1
100	0.3730	0.3010	0.1950
500	0.2040	0.1400	0.0725
1000	0.1675	0.1080	0.0430
Moffit	0.2070	0.1400	0.0805

Table 2 contains the nominal sizes that result from using the bootstrap critical values. The results clearly indicate that 500 repetitions are sufficient to produce quite reasonable sizes.

Table 2: Bootstrap sizes.

Sample	%10	%5	%1
100	0.1045	0.0510	0.0120
500	0.0960	0.0535	0.0120
1000	0.0975	0.0470	0.0120
Moffit	0.0965	0.0545	0.0135

Table 3 contains the results from using the asymptotic critical values for each of the power simulations. The last column contains the Skeels and Vella (1999) results. Note that all their reported results were for the 5% level. Except for the chi-squared mixture, the 95% confidence interval for each of the Moffit cases at the 5% level contains the corresponding result obtained by Skeels and Vella (1999). As discussed in Drukker and Skeels (2002), Skeels and Vella (1999) obtained different results because of software limitations.[7]

(*Continued on next page*)

[7]The software package used by Skeels and Vella (1999) contained a chi-squared inversion algorithm that would sometimes crash when attempting to invert a value in the tails. For this reason, instead of inverting uniform[0,1] pseudo-variates, they generated the χ^2_2 pseudo-variates by inverting uniform[.04,.94] pseudo-variates. Drukker and Skeels (2002) show why throwing the tails causes the difference in the results. The package used by Skeels and Vella (1999) was not Stata, and the other software has been fixed.

Table 3: Asymptotic powers.

Simulation	Sample	%10	%5	%1	S&K
t_1	100	0.9975	0.9940	0.9785	NA
t_1	500	0.9995	0.9995	0.9995	NA
t_1	1000	1.0000	1.0000	1.0000	NA
t_1	Moffit	1.0000	1.0000	1.0000	1.0000
t_5	100	0.4490	0.3395	0.1950	NA
t_5	500	0.8610	0.7465	0.4475	NA
t_5	1000	0.9760	0.9335	0.7370	NA
t_5	Moffit	0.9170	0.8520	0.6555	.8523
$\chi^2_1 - 1$	100	1.0000	1.0000	1.0000	NA
$\chi^2_1 - 1$	500	1.0000	1.0000	1.0000	NA
$\chi^2_1 - 1$	1000	1.0000	1.0000	1.0000	NA
$\chi^2_1 - 1$	Moffit	1.0000	0.9995	0.9965	.9997
$\chi^2_5 - 5$	100	0.9345	0.8970	0.7650	NA
$\chi^2_5 - 5$	500	1.0000	1.0000	1.0000	NA
$\chi^2_5 - 5$	1000	1.0000	1.0000	1.0000	NA
$\chi^2_5 - 5$	Moffit	0.9270	0.8755	0.6880	NA
Mixture	100	0.5230	0.4260	0.2355	NA
Mixture	500	0.9200	0.8755	0.7165	NA
Mixture	1000	0.9950	0.9915	0.9700	NA
Mixture	Moffit	0.4605	0.3210	0.1275	.9631

Using the parametric bootstrap critical values produces the nominal powers in Table 4. The table contains three important results. First, although the test appears weak against t_5, most of the powers appear to be quite reasonable. Second, except for the χ^2 mixture and the χ^2_1, all the 95% confidence intervals around the Moffit-5% results contain the exact powers calculated by Skeels and Vella. The difference in the mixture case is due to the same issue discussed above and in Drukker and Skeels (2002). The difference in χ^2_1, while statistically significant, is not qualitatively significant. Both the Skeels and Vella (1999) and the present results indicate that this test has reasonable power against the χ^2_1 alternative in the Moffit case. Third, results for the χ^2_1, χ^2_5 and the mixture of χ^2 variables indicate that this conditional moment test has less power against skewed distributions with real-world data generating processes than with the simple process. Still, the increases in power caused by increasing the sample size with the simple process indicate that this test with bootstrap critical values should have reasonable power against skewed distributions and real-world data generating processes with sample sizes of 1,000 or more. Of course, given some actual data, one could repeat the design of the Moffit simulation to determine exactly how much power this test has against given skewed alternatives.

Table 4: Bootstrap powers.

Simulation	Sample	%10	%5	%1	S&K
t_1	100	0.9130	0.8315	0.6860	NA
t_1	500	0.9995	0.9980	0.9770	NA
t_1	1000	1.0000	0.9995	0.9955	NA
t_1	Moffit	1.0000	1.0000	0.9980	1.0000
t_5	100	0.0775	0.0305	0.0070	NA
t_5	500	0.5725	0.3125	0.0575	NA
t_5	1000	0.9170	0.7460	0.2765	NA
t_5	Moffit	0.7265	0.4830	0.0915	.4733
$\chi_1^2 - 1$	100	0.9970	0.9690	0.4080	NA
$\chi_1^2 - 1$	500	1.0000	1.0000	1.0000	NA
$\chi_1^2 - 1$	1000	1.0000	1.0000	1.0000	NA
$\chi_1^2 - 1$	Moffit	0.9945	0.9490	0.3850	.9868
$\chi_5^2 - 5$	100	0.4755	0.1925	0.0090	NA
$\chi_5^2 - 5$	500	1.0000	1.0000	1.0000	NA
$\chi_5^2 - 5$	1000	1.0000	1.0000	1.0000	NA
$\chi_5^2 - 5$	Moffit	0.7055	0.3855	0.0190	NA
Mixture	100	0.0965	0.0320	0.0070	NA
Mixture	500	0.8020	0.6215	0.2370	NA
Mixture	1000	0.9895	0.9725	0.8175	NA
Mixture	Moffit	0.1640	0.0370	0.0010	.8932

5 Conclusion

Since misspecifying the distribution or skedastic component in a maximum likelihood estimator of a categorical or limited dependent variable model results in inconsistent estimates, the derivation of a class of conditional moment tests against these alternatives originally caused great excitement. However, subsequent studies found that using the asymptotic critical values produced severely oversized tests.

This paper has presented evidence that using critical values obtained from a parametric bootstrap with 500 repetitions essentially removes the size distortions from the standard conditional moment test for non-normal disturbances after tobit estimation. The results also indicate that using the parametric bootstrap critical values with the conditional moment test for normality yield reasonable powers. This paper also replicated most of the tobit/normal results of Skeels and Vella (1999).

The next step is to investigate how much power the other OPG conditional moment tests investigated by Skeels and Vella (1999) have against their alternatives when using bootstrap critical values. This topic is under investigation by the author.

6 Appendix

6.1 The sample moment restrictions

Begin with the tobit estimates $\widehat{\beta}$ and $\widehat{\sigma}$ and residuals $\widehat{u}_i$. The estimates can be used to obtain

$$\widehat{z}_i = \frac{x_i'\widehat{\beta}}{\widehat{\sigma}}$$

$$\widehat{\lambda}_i = \frac{\phi(\widehat{z}_i)}{1-\Phi(\widehat{z}_i)}$$

where $\phi()$ is the standard normal density and $\Phi()$ is the standard normal CDF.

Since the data specify I_i according to

$$I_i = \begin{cases} 1 & \text{if } y_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

the sample moment restrictions are

$$m_i(\widehat{\theta}) = \left\{ \begin{array}{c} I_i\widehat{u}_i^3 - (1-I_i)(\widehat{z}_i^2+2)\widehat{\sigma}^3\widehat{\lambda}_i \\ I_i(\widehat{u}_i^4 - 3\widehat{\sigma}^4) + (1-I_i)(\widehat{z}_i^2+3)\widehat{\sigma}^4\widehat{\lambda}_i\widehat{z}_i \end{array} \right\}$$

Performing these computations for each observation i produces the $(N \times 2)$ matrix,

$$\widehat{M} = m_i(\widehat{\theta})'$$

6.2 An artificial regression for τ

Newey (1985) noted that one of the benefits of τ is that it could be easily calculated via an artificial regression. Specifically, he noted that $\tau = N - SSR$ where N is the number of observation in the artificial regression,

$$\iota = \widehat{M}\Gamma_1 + \widehat{S}\Gamma_2 + e$$

and SSR is the sum of squared residuals from this regression. This paper uses the new command `tobcm`, written by the author to compute τ. `tobcm` uses the artificial regression method to compute τ.

6.3 Generating the simulated data

All the simulations were run in Stata 7.0, and the ado- and do-files are available from the author's user site.

Since the functional forms are given in the text, it remains to discuss how the error terms were generated. For the size simulations, Figure 2 describes the method of generating the simulated normal variates.

Figure 2: Data generation in size simulations.

Case	Mathematical Description	Stata code
Moffit	$34.09747 * u_i \quad u_i \sim N(0,1)$	`gen double ui= se*invnorm(uniform())`
Simple specification	$2 * u_i \quad u_i \sim N(0,1)$	`gen double ui= 2*invnorm(uniform())`

Note that `se` in Figure 2 is a Stata scalar equal to the estimated `_b[_se]` from the tobit. The methods for generating the disturbances for the power simulations are given in Figure 3.

Figure 3: Data generation in power simulations.

Distribution	Mathematical Description	Stata Code	Scale
t_1	$s_1 \left\{ \frac{N(0,1)}{N(0,1)} \right\}$	`gen double t1= s1*(u1/u2)`	See Note 1
t_5	$s_2 \left\{ \frac{N(0,1)}{(\frac{\chi^2(5)}{5})^{\frac{1}{2}}} \right\}$	`gen double t5= s2*(u3/(chic))`	$b\sqrt{(5/3)}$
$\chi^2_1 - 1$	$s_3 \left\{ \chi^2(1) - 1 \right\}$	`gen double chia= s3*(chi2_1-1)`	$b * \sqrt{2}$
$\chi^2_5 - 5$	$s_4 \left\{ \chi^2(5) - 5 \right\}$	`gen double chib= s4*(chi2_5-5)`	$b * \sqrt{10}$
$.4(\chi^2_1 - 1)$ $+.6(\chi^2_{25} - 25)$	$s_5 \left[.4(\chi^2(1) - 1) +.6\{\chi^2(25) - 25\} \right]$	`gen double mix = s5*(.4*(chi2_1-1) +.6*(chi2_25-25))`	$b * \sqrt{.16 * 2 + .36 * 50}$

Note 1: Neither the mean nor the variance exist for the Cauchy distribution. No adjustment was made for location. The generated variable was scaled by its standard deviation and multiplied by b.
Note 2: `u1`, `u2`, `u3` are pseudo-standard normals created by `gen double ui = invnorm(uniform())`,
Note 3: b is the estimated standard error in the Moffit case, and b is 2 in the simple case
Note 4: `chic = sqrt(chi2_5)/5)`
Note 5: `gen double chi2_df = invchi2(df,uniform());` df=$\{1, 5, 25\}$

7 Acknowledgment

I would like to thank Chris Skeels for making the Moffit (1984) data available to me and for hunting down the code he used to perform his simulations.

8 References

Drukker, D. M. and C. L. Skeels. 2002. Monte Carlo evidence on the robustness of conditional moment tests in tobit and probit models: A comment. Tech. rep., Department of Economics, The University of Melbourne.

Efron, B. and R. J. Tibshirani. 1993. *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Ericson, P. and J. Hansen. 1999. A note on the performance of simple specification tests for the tobit model. *Oxford Bulletin of Economics and Statistics* 61: 121–127.

Lee, L. F. and G. S. Maddala. 1985. The common structure of tests for selectivity bias, serial correlation, heteroskedasticity and non-normality in the tobit model. *International Economic Review* 26: 1–20.

Moffit, R. 1984. The estimation of a joint wage-hours labor supply model. *Journal of Labor Economics* 2: 550–566.

Newey, W. 1985. Maximum likelihood specification testing and conditional moment tests. *Econometrica* 53: 1047–1073.

Pagan, A. and F. Vella. 1989. Diagnostic tests for models based on individual data: a survey. *Journal of Applied Econometrics* 4: 229–259.

Skeels, C. and F. Vella. 1997. Monte Carlo evidence on the robustness of conditional moment tests in tobit and probit models. *Econometric Reviews* 16: 69–92.

—. 1999. A Monte Carlo investigation of the sampling behavior of conditional moment tests in tobit and probit models. *Journal of Econometrics* 92: 275–294.

Tauchen, G. 1985. Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics* 30: 415–443.

White, H. 1982. Maximum likelihood estimation of misspecified models. *Econometrica* 50: 1–26.

Wooldridge, J. 2001. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

About the Author

David M. Drukker is the Director of Econometrics at Stata Corporation.