Conditional Moments Tests for Tobit and Probit in STATA

Måns Söderbom*

November 2001

Abstract

It is well known that standard discrete choice models, such as logit and probit, and limited dependent variable models, such as tobit, fail to give consistent parameter estimates if the statistical assumptions concerning the error terms are incorrect (see e.g. Davidson and MacKinnon, Chapter 15). In applied work it is therefore of some interest to explore if such assumptions hold. One popular approach is that of conditional moment (CM) tests (Newey, 1985; Tauchen, 1985), which examine whether the relevant sample moments are supported by the data. This note outlines the theory behind these tests, explains how they can carried out in STATA for tobit and binary and ordered probit, and provides an illustration using real data.

*Centre for the Study of African Economies, Department of Economics, University of Oxford, Manor Road, Oxford OX1 3UQ. Email: `mans.soderbom@economics.ox.ac.uk`. Fax: +44(0)1865-281447. Tel: +44(0)1865-271954. Måns Söderbom is the UNIDO Fellow at the CSAE.

1. Conditional Moment Tests

Let $\ln L(y_i;\beta)$, $i = 1, ..., N$, be the contribution of individual i to the sample log-likelihood, where β is a $k \times 1$ parameter vector and the y_i are independently distributed. Define $S = \frac{\partial \ln L(y_i;\beta)}{\partial \beta}$, the $N \times k$ matrix of scores evaluated at the maximum likelihood estimate $\widehat{\beta}$, and let $m_i(\beta)$ be an $r \times 1$ vector of functions such that $H_0 : E[m_i(\beta_0)] = 0, i = 1, ..., N$, is the null hypothesis to be tested (where β_0 is the 'true' parameter value under H_0). Thus, the null hypothesis is that the expected value of one or several moments is zero. For instance, if we want to test the null hypothesis that the residuals are normally distributed, we may test $H_0 : E[\varepsilon^3] = 0$ (i.e. no skewness) and $E[\varepsilon^4 - 3\sigma^4] = 0$ (no kurtosis). To take expectations, we use all observations in the estimation sample and compute N vectors $m_1(\beta), ..., m_N(\beta)$, and then stack these to form the $N \times r$ matrix $M = \left(m_1\left(\widehat{\beta}\right), ..., m_N\left(\widehat{\beta}\right)\right)^T$. Newey (1985) and Tauchen (1985) showed that asymptotically and under certain regularity conditions,

$$\tau = \iota^T Z \left(Z^T Z\right)^{-1} Z^T \iota \sim \chi_r^2 \tag{1.1}$$

where $\iota = (1, ..., 1)$ is an $N \times 1$ vector of ones, $Z = M - SH\frac{\partial}{\delta\beta}N^{-1}\sum_i m_i(\beta)$, and H is the estimated Hessian matrix. At first sight this expression may look complicated to compute. In fact, it is not. To see this, imagine first that somehow we have been able to calculate the matrix Z. Equipped with this matrix we see directly from 1.1 that τ can be calculated as the explained sum of squares (ESS) from an OLS regression with ι as the 'dependent' variable and Z as the 'independent' variables.[1] This is called an 'artificial regression'. All statistical packages report ESS, so given that we can obtain Z, all we have to do is to generate a 'variable' equal to 1, regress this on Z (without a constant) and look at the calculated ESS. Becuase the dependent variable in the artificial regression is a vector of ones, it follows that $\tau = ESS = N - SSE = N \times R^2$, where SSE is the residual sum of squares and R^2 is the coefficient of determination. Hence there are several numerically identical ways to calculate τ.

So how can we calculate Z? This is slightly complicated, fortunately it has been shown how regressing ι on Z can be thought of as a restricted version of the regression of ι on (S, M) (Pagan and Vella, pp. S33-S34). This will be the route taken here.

To sum up, calculating τ involves the following step: i) estimate the model with maximum likelihood (MLE); ii) calculate the score matrix S, using the MLE estimates; calculate the relevant elements of the moment matrix M, using the MLE estimates; iv) regress ι on (S, M) and calculate the resulting $\tau = N \times R^2$ (for instance). The next section outlines how this can be done in STATA.

2. STATA Code for Calculation of CM Test Statistic

To calculate test statistics such as the one in 1.1, I have found STATA particularly useful. The ado file accompanying this note, cmtest.ado, enables the user to compute various CM tests for tobit and binary and ordered probit, by submitting to STATA only one line of

[1] If we regress ι on Z, we obtain parameter estimates φ. The explained sum of squares from this regression is $(Z^T\varphi)^T (Z^T\varphi) = \left(Z^T (Z^T Z)^{-1} Z^T \iota\right)^T \left(Z^T (Z^T Z)^{-1} Z^T \iota\right)$, which is equal to $\iota^T Z (Z^T Z)^{-1} Z^T \iota$.

instructions.[2] The program can be used with STATA 6.0 or later versions. In this section I explain how the program works, and how to make it run.

2.1. How to make cmtest run

First, some preparations:

1. Copy cmtest.ado to the folder C:\ado - if this folder does not exist, it will have to be created.[3]

2. Organise the dataset. The current version of the program does not automatically discard observations with missing values, so it is strongly recommended that the user deletes all such observations before running the program.

3. Notice that the variable names 'dfdb', and 'D1', 'D2',... are reserved for use within the ado-file. If you happen to have varibles with exactly these names, they would have to be re-named or otherwise the program will not run properly.

The program computes tests for three different models: tobit, binary probit and ordered probit. The user will have to specify which one of these he or she wants.[4] This is done by setting *model* to 1 = tobit, 2 =binary probit or 3 = ordererd probit. Unless this is specified, the program will run the binary probit. Notice also that, for the tobit, the

[2] An ado-file is simply an ASCII text file that contains the code.

[3] The reason is that cmtest.ado is a user-defined program, and STATA by default looks for such files in C:\ado. If you put the file elswhere, STATA will not find it.

[4] Naturally, anything that will result in estimation problems for the underlying model will cause similar problems in the computation of test statistics. It is therefore a good idea to estimate the model separately first, before running cmtest.

current version of the program only accommodates lower censoring at zero. Adjusting the program in order to allow for a more general type of censoring is not difficult.

With the above in mind, cmtest works just like the regression commands in STATA: the user types 'cmtest' followed by the dependent variable and then the explanatory variables. For instance, say that we have binary data on whether a person works or not (variable name: pworks) and we want to model this decision using a binary probit, where the person's education (educ) and age (age) are explanatory variables. To compute CM tests for heteroskedasticity and normality for this specification, we submit to STATA the instruction

cmtest pworks educ age, model(2)

The program will then estimate the probit model, and compute and report a number of test statistics.

2.2. What does the program do?

Here is a brief description of what the cmtest program does. For more details, look at the actual program.

Step 1: Estimate the model and compute the score matrix, S. One nice feature of STATA is that we easily can obtain estimates of $\frac{\partial \ln L(y_i;\beta)}{\partial x_i^T \beta}$, by using the 'score' option. With this in hand, we can then compute the score vector for individual i as $\frac{\partial \ln L(y_i;\beta)}{\partial x_i^T \beta} x_i$, and then stack all N vectors to form the $N \times k$ score matrix.

Step 2: Compute the individual contributions to the moment conditions. For this we need the residuals. However, when data are discrete or censored, conventional residuals (such as $y - x^T \widehat{\beta}$ in the OLS model) cannot be computed. Nevertheless it is possible to calculate *generalised residuals*. Consider the probit model, for instance. If the j^{th} observation of the dependent variable is equal to zero, then we know from the model that $\varepsilon_j < -x_j^T\widehat{\beta}$. With this information, we can obtain an estimate of the residual. Because of the normality assumption, the expected value of the residual, or the generalised residual, is $E\left[\varepsilon_j \mid \varepsilon_j < -x_j^T\widehat{\beta}\right] = \frac{\phi\left(x_j^T\widehat{\beta}\right)}{1-\Phi\left(x_j^T\widehat{\beta}\right)}$, where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and the cumulative density function (CDF) of the standard normal distribution. Similarly, if the k^{th} observation of the dependent variable is equal to one, we have $E\left[\varepsilon_k \mid \varepsilon_k \geq -x_k^T\widehat{\beta}\right] = \frac{\phi\left(x_k^T\widehat{\beta}\right)}{\Phi\left(x_k^T\widehat{\beta}\right)}$. To test for homoskedasticity or normality, we need higher moments. Specifically, the cmtest program uses the following moment conditions

1. Homoskedasticity: $E\left[z_j\left[\left(y_i^* - x_j^T\beta\right)^2 - \sigma^2\right]\right] = 0$,
2. Normality: $E\left[\left(y_i^* - x_j^T\beta\right)^3\right] = 0$, and $E\left[\left(y_i^* - x_j^T\beta\right)^4 - 3\sigma^4\right] = 0$

The generalised residuals for moments 1, 2, 3 and 4 are as follows for the binary probit model:

Residual	Binary Probit
$\widehat{e}^{(1)}$	$-D_1 h\left(x^T\widehat{\beta}\right) + D_2 h\left(-x^T\widehat{\beta}\right)$
$\widehat{e}^{(2)}$	$-x^T\widehat{\beta}\widehat{e}^{(1)}$
$\widehat{e}^{(3)}$	$\left(2 + \left(x^T\widehat{\beta}\right)^2\right)\widehat{e}^{(1)}$
$\widehat{e}^{(4)}$	$\left(3x^T\widehat{\beta} + \left(x^T\widehat{\beta}\right)^3\right)\widehat{e}^{(1)}$

where $D_1 = 1$ if $y = 0$, $D_1 = 0$ if $y > 0$, $D_2 = 1 - D_1$ and $h(\cdot) = \frac{\phi(\cdot)}{1-\Phi(\cdot)}$. For the tobit,

Residual	Tobit
$\widehat{e}^{(1)}$	$-D_1 h\left(\frac{x^T\widehat{\beta}}{\widehat{\sigma}}\right) + D_2 h\left(\frac{y-x^T\widehat{\beta}}{\widehat{\sigma}}\right)$
$\widehat{e}^{(2)}$	$-D_1 \frac{x^T\widehat{\beta}}{\widehat{\sigma}} h\left(\frac{x^T\widehat{\beta}}{\widehat{\sigma}}\right) + D_2\left(\left(\frac{y-x^T\widehat{\beta}}{\widehat{\sigma}}\right)^2 - 1\right)$
$\widehat{e}^{(3)}$	$-D_1\left(2+\left(\frac{x^T\widehat{\beta}}{\widehat{\sigma}}\right)^2\right) h\left(\frac{x^T\widehat{\beta}}{\widehat{\sigma}}\right) + D_2\left(\frac{y-x^T\widehat{\beta}}{\widehat{\sigma}}\right)^3$
$\widehat{e}^{(4)}$	$-D_1\left(3\left(\frac{x^T\widehat{\beta}}{\widehat{\sigma}}\right)+\left(\frac{x^T\widehat{\beta}}{\widehat{\sigma}}\right)^3\right) h\left(\frac{x^T\widehat{\beta}}{\widehat{\sigma}}\right) + D_2\left(\left(\frac{y-x^T\widehat{\beta}}{\widehat{\sigma}}\right)^4 - 3\right)$

The generalised residuals for the ordered probit are generalisations of those of the binary probit, and will not be shown here.

Step 3: With S and M available all we have to do is to run the artificial regressions as described in Section 2, and compute the test statistics. This is done in the final part of the program.

3. An Application

A recent paper by published in *Journal of Applied Econometrics* (Martins, 2001) estimates wage and participation equations, using data on married women in Portugal. The author's dataset is publicly available at http://qed.econ.queensu.ca/jae/2001-v16.1/, and we shall use it here to illustrate how the CM tests work.

We focus on the participation equation, where the decision to work is modelled as a function of CHILD (the number of children under 18 living in the family), YCHILD (the number of children under the age of 3 living in the family), HW (log of monthly husband's wage), EDU (years of education), AGE (age in years divided by 10) and AGE2 (AGE squared) using the probit model.

Table 1 reports estimates of the coefficients and standard errors that replicate those reported by Martins in Table I, and in addition CM tests for homoskedasticity and normality. Clearly, the probit is mis-specified, with ample evidence of both heteroskedasticity and non-normality. The author acknowledges this, and proceeds by estimating a semi-parametric binary choice model.

Table 1: Probit Estimates and CM Tests

Variables	Coeff.	z-value	p-value, homosk.
CHILD	-0.126	4.46	0.744
YCHILD	-0.074	0.99	0.026
HW	-0.076	0.97	0.514
EDU	0.141	14.92	0.092
AGE	0.908	3.57	0.007
AGE2	-0.137	4.33	
Intercept	-0.988	1.04	
log Likelihood	-1371.7		
Homoskedasticity (p-value)	0.002		
Normality (p-value)	0.004		

4. References

Davidson, R. and J. G. MacKinnon (1993). *Estimation and Inference in Econometrics.* New York, Oxford: Oxford University Press.

Martins, M. F. O. (2001). "Parametric and semiparametric estimation of sample selection models: an empirical application to the female labour force in Portugal," *Journal of Applied Econometrics* 16, pp. 23-39.

Newey, W. K. (1985). "Maximum likelihood specification testing and conditional moment tests," *Econometrica* 53, pp. 1047-1073.

Pagan, A. and F. Vella (1989). "Diagnostic tests for models based on individual data: a survey," *Journal of Applied Econometrics* 4, pp. S29-S59.

Tauchen, G. (1985). "Diagnostic testing and evaluation of maximum likelihood models," *Journal of Econometrics* 30, pp. 415-443.