Center
for
Economic Research

No. 2000-28

A MAXIMUM LIKELIHOOD ESTIMATOR BASED ON FIRST DIFFERENCES FOR A PANEL DATA TOBIT MODEL WITH INDIVIDUAL SPECIFIC EFFECTS

By Adriaan S. Kalwij

March 2000

ISSN 0924-7815

A Maximum Likelihood Estimator Based on First Differences for a Panel Data Tobit Model with Individual Specific Effects

Adriaan S. Kalwij *

Department of Economics
University of Oxford

February 2000

Abstract

This paper proposes an alternative estimation procedure for a panel data Tobit model with individual specific effects based on taking first differences of the equation of interest. This helps to alleviate the sensitivity of the estimates to a specific parameterization of the individual specific effects and some Monte Carlo evidence is provided in support of this. To allow for arbitrary serial correlation estimation takes place in two steps: Maximum Likelihood is applied to each pair of consecutive periods and then a Minimum Distance estimator is employed.

JEL-classification: C23

* Mailing address: Department of Economics, University of Oxford, Manor road, Oxford OX1 3UL, United Kingdom. Email: adriaan.kalwij@economics.ox.ac.uk

1. Introduction

This paper is concerned with the estimation of a panel data Tobit model in which the unobserved individual specific effects are allowed to correlate with the explanatory variables. More specifically, this paper proposes a Maximum Likelihood estimator based on taking first differences of the equation of interest in order to alleviate the sensitivity of the estimates to a specific parameterization of the individual specific effects. With respect to previously proposed parametric estimators for censored regression panel data model this set up can be regarded a bias reduction strategy for the possible bias caused by misspecification of the individual specific effects.

Nijman and Verbeek (1992) and Zabel (1992) propose a Maximum Likelihood estimator for panel data selection models in which the individual specific effects are allowed to correlate with the explanatory variables. To estimate such models Wooldridge (1995) proposes a two-step estimator, in the spirit of the Heckman (1976), using fewer distributional assumptions and allowing for arbitrary serial correlation. These estimators can be applied to a panel data Tobit model with individual specific effects. To allow for correlation between the individual specific effects and the explanatory variables both estimators, following Mundlak (1978) and Chamberlain (1984), explicitly model this correlation by assuming a specific parameterization of the individual specific effects as a function of the explanatory variables and random individual specific effects. A convenient and often made choice is to model the individual specific effects as a linear combination of the averages over time of the explanatory variables plus random individual specific effects. Intuitively this is an appealing approach since in the absence of censoring this yields the familiar ‘within’ estimates (see Mundlak, 1978).

Unlike in a linear regression model in a censored regression model consistency of the estimates is based on the assumption of correctly specified individual specific effects. To overcome this problem, Honoré (1992) proposes a trimmed Lead Absolute Deviations estimator requiring less parametric assumptions and allowing for arbitrary correlation between the explanatory variables and the individual specific effects (i.e. a fixed effects specification). The costs of applying this estimator are considerable. Firstly, in contrast to the parametric approaches discussed above, one cannot identify the marginal effects of the explanatory variables on the dependent variable. From a policy point of view it may be insufficient to have only parameter estimates. Secondly, this estimator depends on other assumptions concerning the error terms and, consequently, does not allow for arbitrary serial correlation.[1] And thirdly, from an empirical point

[1] The estimator proposed by Honoré (1992) relies on the so-called conditional symmetry assumption.

of view it is fair to say that this estimator is relatively difficult to implement which considerably restricts widespread application.[2] A possible parametric solution to alleviate the sensitivity of the parameter estimates to the parameterization of the individual specific effects is to allow for a more flexible parameterization[3]. This may, however, increase the number of parameters dramatically.

An alternative parametric solution proposed in this paper is to start by eliminating the individual specific effects from the equation of interest and setting up the likelihood function based on taking first differences of the equation of interest. Following the studies mentioned above, in the selection part of the model the individual specific effects are parameterized as a function of the explanatory variables and random individual specific effects. To allow for arbitrary serial correlation estimation takes place in two steps: Maximum Likelihood is applied to each pair of consecutive periods and then a Minimum Distance estimator is employed to obtain estimates of the parameters of interest. This alternative parametric approach yields parameter estimates that are less sensitive to a specific parameterization of the individual specific effects relatively to using a standard Tobit model. Monte Carlo evidence is provided in support of this. Also the estimation procedure is relatively easy to carry out, hence may provide a powerful tool for analyzing censored panel data.[4]

[2] An application of this estimator can be found in Charlier et al. (2000).

[3] For instance, following Chamberlain (1984) and Wooldridge (1995) one can parameterise the individual specific effects as a linear function of all past and future exogenous variables. See also Zabel (1992).

[4] A Gauss program is available from the author upon request.

2. A Panel Data Tobit Model with Individual Specific Effects

The model of interest is formulated as follows:

$$
\begin{aligned}
(1) \qquad & y_{it}^{*} = X_{it} b + a_i + e_{it} , \\
& y_{it} = \max\left(0, y_{it}^{*}\right), \qquad\qquad i = 1,..,N \ , \ t = 1,..,T \ ,
\end{aligned}
$$

Where the individual is indexed by *i*, the time period by *t*. X_{it} is a (1xK) vector of exogenous variables, b is a (Kx1) vector of the parameters of interest and a_i is an unobserved individual specific effect that may be correlated with X_{it}. The latent dependent variable is censored at zero and only y_{it} is observed. The error term e_{it} is assumed to be Normal distributed with mean zero and variance $s_{e,t}^{2}, e_{it} \sim N\left(0, s_{e,t}^{2}\right)$, and is allowed to be arbitrary serially correlated. The panel data is characterized by having a large number of individuals over a short period of time.

Following Mundlak (1978), Zabel (1992) and Nijman and Verbeek (1992) specify the individual specific effect as a linear function of the averages over time of all exogenous variables plus a random individual specific effect that is assumed to be independent of the explanatory variables:

$$
(2) \qquad a_i = \overline{X}_i g + m_i , \qquad\qquad \overline{X}_i = \frac{1}{T} \sum_{s=1}^{T} X_{is} \ .
$$

The random individual specific effect, m_i, is assumed Normal distributed with mean zero and variance $s_{m}^{2}, m_i \sim N(0, s_{m}^{2})$.

2.1 Model A

Substituting equation (2) in model (1) yields:

$$
\begin{aligned}
(3) \qquad & y_{it}^{*} = X_{it} b + \overline{X}_i g + u_{it} \ , \\
& y_{it} = \max(0, y_{it}^{*}) , \qquad\qquad i = 1,..,N \ , \ t = 1,..,T \ .
\end{aligned}
$$

Where $u_{it} = m_i + e_{it}$ with, given the distributional assumptions, $u_{it} \sim N(0, s_t^2)$ where $s_t^2 = s_m^2 + s_{e,t}^2$. The estimation is done in two steps in order to take into account arbitrary serial correlation. First one obtains Maximum Likelihood estimates per period (see, e.g., Tobin, 1958):

(4) $$\hat{q}_t = \arg\max_q \sum_{i=1}^{N} (1 - I_{(y_{it}>0)}) \ln\left(L_{it}^1\right) + I_{(y_{it}>0)} \ln\left(L_{it}^2\right)$$

Where $q_t = (b_t, g_t, s_t)$, $L_{it}^1 = \Phi\left(\frac{-(X_{it}b + \overline{X}_i g)}{s_t}\right)$ and $L_{2i} = \frac{1}{s_t} f\left(\frac{y_{it} - (X_{it}b + \overline{X}_i g)}{s_t}\right)$.

The cumulative standard Normal distribution is denoted by $\Phi(.)$ and the standard Normal distribution by $f(.)$. Next, a Minimum Distance estimator using the optimal weighting matrix is employed to impose the restrictions { $b_t = b$ and $g_t = g$, $\forall t$ } (see, e.g., Chamberlain, 1984). From an empirical point of view, this estimator is quite appealing since it is relatively easy to implement. As discussed in the introduction, consistency depends on correctly specified individual specific effects (equation (2)).

2.2 Model B

In order to alleviate the sensitivity of the parameter estimates to a specific parameterization of the individual specific effects this paper proposes to start by eliminating the individual specific effects from the main equation by taking first differences:

(5) $$\Delta y_{it}^* = \Delta X_{it} b + h_{it} \qquad i = 1,..,N,\ t = 1,..,T,$$

$$\Delta y_{it} = \begin{cases} \Delta y_{it}^* & if \quad y_{is}^* > 0 \quad and \quad y_{it}^* > 0 \\ unobserved & otherwise \end{cases}.$$

Where *s=t-1*, $\Delta y_{it}^* = y_{it}^* - y_{is}^*$, $\Delta X_{it} = X_{it} - X_{is}$ and $h_{it} = u_{it} - u_{is} (\equiv e_{it} - e_{is})$. The correlation between u_{is} and u_{it} is denoted by r_t and, as previously defined, $u_{it} \sim N(0, s_t^2)$. Given the distribution assumptions: $h_{it} \sim N\left(0, s_{h,t}^2\right)$ with $s_{h,t}^2 = s_s^2 - 2 r_t s_s s_t + s_t^2$. Note that applying a Least Squares to equation (5) on a sample of positive values of y_{is} and y_{it} yields inconsistent estimates since $E\left[h_{it} \mid y_{is}^{*} > 0, y_{it}^{*} > 0\right] \neq 0$.

The probability of observing positive values of the dependent variable in both period s and t is given by:

$$(6) \qquad L_{it}^{1} \equiv \Pr(y_{is}^{*} > 0, y_{it}^{*} > 0 \mid X_{i1},..,X_{iT};q) = \Phi_{2}\left(\frac{X_{is}b + \overline{X}_{i}g}{s_{s}}, \frac{X_{it}b + \overline{X}_{i}g}{s_{t}}, r_{t} \right).$$

Where $q = (b, g, s_{1},.., s_{T}, r_{2},.., r_{T})$ and $\Phi_2(.)$ denotes the cumulative bivariate standard Normal distribution.

The truncated distribution of $\left(\Delta y_{it} - \Delta X_{it} b\right)$ is given by (see appendix):

$$(7) \qquad L_{it}^{2} = \frac{1}{s_{h,t}} f\left(\frac{\Delta y_{it} - \Delta X_{it} b}{s_{h,t}} \right) \times \Phi\left(\min\left(\frac{X_{is}b + \overline{X}_{i}g + \frac{s_{hs}}{s_{h,t}^{2}}\left(\Delta y_{it} - \Delta X_{it} b\right)}{\sqrt{s_{s}^{2} - s_{hs}^{2} s_{h,t}^{-2}}}, \frac{X_{it}b + \overline{X}_{i}g + \frac{s_{ht}}{s_{h,t}^{2}}\left(\Delta y_{it} - \Delta X_{it} b\right)}{\sqrt{s_{t}^{2} - s_{ht}^{2} s_{h,t}^{-2}}} \right)\right),$$

Where $s_{hs} = r_{t} s_{s} s_{t} - s_{s}^{2}$ and $s_{ht} = s_{t}^{2} - r_{t} s_{s} s_{t}$.

Based on equations (6) and (7), the Maximum Likelihood estimate of $q_{t} = (b_{t}, g_{t}, s_{s}, s_{t}, r_{t})$ is given by:

$$(8) \qquad \hat{q}_{t} = \arg\max_{q} \sum_{i=1}^{N} (1 - I_{(y_{is} > 0, y_{it} > 0)}) \ln\left(1 - L_{it}^{1}\right) + I_{(y_{is} > 0, y_{it} > 0)} \ln\left(L_{it}^{2}\right)$$

Next, a Minimum Distance estimator is employed to impose the restriction { $b_{t} = b$ and $g_{t} = g$, $\forall t$ } using the optimal weighting matrix. Note that only the β parameters corresponding to the time varying regressors are identified. For a time constant regressor only the sum (say, $\beta_2 + \gamma_2$) is identified.

Table: Simulation results. MB is the mean bias, RMSE is the root mean squared error, MedB is the median bias and MAD is the median absolute deviation.

β=1		Model A				Model B			
N	T	MB	RMSE	MedB	MAD	MB	RMSE	MedB	MAD
500	2	0.104	0.079	0.106	0.106	0.010	0.093	0.005	0.066
500	4	0.141	0.064	0.136	0.136	-0.005	0.051	-0.006	0.029
500	8	0.102	0.040	0.100	0.100	-0.003	0.032	-0.004	0.022
1000	2	0.084	0.065	0.087	0.090	-0.008	0.084	-0.014	0.055
1000	4	0.142	0.043	0.137	0.137	0.001	0.033	-0.000	0.019
1000	8	0.095	0.025	0.096	0.096	-0.002	0.021	-0.001	0.012

The data is generated as follows: $y_{it} = \max(0, 0.2 + X_{it} + a_i + e_{it})$. So the true value of β is 1. $X_{it} = 0.8X_{it-1} + x_{it}$, $e_{it} = 0.4e_{it-1} + z_{it}$, $X_{i1} = x_{i1}$, $e_{i1} = z_{i1}$, All three error terms (e_{it}, x_{it}, z_{it}) are N(0,1) distributed. The individual specific effect is <u>non-linear</u> in the time-averages of the explanatory variables: $a_i = \bar{X}_i \mid \bar{X}_i \mid + m_i$, $m_i \sim N(0,1)$, $i = 1,..,N$, $t = 1,..,T$.

3. A Monte Carlo experiment

The main idea behind setting up a panel data Tobit model in first differences has been to reduce the bias due to misspecification of the individual specific effects (equation (2)). A Monte Carlo study is carried out to provide some empirical support for this notion. Studies referring to a parametric estimator for a censored regression model usually have Model A of section 2.1 in mind. Therefore the estimator based on first differences, i.e. Model B of section 2.2, is compared with Model A. Details of the design are given at the bottom of the table. The simulations are based on 100 replications and the values chosen for N and T are, respectively, {500, 1000} and {2, 4, 8}. The models A and B as outlined in section 2 are estimated using the parameterization of the individual effects as specified in equation (2). Given the design, both models are misspecified and the simulation results reported in the table provide some measure of the relative performance of the two models under misspecification of the individual specific effects. Of course, the results in the table have to be interpreted with caution since they may depend on the design chosen.

As has been put forward in the literature the simulation results show that the parameter estimate of β using Model A is sensitive to misspecification of the individual specific effects. Although the simulation results (in particular the MAD) show that both estimators yield inconsistent estimates, the estimator based on first differences (i.e. model B) is less sensitive to misspecification of the individual specific effect. The bias reduction when using model B instead of model A is substantial (up to 80%) in this particular example. Of course, in empirical studies one can test, similar to the suggestion of Zabel (1992), whether or not a more flexible parameterization of the individual specific effects is needed.

4. Some concluding remarks

The benefit of using model B instead of model A is that the parameter estimates are less sensitive to a specific parameterization of the individual specific effects. The cost associated with this is that estimating model B demands more from the data than estimating model A since one only uses those individuals that are observed in two consecutive periods and identification of b is largely based on observing positive values of the dependent variables in two consecutive periods. While Model A is straightforward extension of the normal censored regression model as formulated by Tobin (1958) by exploiting the fact one has panel data, it seems inappropriate to classify Model B as such. The main equation of Model B, i.e. equation (5), only includes time varying regressors and the selection part also includes time constant variables. For this reason one should perhaps classify this type of model as a hybrid model, i.e. a model in between the classical Tobit and sample selection models.

As in the standard Tobit model formulated by Tobin (1958) the Normality assumption is needed for consistency. If Normality is too strong of an assumption then for both models a two-step estimator in each period yields consistent estimates under less distributional assumptions (only Normality in the first step). The estimator of Wooldridge (1995) can be taken as a two-step estimator of model A and a parametric version of the estimator proposed by Rochina-Barrachina (2000) can be taken as a two-step estimator for model B. Of course, if Normality is not too strong of an assumption, a two-step estimator leads to severe loss of efficiency.[5] In this respect, the Maximum Likelihood estimator proposed in this paper (model B) is considered complementary to these two-step estimators in the specific case of a panel data Tobit model.

[5] Simulation results not reported here show that the estimates are not very sensitive to violations of the Normality assumption.

Acknowledgements

I would like to thank Maria Rochina-Barrachina and Jeff Wooldridge for very helpful comments. Financial support of the Leverhulme Trust project “The labour market consequences of technical and structural change” is gratefully acknowledged.

References

Chamberlain, G., 1984, Panel Data, in Z. Griliches and M.D. Intriligator (eds), Handbook of Econometrics, Vol. II, North-Holland, Amsterdam, 1247-1318.

Charlier, E., B. Melenberg, and A. van Soest, 2000, Estimation of a censored regression panel data model using conditional moment restrictions efficiently, Journal of Econometrics, 95, 25-56.

Heckman, J.J., 1976, The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models, Annals of Economic and Social Measurement, 5, 475-492.

Honoré, B.E., 1992, Trimmed LAD and Least Squares estimation of truncated and censored regression models with fixed effects, Econometrica, 60, 533-565

Mundlak, Y., 1978, On the Pooling of Time Series and Cross Section Data, Econometrica, 46, 69-86.

Nijman, T. and M. Verbeek, 1992, Nonresponse in panel data: the impact on estimates of a life-cycle consumption function, Journal of Applied Econometrics, 7, 243-257.

Rochina-Barrachina, M.E., 2000, A new estimator for panel data sample selection models, Annales d’Economie et de Statistique, 55/56, 153-181.

Tobin, J., 1958, Estimation of relationships for limited dependent variables, Econometrica, 26, 24-36.

Wooldridge, J.M., 1995, Selection corrections for panel data models under conditional mean independence assumptions, Journal of Econometrics, 68, 115-132.

Zabel, J., 1992, Estimating fixed and random effects with selectivity, Economics Letters 40, 269-272.

Appendix: The truncated distribution of h_{it}

A density function is denoted by *f*(.), the cumulative standard Normal distribution is denoted by Φ(.) and the standard Normal distribution is denoted by f(.). The truncated distribution of h_{it} is denoted by $g(h_{it})$ and is given by:

$$g(h_{it}) = f\left(h_{it} \mid u_{is} > \left(-X_{is}b - \overline{X}_i g\right), u_{it} > \left(-X_{it}b - \overline{X}_i g\right)\right)$$

$$,, \quad = \int_{-X_{it}b-\overline{X}_i g}^{\infty} \int_{-X_{is}b-\overline{X}_i g}^{\infty} f(h_{it}, u_{is}, u_{it}) \partial u_{is} \partial u_{it}$$

$$,, \quad = f(h_{it}) \int_{-X_{it}b-\overline{X}_i g}^{\infty} \int_{-X_{is}b-\overline{X}_i g}^{\infty} f(u_{is}, u_{it} \mid h_{it}) \partial u_{is} \partial u_{it} \; .$$

Given the distributional assumptions made in section 2, the first density function at the right hand side is a Normal density function with mean 0 and variance $s^2_{h,t}$. The second term on the right hand side involves a degenerative distribution since $h_{it} = u_{it} - u_{is}$ and can be written as follows:

$$\int_{-X_{it}b-\overline{X}_i g}^{\infty} \int_{-X_{is}b-\overline{X}_i g}^{\infty} f(u_{is}, u_{it} \mid h_{it}) \partial u_{is} \partial u_{it} = \int_{-X_{it}b-\overline{X}_i g}^{\infty} f(u_{it} \mid h_{it}) \int_{-X_{is}b-\overline{X}_i g}^{\infty} f(u_{is} \mid u_{it}, h_{it}) \partial u_{is} \partial u_{it} \, ,$$

$$,, \quad = \int_{-X_{it}b-\overline{X}_i g}^{\infty} f(u_{it} \mid h_{it}) \int_{-X_{is}b-\overline{X}_i g}^{\infty} I\left(u_{is} = x\right) \partial u_{is} \partial u_{it} \, ,$$

$$\text{With} \quad x = \frac{s_{hs}}{s^2_h} h_{it} + \left(s^2_s - \frac{s^2_{hs}}{s^2_{h,t}}\right)^{1/2} \left(s^2_t - \frac{s^2_{ht}}{s^2_{h,t}}\right)^{-1/2} \left(u_{it} - \frac{s_{ht}}{s^2_{h,t}} h_{it}\right),$$

$$,, \quad = \int_{-X_{it}b-\overline{X}_i g}^{\infty} f(u_{it} \mid h_{it}) I\left(\frac{u_{it} - \frac{s_{ht}}{s^2_{h,t}} h_{it}}{\sqrt{s^2_t - s^2_{ht} s^{-2}_{h,t}}} > \frac{-X_{is}b - \overline{X}_i g - \frac{s_{hs}}{s^2_{h,t}} h_{it}}{\sqrt{s^2_s - s^2_{hs} s^{-2}_{h,t}}} \right) \partial u_{it} \, ,$$

$$,, \quad = \int_{\frac{-X_{it}b-\overline{X}_i g-\frac{s_{ht}}{s_{h,t}^{2}}h_{it}}{\sqrt{s_t^2-s_{ht}^2 s_{h,t}^{-2}}}}^{\infty} f\left(u_{it}^{*}\right) I\left(u_{it}^{*} > \frac{-X_{is}b-\overline{X}_i g-\frac{s_{hs}}{s_{h,t}^{2}}h_{it}}{\sqrt{s_s^2-s_{hs}^2 s_{h,t}^{-2}}} \right) \partial u_{it}^{*},$$

$$\text{With} \quad u_{it}^{*} = \frac{u_{it}-\frac{s_{ht}}{s_{h,t}^{2}}h_{it}}{\sqrt{s_t^2-s_{ht}^2 s_{h,t}^{-2}}},$$

$$,, \quad = 1-\Phi\left(\max\left(\frac{-X_{is}b-\overline{X}_i g-\frac{s_{hs}}{s_{h,t}^{2}}h_{it}}{\sqrt{s_s^2-s_{hs}^2 s_{h,t}^{-2}}}, \frac{-X_{it}b-\overline{X}_i g-\frac{s_{ht}}{s_{h,t}^{2}}h_{it}}{\sqrt{s_t^2-s_{ht}^2 s_{h,t}^{-2}}} \right)\right).$$

Hence,

$$g(h_{it}) = \frac{1}{s_{h,t}} f\left(\frac{h_{it}}{s_{h,t}}\right) \Phi\left(\min\left(\frac{X_{is}b+\overline{X}_i g+\frac{s_{hs}}{s_{h,t}^{2}}h_{it}}{\sqrt{s_s^2-s_{hs}^2 s_{h,t}^{-2}}}, \frac{X_{it}b+\overline{X}_i g+\frac{s_{ht}}{s_{h,t}^{2}}h_{it}}{\sqrt{s_t^2-s_{ht}^2 s_{h,t}^{-2}}} \right)\right),$$

With $s_{hs} = r_t s_s s_t - s_s^2$ and $s_{ht} = s_t^2 - r_t s_s s_t$.