A Two-Step First Difference Estimator for a Panel Data Tobit Model under Conditional Mean Independence Assumptions
Kalwij, A.S.

Publication date:
2004

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright, please contact us providing details, and we will remove access to the work immediately and investigate your claim.
A TWO-STEP FIRST DIFFERENCE ESTIMATOR FOR A PANEL DATA TOBIT MODEL UNDER CONDITIONAL MEAN INDEPENDENCE ASSUMPTIONS

By A.S. Kalwij

July 2004
A TWO-STEP FIRST-DIFFERENCE ESTIMATOR FOR A PANEL DATA TOBIT
MODEL UNDER CONDITIONAL MEAN INDEPENDENCE ASSUMPTIONS

Adriaan S. Kalwij*

Department of Economics, Tilburg University, Tilburg, The Netherlands

23 July 2004

ABSTRACT

This study develops a two-step estimator for a panel data Tobit model based on taking first-differences of
the equation of interest, under conditional mean independence assumptions. The necessary correction
terms are non-standard and a substantial part is therefore devoted to the formal derivation of these
correction terms. The main advantage of this estimator is that it yields estimates that are far less
sensitivity to misspecification of the conditional mean independence assumption than an estimation
procedure set up in levels. Monte Carlo simulations are provided in support of this.

Keywords: Two-Step Estimator, Panel data, Tobit model
JEL-classification: C12, C23, C24

*Address for correspondence: Room B933, Department of Economics, Tilburg University,
PO Box 90153, 5000 LE Tilburg, The Netherlands. (Email: kalwij@uvt.nl)
1. INTRODUCTION

The main contribution to the literature of this study is the derivation of a two-step panel data Tobit model based on taking first-differences of the equation of interest, under conditional mean independence assumptions (Wooldridge, 1995). The main advantage of this new estimator is that it yields estimates that are far less sensitivity to misspecification of the conditional mean independence assumption than an estimation procedure set up in levels, as proposed by Wooldridge (1995, section 4).

The literature on nonlinear panel data models, surveyed by Honoré (2002), suggests two approaches to estimating a censored regression model in which the individual specific effects are allowed to correlate with the explanatory variables. The first is a fixed effects approach (Honoré, 1992, Kyriazidou, 1997) for which consistency does not require any assumptions on the individual specific effects but does require an additional distributional assumption, for instance the stationarity assumption in Honoré (1992). The second is a random effects approach (Nijman and Verbeek, 1992, Wooldridge, 1995, Rochina-Barrachina, 1999, Kalwij, 2003) for which consistency does not require an additional distributional assumption but does require a correct specification of the correlation between the individual specific effects and the explanatory variables, the so-called conditional mean independence assumption (Wooldrigde, 1995). This latter approach is followed in this study. The main reason for preferring this random effects approach is that, unlike the fixed effects approach, it yields a fully specified model that can be used for calculating marginal effects. Consistency of the estimates is based on the assumption of correctly specified correlation between the individual specific effects and the explanatory variables and, as a consequence, parameter estimates turn out to be sensitivity to the chosen parameterization. For this reason one may wish to use a very flexible parameterization (Zabel, 1992, Wooldridge, 1995). This can, however, increase the number of parameters dramatically. An alternative solution proposed in Kalwij (2003) is to start by eliminating the individual specific effects from the equation of interest by taking first-differences. He develops a Maximum Likelihood estimator based on this idea. The disadvantage of his estimator is that it is difficult to use in practice for applied researchers since it demands programming of
non-standard likelihood contributions. For this reason this study builds on the work of Kalwij (2003) and constructs a relatively easy-to-use two-step estimator for a panel data Tobit model based on taking first-differences of the equation of interest. Another reason for preferring a two-step procedure is to reduce the sensitivity of the parameter estimates with respect to the distribution assumptions. The procedure for obtaining the necessary correction terms is conceptually similar to the derivation of the inverse Mill’s ratio in case of a standard Tobit model in levels. The correction terms for a Tobit model based on taking first-differences are, however, non-trivial to derive and for this purpose a large part of this study is devoted to this derivation. Section 2 derives the Moment Generating Function of a random variable that is defined as the difference of two censored normal random variables and derives the expectation of this random variable. Based on this result, section 3 constructs the correction terms and formulates the two-step first-difference estimator for a panel data Tobit model under conditional mean independence assumptions. Section 4 carries out simulations to examine how this estimator behaves, relatively to the two-step Tobit estimator of Wooldridge (1995), when the conditional mean independence assumption does not hold. Section 5 concludes.

2. THE FIRST MOMENT OF THE DIFFERENCE OF TWO CENSORED NORMAL RANDOM VARIABLES

$X$ and $Y$ are two jointly distributed normal random variables:

$$
\begin{bmatrix}
X \\
Y
\end{bmatrix} \sim N\left(\begin{bmatrix}
\mu_X \\
\mu_Y
\end{bmatrix}, \begin{bmatrix}
\sigma_X^2 & \rho \sigma_X \sigma_Y \\
\rho \sigma_X \sigma_Y & \sigma_Y^2
\end{bmatrix}\right)
$$

The density function of $(X,Y)$ is the bivariate normal density function and is denoted by $f_{XY}(x,y)$. In the remainder of this study $\phi(.)$ denotes the standard normal density function, $\Phi(.)$ the standard normal
distribution function and \( \Phi_2(...) ; \rho \) the bivariate standard normal distribution function with a correlation coefficient \( \rho \) (\( \equiv \rho_{xy} \), subscript \( XY \) is dropped).

\( X \) is censored for values below \( a \) and \( Y \) is censored for values below \( b \). The density function of the random variable \( Z = X - Y \) conditional on \( X > a \) and \( Y > b \) is given by (see Kalwij, 2003):

\[
f_{Z|X>a,Y>b}(z \mid X > a, Y > b) = \frac{1}{\sigma_Z} \phi\left( \frac{z - \mu_Z}{\sigma_Z} \right) \Phi\left( -\frac{\max\{a, b + z\} - \mu_{X|Z}}{\sigma_{X|Z}} \right),
\]

where \( \mu_Z = \mu_X - \mu_Y \), \( \sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 - 2 \rho \sigma_X \sigma_Y \), and \( \rho_{xz} = \frac{\sigma_x - \rho \sigma_y}{\sigma_z} \).

The Moment Generating Function of the random variable \( Z = X - Y \) conditional on \( X > a \) and \( Y > b \) is given by (see Appendix A):

\[
\int_{-\infty}^{+\infty} e^{\mu_Z + \frac{t^2}{2}} \phi(s) \Phi\left( -\frac{\max\{a, b + \mu_z + \sigma_z(s + t \sigma_z)\} - \mu_X - \rho_{xz}(s + t \sigma_z)}{\sqrt{1 - \rho_{xz}^2}} \right) ds
\]

The expectation of \( Z = X - Y \) conditional on \( X > a \) and \( Y > b \) is obtained by taking the derivative of the Moment Generating Function, i.e. equation (3), with respect to \( t \) and evaluate this function in \( t \) equals 0. This yields (see Appendix B):

\[
\mu_{Z|X>a,Y>b} = \Phi_2\left( \frac{-a - \mu_X}{\sigma_X}, \frac{-b - \mu_Y}{\sigma_Y}; \rho \right)
\]

\[
	imes \int_{-\infty}^{a-b-\mu_x} \phi(s) \Phi\left( -\frac{a - \mu_X}{\sigma_X} - \rho_{xz}s \right) ds
\]
\[ + \mu_z \int_{a-b-\mu_x \sqrt{1-\rho^2}}^{+\infty} \phi(s) \phi \left( \frac{x}{\sqrt{1-\rho^2}} \right) \Phi \left( \frac{b-\mu_y}{\sigma_y} \right) \Phi \left( \frac{a-\mu_x}{\sigma_x} \right) \Phi \left( \frac{b-\mu_y}{\sigma_y} \right) \]

3. A TWO-STEP FIRST-DIFFERENCE ESTIMATOR FOR
A PANEL DATA TOBIT MODEL

The model of interest is formulated as follows, assuming \( T \) time periods and \( N \) individuals:

\[ y_{it}^* = x_{it} \beta + \alpha_i + \epsilon_{it}, \]

\[ y_{it} = \max(0, y_{it}^*) \quad t = \{1, \ldots, T\}, i = \{1, \ldots, N\}. \]

This is a panel data Tobit model and is a straightforward extension of the standard Tobit model (Tobin, 1958) by including an additional dimension (time or individuals). The individual is indexed by \( i \) and the time period by \( t \). \( x_{it} \) is a \( (1 \times K) \) vector of exogenous variables, \( \beta \) is a \( (K \times 1) \) vector of the parameters of interest and \( \alpha_i \) is an unobserved individual specific effect that may correlate with \( x_{it} \). The latent dependent variable \( y_{it}^* \) is censored at zero and only \( y_{it} \) is observed. The error term \( \epsilon_{it} \) is assumed to be a normal
random variable with mean zero and variance $\sigma^2_t, \varepsilon_{it} \sim N\left(0, \sigma^2_t\right)$, and is allowed to be serially correlated.

The panel data are characterized by having a large number of individuals over a short period of time.

The two-step estimator proposed by Wooldridge (1995) is as follows. Based on estimates of a probit model explaining whether or not the dependent variable is positive, given $x_{it}$, one constructs the familiar inverse Mills ratios for each time period. Next, one estimates equation (5) on a sub-sample of observations for which the dependent variable is positive with the period-specific inverse Mills ratios as additional explanatory variables. A conditional mean independence assumption is made to deal with the possible correlations between the unobserved individual specific effects and the explanatory variables. This procedure allows for arbitrary serial correlation.

The main idea behind the Maximum Likelihood estimator of Kalwij (2003) is that one starts by eliminating the unobserved individual specific effects from equation (5) in order to make the resulting estimates less sensitivity to a conditional mean independence assumption. First-differencing equation (5) yields the following model:

\begin{equation}
\Delta y^*_t = \Delta x_{it} \beta + \eta_{it}
\end{equation}

\begin{equation}
\Delta y_{it} = \begin{cases} 
\Delta y_{it}^* & \text{if } y_{it}^* > 0 \text{ and } y_{it}^* > 0 \\
0 & \text{if } y_{it}^* \leq 0 \text{ and } y_{it}^* \leq 0 \\
y_{it}^* & \text{if } y_{it}^* \leq 0 \text{ and } y_{it}^* > 0 \\
- y_{it}^* & \text{if } y_{it}^* > 0 \text{ and } y_{it}^* \leq 0 
\end{cases}
\end{equation}

Where $\Delta y^*_t = y^*_t - y^*_i$, $\Delta x_{it} = x_{it} - x_i$, and $\eta_{it} = \varepsilon_{it} - \varepsilon_{is}$. The correlation coefficient of $\varepsilon_{is}$ and $\varepsilon_{it}$ is denoted by $\rho_t$. Given the distributional assumptions $\eta_{it} \sim N\left(0, \sigma^2_{\eta_t}\right)$ with $\sigma^2_{\eta_t} = \sigma^2_\eta - 2 \rho_t \sigma_{\varepsilon} \sigma_t + \sigma^2_t$, one can obtain Maximum Likelihood estimates of the parameters of interest using the density function as formulated in equation (2) (see Kalwij, 2003). An alternative is to construct a two-step estimator in the spirit of Rochina-Barrachina (1999) who develops a two-step estimator for panel data selection models. For the case of a first-difference panel data Tobit model (equations (6) and (7)) a two-step procedure is developed underneath, using the fundamental statistical results of Section 2.
Considering the model given by equation (6) and the nonzero observations on the dependent variable in both periods, the expectation of $\Delta y_{it}$ conditional on $y_{is} > 0$ and $y_{it} > 0$ is given by:

$$E[\Delta y_{it} | y_{is} > 0, y_{it} > 0] = \Delta x_{it} \beta + E[\eta_{it} | \epsilon_{is} > -x_{is} \beta - \alpha_i, \epsilon_{it} > -x_{it} \beta - \alpha_i].$$

Using equation (4), the expectation at the RHS of equation (8) is written as follows:

$$E[\eta_{it} | \epsilon_{is} > -x_{is} \beta - \alpha_i, \epsilon_{it} > -x_{it} \beta - \alpha_i] = \pi_s \Lambda_s(M_{is}, M_{it}, \rho_i) - \pi_s \Lambda_s(M_{is}, M_{it}, \rho_i).$$

With

$$\pi_s = (\sigma_s - \rho, \sigma_s),$$

$$\pi_t = (\sigma_t - \rho, \sigma_s),$$

$$M_{is} = (x_{is} \beta + \alpha_i) / \sigma_s,$$

$$M_{it} = (x_{it} \beta + \alpha_i) / \sigma_t,$$

$$\Lambda_s(M_{is}, M_{it}, \rho) = \frac{\phi(M_{is}) \Phi(M_{it} - \rho M_{it}) \sqrt{1 - \rho_t^2}}{\Phi_2(M_{is}, M_{it}; \rho)},$$

$$\Lambda_t(M_{is}, M_{it}, \rho) = \frac{\phi(M_{it}) \Phi(M_{is} - \rho M_{is}) \sqrt{1 - \rho_t^2}}{\Phi_2(M_{is}, M_{it}; \rho)}.$$

The correction terms are now added and subtracted from equation (6):

$$\Delta y_{it} = \Delta x_{it} \beta + \pi_s \Lambda_s(M_{is}, M_{it}, \rho_i) - \pi_s \Lambda_s(M_{is}, M_{it}, \rho_i) + \xi_{it}$$

if $y_{is} > 0$ and $y_{it} > 0$, with

$$\xi_{it} = -\pi_s \Lambda_s(M_{is}, M_{it}, \rho_i) + \pi_s \Lambda_s(M_{is}, M_{it}, \rho_i) + \eta_{it}.$$

$\xi_{it}$ is an error term with expectation zero (conditional on $y_{is} > 0$ and $y_{it} > 0$). If the correction terms are known then equation (16) can be estimated using least squares on the sample of individuals for which the dependent variable is positive in both periods.

3.1 A two-step estimation procedure
A two-step estimation procedure for estimating equation (16) is constructed as follows. In the first step we estimate the following bivariate probit model for every time-pair, say two subsequent periods $s$ and $t$:

\begin{align*}
(18) \quad & y_{is}^* = x_{is} \beta + \alpha_i + \varepsilon_{is}, \\
(19) \quad & y_{it}^* = x_{it} \beta + \alpha_i + \varepsilon_{it},
\end{align*}

\begin{align*}
(20) \quad & I_{is} = \begin{cases} 
0 & \text{if } y_{is}^* \leq 0, \\
1 & \text{if } y_{is}^* > 0,
\end{cases} \\
(21) \quad & I_{it} = \begin{cases} 
0 & \text{if } y_{it}^* \leq 0, \\
1 & \text{if } y_{it}^* > 0.
\end{cases}
\end{align*}

We use the conditional mean independence assumptions as discussed in Wooldridge (1995) to deal with correlations between the unobserved individual specific effects and the explanatory variables. This approach essentially models the correlation between the unobserved individual specific effect and the explanatory variables as a pre-specified function of the explanatory variables and a random individual specific error term, for instance:

\begin{align*}
(22) \quad & \alpha_i = h(x_i) \gamma + \mu_i, \\
\end{align*}

with $\mu_i \mid x_i \sim N(0, \sigma^2_{\mu})$ and $x_i \equiv (x_{i1}, \ldots, x_{iT})$. The parameter $\gamma$ is an additional parameter of the model to be estimated. A popular choice for $h(x_i)$ in empirical research is to take the average over time.

Substituting the parameterisation of the individual specific effects in equations (18) and (19) yields:

\begin{align*}
(23) \quad & y_{is}^* = x_{is} \beta + h(x_i) \gamma + u_{is}, \\
(24) \quad & y_{it}^* = x_{it} \beta + h(x_i) \gamma + u_{it},
\end{align*}

with $u_{is} = \mu_i + \varepsilon_{is}$ and $u_{it} = \mu_i + \varepsilon_{it}$, and variances $\sigma^2_{u,s} = \sigma^2_{\mu} + \sigma^2_{\varepsilon,s}$ and $\sigma^2_{u,t} = \sigma^2_{\mu} + \sigma^2_{\varepsilon,t}$, respectively. The correlation between $u_{is}$ and $u_{it}$ is denoted by $\rho_{st}$. Using Maximum Likelihood for
estimating the bivariate probit model yields estimates for 
\[ \beta_s = \frac{\beta}{\sigma_{u,s}}, \gamma_s = \frac{\gamma}{\sigma_{u,s}}, \beta_t = \frac{\beta}{\sigma_{u,t}}, \gamma_t = \frac{\gamma}{\sigma_{u,t}} \]
and \( \rho_{st} \).

In the second step we substitute \( \hat{M}_{is} = x_{it}\hat{\beta}_s + h(x_i)\hat{\gamma}_s \), \( \hat{M}_{it} = x_{it}\hat{\beta}_t + h(x_i)\hat{\gamma}_t \) and \( \hat{\rho}_{st} \) in equation (16):
\[
(25) \quad \Delta y_{it} = \Delta x_{it}\beta + \pi^*_{s,t}\Lambda_t\left(\hat{M}_{is}, \hat{M}_{it}, \hat{\rho}_{st}\right) - \pi^*_{s,t}\Lambda_t\left(\hat{M}_{is}, \hat{M}_{it}, \hat{\rho}_{st}\right) + \xi_{it}, \quad \text{if } y_{it} > 0 \text{ and } y_{it} > 0.
\]
This equation is estimated using Least Squares on a sample of individuals for which the dependent variable is positive in both periods (\( y_{it} > 0 \) and \( y_{it} > 0 \)). The auxiliary parameters \( \pi^*_{s,t} \) and \( \pi^*_{t,s} \) are defined as \( \pi^*_{s,t} = (\sigma_{u,s} - \rho_{st}\sigma_{u,t}) \) and \( \pi^*_{t,s} = (\sigma_{u,t} - \rho_{st}\sigma_{u,s}) \).

### 4. A MONTE CARLO STUDY

The main idea behind constructing a new two-step estimator for a panel data Tobit model based on taking first-differences of the equation of interest is to reduce the bias that may arise if the conditional mean independence assumption does not hold, i.e. a misspecification of the individual specific effects (equation (22)). To examine this a Monte Carlo study is carried out to compare the two-step first-difference Tobit estimator of Section 3 with the 'standard' two-step panel data Tobit estimator as proposed by Wooldridge (1995), both under conditional mean independence assumptions.

The design of the Monte Carlo simulations is as follows. The data is generated by:
\[
y_{it} = \max(0, 0.2 + x_{it} + \alpha_i + \epsilon_{it}), \quad x_{it} = 0.8x_{it-1} + \xi_{it}, \quad \epsilon_{it} = 0.4\epsilon_{it-1} + \zeta_{it}, \quad x_{i1} = \xi_{i1}, \quad \epsilon_{i1} = \zeta_{i1} \text{ and the}\]
three error terms \( \epsilon_{it}, \xi_{it} \) and \( \zeta_{it} \) are \( N(0,1) \) distributed. So the true value of \( \beta \) is 1. The individual specific effects are generated using a nonlinear function of the time-averages of the explanatory variables:
\[ \alpha_i = \bar{x}_i | i + \mu_i, \quad \mu_i \sim N(0,1). \]
The simulations are based on 100 replications and the values chosen for
and \( T \) are, respectively, \( \{500, 1000\} \) and \( \{2, 4, 8, 12\} \). These are plausible values in empirical research using panel data. We have experimented with several different designs and the main conclusion of this study remained unchanged. Nevertheless, the results below have to be interpreted with caution since they may depend on the particular design chosen.

We estimate \( \beta \) using three different models and using the two-step estimators outlined in Section 3 and in Wooldridge (1995). The three models differ with respect to the conditional mean independence assumption chosen, i.e. the choice of \( h(.) \) in equation (22). The first model is a correctly specified model using \( h(x_i) = x_i | \bar{x}_i | \). The second model is a random effects panel data Tobit model assuming no correlation between the individual specific effect and the explanatory variable: \( h(x_i) = 0 \). The third model uses an in empirical research often made conditional mean independence assumption that the correlation between the unobserved individual specific effect and the explanatory variable is a linear function of the averages over time of the explanatory variable: \( h(x_i) = \bar{x}_i \). Hence, the second and third models are misspecified. When estimated in levels the second model is commonly estimated when using a standard statistical package such as STATA (xttobit command).

The simulation results are reported in table 1. The simulation results of model 1 show that both estimators yield, of course, unbiased estimates of the parameter of interest (\( \beta=1 \)). The root mean squared error (RMSE) shows that the first-difference Tobit estimator is often less efficient but the differences with Wooldridge’s Tobit estimator are relatively small. Nevertheless, using the first-difference Tobit estimator instead of Wooldridge’s Tobit estimator is likely to result in a loss of efficiency. The simulation results when using models 2 and 3 provide some measure of the relative performance of the two estimators when the conditional mean independence assumption, as specified in equation (22), does not hold. When estimating model 2, both estimators yield biased estimates but the magnitude of the biases of the two estimators differ considerably. The estimator based on first-differences yields an estimate that is far less sensitive to misspecification of the individual specific effects, relatively to Wooldridge’s Tobit estimator. Finally we estimate the model with a commonly chosen conditional mean independence assumption, i.e.
model 3. Here we see substantial bias reductions for both estimators but the relative bias when using Wooldridge’s estimator remains high.

We experimented with alternative simulation designs but they all yielded the same conclusion: using the two-step first-difference Tobit estimator substantially reduces the sensitivity of the estimates of the parameters of interest with respect to the conditional mean independence assumption.

5. CONCLUSIONS

This study developed a new two-step first-difference estimator for a panel data Tobit model under conditional mean independence assumptions. In particular, since the necessary correction terms are non-standard, a substantial part of this study has been devoted to the formal derivation of these correction terms. This new estimator yields estimates that are far less sensitivity to a misspecification of the conditional mean independence assumption than an estimation procedure set up in levels. The results of the Monte Carlo study in Section 4 support this.

We conclude that the two-step first-difference Tobit estimator proposed in this study can be regarded as a very attractive alternative to the fixed effects Tobit estimator of Honoré (1992), since it allows one to calculate marginal effects, and as a bias reduction technique when applied to a random effects panel data Tobit model under conditional mean independence assumptions (Wooldridge, 1995). We successfully applied this new two-step first-difference Tobit estimator in Kalwij and Gregory (2005) to an analysis of paid overtime work in Britain, hereby comparing the empirical results in detail with alternative approaches commonly used in the literature.
ACKNOWLEDGEMENTS

I would like to thank Mary Gregory, Maria Rochina-Barrachina, Jeff Wooldridge and seminar participants at the 9th International Conference on Panel Data for very helpful comments.

REFERENCES

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


APPENDIX A: THE MOMENT GENERATING FUNCTION

The Moment Generating Function of the random variable $Z$ given $X>a$ and $Y>b$ is defined as:

(A1) \[ m(t) = \int_{a}^{\infty} \left[ \int_{b}^{\infty} e^{\mathbf{t} \cdot \mathbf{z}} f_{XZ}(x, z) dx \right] dz = \int_{\min\{a,b+z\}}^{\infty} \int_{a}^{\infty} e^{\mathbf{t} \cdot \mathbf{z}} f_{XY}(x, y) dy dx \]

(A2) \[ m(t) = \int_{a}^{\infty} \left[ \int_{b}^{\infty} e^{\mathbf{t} \cdot \mathbf{z}} f_{XZ}(x, z) dz \right] dx = \int_{a}^{\infty} \int_{b}^{\infty} e^{\mathbf{t} \cdot \mathbf{z}} f_{XY}(x, y) dy dx \]

The denominator of equation (A2) can be written as:

\[ \int_{a}^{\infty} \int_{b}^{\infty} f_{XY}(x, y) dy dx = \Phi_{2}\left( \frac{-(a - \mu_{X})}{\sigma_{X}}, \frac{-(b - \mu_{Y})}{\sigma_{Y}}, \rho \right) \]

The numerator of equation (A2) can be written as:

(A3) \[ \int_{-\infty}^{\infty} \left[ \int_{\min\{a,b+z\}}^{\infty} e^{\mathbf{t} \cdot \mathbf{z}} f_{XZ}(x, z) dx \right] dz \]

To simplify the integral we substitute

(A4) \[ u = \frac{x - \mu_{X}}{\sigma_X} \quad \text{and} \quad v = \frac{z - \mu_{Z}}{\sigma_Z} \]

Furthermore, to adjust the borders of integration we define:

(A5) \[ c_{1}(v) = \frac{a - \mu_{X}}{\sigma_X} \quad \text{and} \quad c_{2}(v) = \frac{b + (\mu_{Z} + \sigma_{Z}v) - \mu_{X}}{\sigma_X} \]

Now (A3) becomes:

(A6) \[ \int_{-\infty}^{\infty} \left[ \int_{\min\{c_{1}(v),c_{2}(v)\}}^{\infty} e^{i(\mu_{Z} + \sigma_{Z}v)} \phi_{2}(u, v; \rho_{XZ}) du \right] dv , \]
where $\phi_2(\ldots)$ denoted the bivariate standard normal distribution. Equation (A6) is rewritten as follows:

\[
(A7) \quad e^{iu_z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi \sqrt{1 - \rho_{xz}^2}} \exp \left\{ -\frac{1}{2(1 - \rho_{xz}^2)} \left[ u^2 - 2\rho_{xz}uv + v^2 - 2(1 - \rho_{xz}^2)\sigma_z v \right] \right\} du \, dv
\]

To simplify this integral we substitute:

\[
(A8) \quad w = \frac{u - \rho_{xz}v}{\sqrt{1 - \rho_{xz}^2}}, \quad \text{and} \quad s = v - t\sigma_z.
\]

The new borders of integration become:

\[
(A9) \quad c_3(s,t) = \frac{a - \mu_x}{\sigma_x} - \rho_{xz} (s + t\sigma_z) \quad \text{and} \quad c_4(s,t) = \frac{b + (\mu_z + \sigma_z (s + t\sigma_z) ) - \mu_x}{\sigma_x} - \rho_{xz} (s + t\sigma_z)
\]

Equation (A6) now becomes:

\[
(A10) \quad e^{iu_z + \frac{4\sigma_z^2}{2}} \int_{-\infty}^{\infty} \int_{\max\{c_3(s,t),c_4(s,t)\}} \phi(s) \phi(w) dw \, ds
\]

with

\[
(A11) \quad \max\{c_3(s,t),c_4(s,t)\} = \begin{cases} 
\frac{c_3(s,t)}{\sigma_z} & \text{if } s < \frac{a - b - \mu_z - t\sigma_z^2}{\sigma_z} \\
\frac{c_4(s,t)}{\sigma_z} & \text{if } s \geq \frac{a - b - \mu_z - t\sigma_z^2}{\sigma_z}
\end{cases}
\]
APPENDIX B

$m_l(t)$ denotes the numerator of the Moment Generating Function of equation (3), i.e. equation (A10):

\[
(B1) \quad m_l(t) = e^{s \sqrt{\mu_x + \frac{1}{2} \sigma_z^2}} \int_{-\infty}^{\infty} \phi(s) \left[ \int_{\max\{c_3(s,t),c_4(s,t)\}}^{\infty} \phi(w) dw \right] ds.
\]

The borders $c_3(s,t)$ and $c_4(s,t)$ are defined as in Appendix A (equation (A9)). First we need the derivative of $m_l(t)$ with respect to $t$:

\[
(B2) \quad \frac{\partial m_l(t)}{\partial t} = \left( \mu_x + \sigma_z^2 \right) e^{s \sqrt{\mu_x + \frac{1}{2} \sigma_z^2}} \int_{-\infty}^{\infty} \left[ \int_{\max\{c_3(s,t),c_4(s,t)\}}^{\infty} \phi_2(w,s;0) dw \right] ds \\
+ e^{s \sqrt{\mu_x + \frac{1}{2} \sigma_z^2}} \int_{-\infty}^{\infty} \left[ \frac{\partial \max\{c_3(s,t),c_4(s,t)\}}{\partial t} \phi_2 \left( \max\{c_3(s,t),c_4(s,t)\},s;0 \right) \right] ds.
\]

\[
(B3) \quad \frac{\partial m_l(t)}{\partial t} = \left( \mu_x + \sigma_z^2 \right) e^{s \sqrt{\mu_x + \frac{1}{2} \sigma_z^2}} \int_{-\infty}^{\infty} \left[ \int_{\max\{c_3(s,t),c_4(s,t)\}}^{\infty} \frac{1}{2\pi} \exp \left\{ -\frac{1}{2} \left[ w^2 + s^2 \right] \right\} dw \right] ds \\
+ e^{s \sqrt{\mu_x + \frac{1}{2} \sigma_z^2}} \int_{c(t)}^{\infty} \left[ \frac{\partial c_3(s,t)}{\partial t} \frac{1}{2\pi} \exp \left\{ -\frac{1}{2} \left[ (c_3(s,t))^2 + s^2 \right] \right\} ds \\
+ e^{s \sqrt{\mu_x + \frac{1}{2} \sigma_z^2}} \int_{c(t)}^{\infty} \left[ \frac{\partial c_4(s,t)}{\partial t} \frac{1}{2\pi} \exp \left\{ -\frac{1}{2} \left[ (c_4(s,t))^2 + s^2 \right] \right\} ds \right]
\]

with

\[
(B4) \quad \frac{\partial c_3(s,t)}{\partial t} = -\rho_{xz} \sigma_z \frac{\rho_{xZ}}{\sqrt{1 - \rho_{xz}^2}} \quad \text{and} \quad \frac{\partial c_4(s,t)}{\partial t} = \frac{\rho_{xZ} \sigma_z}{\sqrt{1 - \rho_{xz}^2}}.
\]
(B5) \[ (c_3(s,t))^2 + s^2 \bigg|_{t=0} = \left( s - \frac{a - \mu_X}{\sigma_X} \rho_{xz} \right)^2 + \left( \frac{a - \mu_X}{\sigma_X} \right)^2, \]

and

(B6) \[ (c_4(s,t))^2 + s^2 \bigg|_{t=0} = \left( s - \frac{b - \mu_Z \times \rho_{xz} \sigma_X - \sigma_Z}{\sigma_y} \right)^2 + \left( \frac{b - \mu_Y}{\sigma_y} \right)^2. \]

Next we evaluate equation (B3) in \( t=0 \):

(B7) \[ \frac{\partial m_1(t)}{\partial t} \bigg|_{t=0} = (\mu_Z) \int_{c(0)}^{c(0)} \phi(s) \Phi(-c_3(s,0))ds + (\mu_Z) \int_{c(0)}^{c(0)} \phi(s) \Phi(-c_4(s,0))ds \]

\[ + \rho_{xz} \sigma_Z \Phi \left( \frac{a - \mu_X}{\sigma_X} \right) \left( \frac{c(0) - \frac{a - \mu_X}{\sigma_X} \rho_{xy}}{\sqrt{1 - \rho_{xz}^2}} \right) \]

\[ + \frac{\sigma_X}{\sigma_Y} \left( \rho_{xz} \sigma_Z - \frac{\sigma_Z^2}{\sigma_X} \right) \Phi \left( \frac{b - \mu_Y}{\sigma_Y} \right) \left( - \frac{c(0) - \frac{b + \mu_Z - \mu_X \times \rho_{xz} \sigma_X - \sigma_Z}{\sigma_Y} \sigma_X}{\sqrt{1 - \rho_{xz}^2}} \right), \]

with\( c(t) = \frac{a-b-\mu_X - t \sigma_Z^2}{\sigma_Z} \). Substituting \( c(0) = \frac{a-b-\mu_Z}{\sigma_Z} \) in equation (B7) yields:

(B8) \[ \frac{\partial m_1(t)}{\partial t} \bigg|_{t=0} = (\mu_Z) \int_{c(0)}^{c(0)} \phi(s) \Phi(-c_3(s,0))ds + (\mu_Z) \int_{c(0)}^{c(0)} \phi(s) \Phi(-c_4(s,0))ds \]

\[ + \left( \sigma_X - \rho \sigma_Y \right) \Phi \left( \frac{a - \mu_X}{\sigma_X} \right) \left( \frac{\rho_X \sigma_X}{\sqrt{1 - \rho_{xz}^2}} \right) \Phi \left( \frac{b - \mu_Y}{\sigma_Y} \right) \left( \frac{\sigma_Y}{\sqrt{1 - \rho_{xz}^2}} \right), \]
\[ + (\rho \sigma_x - \sigma_y) \left( \frac{b - \mu_y}{\sigma_y} \right) \Phi \left( \frac{\rho \left( \frac{b - \mu_y}{\sigma_y} \right) - \left( \frac{a - \mu_x}{\sigma_x} \right)}{\sqrt{1 - \rho^2}} \right). \]

Combining equation (B8) with the denominator of equation (3), which does not depend on \( t \), yields equation (4), i.e. the derivative of equation (3) (the Moment Generating Function) with respect to \( t \) and evaluated in \( t=0 \).
Table 1: Simulation results. Three model specifications are estimated with the two-step estimator of Wooldridge (1995) and the two-step first-difference Tobit estimator of Section 3. The specifications differ with respect to the conditional mean independence assumption (equation (22)). RMSE is the root mean squared error and MAD is the median absolute deviation. The true value of $\beta$ is 1.

Model 1: correct specification \( h(x_i) = \bar{x}_i \mid \tilde{x}_i \mid \)

<table>
<thead>
<tr>
<th></th>
<th>Two-step Tobit (Wooldridge, 1995)</th>
<th>Two-step first-difference Tobit (Section 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean bias RMSE Median bias MAD</td>
<td>Mean bias RMSE Median bias MAD</td>
</tr>
<tr>
<td>N T</td>
<td>500 2 -0.001 0.111 0.008 0.078</td>
<td>500 2 0.019 0.132 0.002 0.093</td>
</tr>
<tr>
<td>500 4 -0.001 0.058 -0.011 0.035</td>
<td>500 4 -0.007 0.078 -0.010 0.052</td>
<td></td>
</tr>
<tr>
<td>500 8 -0.007 0.051 -0.001 0.032</td>
<td>500 8 -0.006 0.051 -0.010 0.036</td>
<td></td>
</tr>
<tr>
<td>500 12 -0.006 0.046 -0.005 0.030</td>
<td>500 12 -0.003 0.038 -0.003 0.027</td>
<td></td>
</tr>
</tbody>
</table>

Model 2, incorrect specification: \( h(x_i) = 0 \), a standard random effects panel data model

<table>
<thead>
<tr>
<th></th>
<th>Two-step Tobit (Wooldridge, 1995)</th>
<th>Two-step first-difference Tobit (Section 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean bias RMSE Median bias MAD</td>
<td>Mean bias RMSE Median bias MAD</td>
</tr>
<tr>
<td>N T</td>
<td>500 2 2.264 0.309 2.247 2.247</td>
<td>500 2 0.077 0.167 0.065 0.118</td>
</tr>
<tr>
<td>500 4 1.907 0.286 1.929 1.929</td>
<td>500 4 0.038 0.094 0.037 0.065</td>
<td></td>
</tr>
<tr>
<td>500 8 1.512 0.278 1.468 1.468</td>
<td>500 8 0.024 0.064 0.016 0.039</td>
<td></td>
</tr>
<tr>
<td>500 12 1.183 0.190 1.136 1.136</td>
<td>500 12 0.023 0.046 0.021 0.034</td>
<td></td>
</tr>
</tbody>
</table>

Model 3, incorrect specification: \( h(x_i) = \bar{x}_i \), an often chosen empirical specification

<table>
<thead>
<tr>
<th></th>
<th>Two-step Tobit (Wooldridge, 1995)</th>
<th>Two-step first-difference Tobit (Section 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean bias RMSE Median bias MAD</td>
<td>Mean bias RMSE Median bias MAD</td>
</tr>
<tr>
<td>N T</td>
<td>500 2 0.709 0.150 0.687 0.687</td>
<td>500 2 0.023 0.133 0.004 0.092</td>
</tr>
<tr>
<td>500 4 0.591 0.117 0.577 0.577</td>
<td>500 4 -0.003 0.079 -0.007 0.049</td>
<td></td>
</tr>
<tr>
<td>500 8 0.510 0.112 0.492 0.492</td>
<td>500 8 -0.005 0.051 -0.007 0.037</td>
<td></td>
</tr>
<tr>
<td>500 12 0.440 0.089 0.429 0.429</td>
<td>500 12 0.000 0.038 0.000 0.025</td>
<td></td>
</tr>
</tbody>
</table>

18