MODIFIED MAXIMUM LIKELIHOOD ESTIMATION OF TOBIT MODELS WITH FIXED EFFECTS: THEORY AND AN APPLICATION TO EARNINGS EQUATIONS

GABRIEL JIMÉNEZ ZAMBRANO
Banco de España

This paper starts from the orthogonalization method proposed by Cox and Reid which is applied to the Tobit model panel for data with fixed effects. Neyman and Scott showed that, generally, the maximum likelihood estimator is inconsistent (the incidental parameter problem*). The methodology explained here recovers the use of the log-likelihood function to solve this problem taking advantage of the time-series dimension of panel data. For the Tobit model we show when is it possible to recover the orthogonal parameters, and study the characteristics of the estimators obtained with simulation methods. Also, an illustration for earnings equations has been performed.*

Keywords: Orthogonal parameters, modified profile likelihood, panel data, tobit model.

(*JEL* C15, C24, C23)

1. Introduction

One of the main applications of the use of panel data is in controling for unobserved time-invariant heterogeneity. Unfortunately, there are few models for which this is easy to deal with. In particular, Neyman and Scott (1948) showed that estimating the parameters of interest and the 'fixed effects' by maximum likelihood is not a good solution, because in general such estimators do not have good asymptotic properties[1] . The

I wish to thank M. Arellano for his interest in this study and his help, an also to the anonymous referee whose ideas have improved the paper. I am also grateful to A. Encina and J. M. Carro for their generous technical support. And particularly, I cannot forget my family for always being there: this paper is more yours than mine. I am indebted specially to *our* man in Holland, who always has some words or a bike ready, and to P. Ysanda. My heartfelt thanks to them all. E-mail: gabriel.jimenez@bde.es.

[1] When the number of individuals (N) increases and the number of observations per individual (T) stays fixed.

reason is that the number of parameters to be estimated approaches infinity (this is the so-called *incidental parameters problem*). In this sense, Liang (1987) proved that this class of estimators is in general inconsistent of order $O(1/T)$.

The literature on panel data has until now faced this problem solving the models *ad hoc* or using particular solutions for a limited class of models. Especially, there are two widely used techniques: 'Differencing Out' and 'Suficient Statistic'. More recently Cox and Reid (1987) have explored orthogonalization as a way to obtain their conditional profile method. Then, Lancaster (1997) took advantage of this ortogonalization from a Bayesian point of view, applying it to different panel data models with fixed effects. Arellano (2003a) has also used orthogonalization to deal with panel data binary choice models with individual effects, suggesting a different methodology from that of Lancaster. In this context, I use the method exposed in Arellano (2003a) to deal with panel data a Tobit model which also includes fixed effects. This is a non-linear model that has not been analysed before from this point of view. In the paper, I show when is it possible to derive an explicit functional form for the orthogonal effect and I obtain it (which is unusual for this kind of problem). I also show that the results derived for the Tobit model with fixed effects encompass those for the linear model already found in the literature. Finally, I use a Monte Carlo study to show the advantages of this approach, in terms of order of the bias in T, vis-à-vis, other methods.

The first method for panel data described previously, namely Differencing Out, allows us to derive a consistent estimator for the relevant parameters as N tends to infinity, whereas the other methods have biases in T. In particular, the first method consists of differencing out the fixed effect where possible. For instance, this is the case of linear models (see Arellano, 2001; and Arrelano, 2003b) and the Weibull and exponential hazard model (both with exogenous regressors).

The Sufficient Statistic method is based on the conditional maximum likelihood approach. This approach can be used, when applicable, to construct consistent estimators of the parameters of interest. It is based on finding a sufficient statistic of the fixed effects and then using it to construct the conditional distribution. In this way, it is possible to achieve a distribution that does not depend on the nuisance parameters, so that we could in principle estimate the rest of them, maximizing this conditional likelihood. In practice, unfortunately, this is a very difficult method to implement. In fact, it seems to be limited to the logit, Poisson, Weibull and linear models with known variance; and so we cannot apply it systematically.

Due to the small class of models to which these procedures can be applied, there was a widespread abandonment by econometricians of likelihood-based methods until the publication in 1987 of the important piece of work of Cox and Reid. These authors developed the conditional profile likelihood method, based on the idea of the sufficient statistic exposed above. This procedure uses as a starting-point an orthogonalization as an approximation to the conditional likelihood method. Lancaster (1997) also applies an orthogonalization, although focusing on panel data. He orthogonalizes the fixed effects and then presents an *integrated likelihood estimator*, where a prior distribution is needed to integrate out the orthogonalized fixed effects. Lancaster solves some models explicitly, in particular he obtains an orthogonal reparametrization of the nuisance parameter for the linear case. The model I propose embeds the former as a special case.

One of the problems of orthogonalization is that it requires solving a differential equation, which it is not always immediate. Arellano (2003a) proves that it is possible to solve the problem numerically. He applies this method to a logit model with fixed effects, showing that for this particular case an explicit solution can be achieved. Moreover, he finds estimators with lower biases of order T with this methodology. He shows the advantages from working with this kind of estimator as long as T increases, since fixed-T consistent estimators rely on exactly unbiased moment conditions.

Taking this paper as reference point, I work with a more complex model. Here one needs to perform a simulation to study the asymptotic properties of the estimators obtained. The first objective of this paper is to apply the technique developed by Cox and Reid to the Tobit model, or *censored regression model*, to panel data with fixed effects. This methodology relies on two ideas. The first one is *orthogonalization*, which is a partial first step to try to solve the incidental parameters problem. What is achieved with this step is that the estimation by maximum likelihood of the orthogonal reparametrization is not strongly dependent on the rest of the parameters.

The second idea is to use an *approximation*. This idea is associated with the modification of the concentrated (or profile) log-likelihood, which is the log-likelihood evaluated at the Maximum Likelihood Estimator (MLE) of the orthogonal fixed effect. The objective is to approximate this function by the conditional log-likelihood given the MLE of the orthogonal fixed effect. Arellano (2003a) shows that this idea will lead to an improvement in T of the scores' bias. Such a result will lead us to consider the possibility that the estimators obtained with this method could have lower biases than those obtained by maximum likelihood. To check this statement I will carry out a small scale Monte Carlo study designed to illustrate the properties of these estimators for different values of T.

Additionally, we will include, as an illustration, an application of the method developed to the case of earnings in Spain in the 1980s, where the data set has the structure of a balanced panel subject to censoring. Indeed, what we observe is the taxable earnings base rather than actual earnings, which are subject to floors and ceilings that depend on the worker's occupation and they vary over time. Taking the paper of Bover, Bentolila and Arellano (2002) as the reference, I compare their estimates of earnings returns with those obtained when I control for unobserved heterogeneity, taking advantage of the large T of this database, namely 8 years.

The paper is organized as follows. In Section 2 the Tobit model with panel data is introduced, and the incidental parameters problem described. In Section 3 we explain the methodology proposed by Cox and Reid (1987), and we apply it to the Tobit model. Section 4 includes a Monte Carlo study to analyze the asymptotic properties of the estimators proposed in Section 2, comparing the results with those obtained using other methods. An empirical exercise is done in Section 5, and the estimation results are shown. Section 6 contains the conclusions.

2. The model

We consider a Tobit model with individual effects given by

$$\begin{aligned} y_{it}^* &= x_{it}'\beta + \eta_i + v_{it} \qquad i = 1, ..., N; \text{ where } v_i \,|x_i, \eta_i \overset{d}{\sim} N(0, \sigma^2 I_T) \\ y_{it} &= \max\{y_{it}^*, c\} \qquad t = 1, ..., T \end{aligned} \qquad [1]$$

where c, which is in the interval $[-\infty, \infty)$, is the lowest possible value that some observed economic variable y can take, and y^* is the desired level of that variable in the absence of a constraint, also called the latent variable.

It is important to note that if $c \to -\infty$, then $y_{it} = y^*_{it} = x'_{it}\beta + \eta_i + \varepsilon_{it}$, so that the linear regression model is a particular case of the Tobit model. This fact will help us understand the results that we are going to obtain, because the linear model has been widely studied and the properties of its estimators have been proved.

In order to simplify notation we denote the expression $\left(c - x'_{it}\beta - \eta_i\right)/\sigma$ by e_{it}, so that $E\left(\frac{y_{it} - x'_{it}\beta - \eta_i}{\sigma} \,|x_i, \eta_i\right) = \phi\left(e_{it}\right) + e_{it}\Phi\left(e_{it}\right)$, where ϕ is the density function and Φ is the cumulative distribution function of a standardized normal, $N(0,1)$, distribution.

On the other hand, if d_{it} is the binary variable related to the censoring of our model, i.e. $d_{it} = I\left(y_{it} > c\right)$, we have

$$E\left(d_{it}\,|x_i, \eta_i\right) = 1 - \Phi\left(e_{it}\right). \qquad [2]$$

Moreover,

$$E\left[d_{it}\left(\frac{y_{it} - x'_{it}\beta - \eta_i}{\sigma}\right)|x_i, \eta_i\right] = \phi(e_{it}). \qquad [3]$$

Furthermore, the density function of y_{it} given x_i and η_i takes the form

$$\begin{cases} \frac{1}{\sigma}\phi\left(\frac{y_{it} - x'_{it}\beta - \eta_i}{\sigma}\right) & \text{if } d_{it}=1, \\ \Phi\left(e_{it}\right) & \text{if } d_{it}=0. \end{cases}$$

From this expression we have that the log-likelihood for individual i is[2]

$$l_i(\theta, \eta_i) = \sum_{t=1}^{T} \left\{ \begin{array}{c} d_{it}\log\left[\frac{1}{\sigma}\phi\left(\frac{y_{it} - x'_{it}\beta - \eta_i}{\sigma}\right)\right] \\ + \left(1 - d_{it}\right)\log\Phi\left(\frac{c - x'_{it}\beta - \eta_i}{\sigma}\right) \end{array} \right\}. \qquad [4]$$

And under the assumption that the y_{it} variables are independent conditional on x_i and η_i, the log-likelihood function is given by

$$\sum_{i=1}^{N} l_i(\theta, \eta_i). \qquad [5]$$

2.1 The problem

Let $l_i(\theta, \eta_i)$ be the log-likelihood function for individual i conditional on $x_i = (x'_{i1}, ..., x'_{iT})'$ and η_i. A widely used method for inference on

[2] Note that $\theta' = (\beta' \; \sigma^2)$.

the parameters when we are working with unobserved heterogeneity consists of replacing the estimation by maximum likelihood of these parameters into the log-likelihood of the data, and then to analyze the resulting function of the parameters of interest.

Thus, the Maximum Likelihood Estimator (MLE) of η_i given θ is

$$\widehat{\eta}_i(\theta) = \arg\max_{\eta_i} l_i(\theta, \eta_i), \tag{6}$$

which satisfies

$$d_{\eta_i}(\theta, \widehat{\eta}_i(\theta)) = \frac{\partial l_i(\theta, \widehat{\eta}_i(\theta))}{\partial \eta_i} = 0. \tag{7}$$

If $L(\theta) = \sum_{i=1}^{N} l_i(\theta, \widehat{\eta}_i(\theta))$ is the concentrated log-likelihood function, the MLE of θ is given by

$$\widehat{\theta} = \arg\max_{\theta} \sum_{i=1}^{N} l_i(\theta, \widehat{\eta}_i(\theta))$$

and its first order condition is therefore

$$\sum_{i=1}^{N} \left\{ d_{\theta i}(\theta, \widehat{\eta}_i(\theta)) + d_{\eta_i}(\theta, \widehat{\eta}_i(\theta)) \frac{\partial \widehat{\eta}_i(\theta)}{\partial \theta} \right\} = \sum_{i=1}^{N} d_{\theta i}(\theta, \widehat{\eta}_i(\theta)), \tag{8}$$

where $d_{\theta i}(\theta, \eta_i) = \partial l_i(\theta, \eta_i)/\partial \theta$.

Since the publication of the paper by Neyman and Scott (1948) it has been well known that, since the number of individual effects to be estimated increases with N, this procedure yields inconsistent estimators of θ when $N \to \infty$ for T fixed (although it is consistent when $T \to \infty$). This situation is known as the *incidental parameters problem*.

For example, in the linear regression model given by $y_{it} = x'_{it}\beta + \eta_i + v_{it}$, where $v_i \mid x_i, \eta_i \overset{d}{\sim} N(0, \sigma^2 I_T)$, although consistent estimators of β are obtained (the Within-Group estimator), it can be checked that $p \lim_{N\to\infty} \widehat{\sigma}^2 = \sigma^2 - \frac{\sigma^2}{T}$ (see Arellano, 2003b). More generally, Liang (1987) proved that this class of estimator are in general inconsistent of order $O(1/T)$.

3. Modified profile likelihood

3.1 Orthogonalization

The incidental parameters problem will not have any effect if the log-likelihood for individual i has the form

$$l_i(\theta, \eta_i) = l_{1i}(\theta) + l_{2i}(\eta_i), \tag{9}$$

where $l_{1i}(\theta)$ and $l_{2i}(\eta_i)$ are log-likelihood functions. If this factoring is true for every member of the population, and the parameter space for η does not depend on the value taken by θ, then the presence of η_i in the likelihood is of no consequence. In other words, $\widehat{\eta}_i$ does not depend on the value of θ.

Unfortunately few models allow for this factorization, which leads us to introduce a weaker kind of orthogonality, *information orthogonality*, whose verification requires the average of the cross-scores to be equal to zero. That is, if for the case of the stronger orthogonality we have $l_i(\theta, \eta_i) = l_{1i}(\theta) + l_{2i}(\eta_i) \Longleftrightarrow \frac{\partial^2 l_i(\theta,\eta_i)}{\partial\theta\partial\eta_i} = 0$, then we define *information orthogonality* based on the verification on average of this property: $E\left(\frac{\partial^2 l_i(\theta,\eta_i)}{\partial\theta\partial\eta_i}\right) = 0$.

In spite of this relaxation it is possible that θ and η_i do not satisfy the information orthogonality property, so what we can do is a reparametrization from (θ, η_i) to (θ, λ_i) to achive this weak form of orthogonality between θ and λ_i. Thus $\eta_i = \eta(\theta, \lambda_i)$ is chosen such that the reparametrized log-likelihood, $l_i^*(\theta, \lambda_i) = l_i(\theta, \eta_i(\theta, \lambda_i))$, satisfies at the true values the condition:

$$E\left(\frac{\partial^2 l_i^*(\theta_0, \lambda_i)}{\partial\theta\partial\lambda_i}\,|x_i, \eta_i\right) = -E\left(\frac{\partial l_i^*(\theta_0, \lambda_i)}{\partial\theta}\frac{\partial l_i^*(\theta_0, \lambda_i)}{\partial\lambda_i}\,|x_i, \eta_i\right) = 0. \tag{10}$$

This condition has different implications. For instance, the information matrix becomes block diagonal and, likewise, η_i must have a particular form. Following Cox and Reid (1987) and Lancaster (1997), the function $\eta_i(\theta, \lambda_i)$ must satisfy the system of partial differential equations

$$\frac{\partial\eta_i}{\partial\theta} = -E\left(\frac{\partial^2 l_i(\theta_0, \eta_i)}{\partial\theta\partial\eta_i}\,|x_i, \eta_i\right) \Big/ E\left(\frac{\partial^2 l_i(\theta_0, \eta_i)}{\partial\eta_i^2}\,|x_i, \eta_i\right). \tag{11}$$

If [11] holds for all θ in the parameter space, it is called global orthogonality. If it only holds for a value θ_0, it is called local orthogonality.

Orthogonality can always be achieved locally, but global orthogonality is only possible in some especial cases (Cox and Reid, 1987). One of the reasons is that global orthogonality is related to an explicit solution. If we were able to obtain one, there would be global orthogonality (a sufficient condition); mathematically, this is known to be very complicated if not impossible. Therefore, it would be possible to recover $l_i(\theta, \eta_i)$. This methodology has been criticized by the followers of the Bayesian approach, because it is not always possible to obtain an explicit function for $\lambda_i = \lambda(\theta, \eta_i)$. Arellano (2003a) shows that this is not a problem, since it is always possible to work with the first order condition of the problem, as we will note in the next section.

- *Orthogonalization in the Tobit Model*

Now, let us apply all these formulae to the Tobit model. To achive more comprehensible expressions we define some real functions that will appear in the subsequent developments[3] :

$$\Lambda(e) = \phi(e)\,/\Phi(e)\ , \qquad [12]$$

$$M(e) = -\Lambda'(e) = \Lambda^2(e) + \Lambda(e)e = \Lambda(e)\left[\Lambda(e) + e\right], \qquad [13]$$

$$H(e) = \Lambda''(e) = -M'(e) = M(e)\left[2\Lambda(e) + e\right] - \Lambda(e), \qquad [14]$$

$$\varphi(e) = 1 + \Phi(e)\left[M(e) - 1\right], \qquad [15]$$

$$\varphi'(e) = \phi(e)\left[M(e) - 1\right] - \Phi(e)H(e), \qquad [16]$$

$$\xi(e) = 2\phi(e) + \Phi(e)\left[M(e)e - \Lambda(e)\right], \qquad [17]$$

$$\xi'(e) = -2\phi(e)e + \phi(e)\left[M(e)e - \Lambda(e)\right] + \Phi(e)\left[2M(e) - H(e)e\right]. \qquad [18]$$

In this way, we are able to calculate the partial derivatives needed for the orthogonalization.

a) With respect to η_i:

$$d_{\eta i} = \frac{1}{\sigma}\sum_{t=1}^{T}\left\{d_{it}\left(\frac{y_{it} - x'_{it}\beta - \eta_i}{\sigma}\right) - (1 - d_{it})\Lambda(e_{it})\right\}, \qquad [19]$$

$$d_{\eta\eta i} = -\frac{1}{\sigma^2}\sum_{t=1}^{T}\left\{d_{it} + (1 - d_{it})M(e_{it})\right\}, \qquad [20]$$

$$d_{\eta\eta\eta i} = -\frac{1}{\sigma^3}\sum_{t=1}^{T}\left\{(1 - d_{it})H(e_{it})\right\}. \qquad [21]$$

[3]Some useful results of these functions for the subsequent developments have been summarized in the Appendix.

b) With respect to β:

$$d_{\beta i} = \frac{1}{\sigma} \sum_{t=1}^{T} \left\{ d_{it} \left(\frac{y_{it} - x'_{it}\beta - \eta_i}{\sigma} \right) - (1 - d_{it})\Lambda(e_{it}) \right\} x_{it}, \qquad [22]$$

$$d_{\beta\eta i} = -\frac{1}{\sigma^2} \sum_{t=1}^{T} \{ d_{it} + (1 - d_{it}) M(e_{it}) \} x_{it}, \qquad [23]$$

$$d_{\eta\eta\beta i} = -\frac{1}{\sigma^3} \sum_{t=1}^{T} \{ (1 - d_{it}) H(e_{it}) \} x_{it}. \qquad [24]$$

c) With respect to σ^2:

$$d_{\sigma^2 i} = \frac{1}{2\sigma^2} \sum_{t=1}^{T} \left\{ d_{it} \left[\left(\frac{y_{it} - x'_{it}\beta - \eta_i}{\sigma} \right)^2 - 1 \right] - (1 - d_{it})\Lambda(e_{it})e_{it} \right\}, \qquad [25]$$

$$d_{\sigma^2 \eta i} = -\frac{1}{2\sigma^3} \sum_{t=1}^{T} \left\{ 2d_{it} \left(\frac{y_{it} - x'_{it}\beta - \eta_i}{\sigma} \right) + (1\text{-}d_{it}) \left[M(e_{it})e_{it}\text{-}\Lambda(e_{it}) \right] \right\}, \qquad [26]$$

$$d_{\eta\eta\sigma^2 i} = \frac{1}{2\sigma^4} \sum_{t=1}^{T} \{ 2d_{it} + (1 - d_{it}) \left[2M(e_{it}) - H(e_{it})e_{it} \right] \}. \qquad [27]$$

Taking conditional expectations in [20] and [23] we obtain

$$E\left(d_{\eta\eta i}(\theta_0, \eta_i) \mid x_i \ \eta_i\right) = -\frac{1}{\sigma^2} \sum_{t=1}^{T} \varphi\left(\frac{c - x'_{it}\beta_0 - \eta_i}{\sigma_0}\right), \qquad [28]$$

$$E\left(d_{\beta\eta i}(\theta_0, \eta_i) \mid x_i \ \eta_i\right) = -\frac{1}{\sigma^2} \sum_{t=1}^{T} \varphi\left(\frac{c - x'_{it}\beta_0 - \eta_i}{\sigma_0}\right) x_{it}. \qquad [29]$$

If $e^0_{it} = \frac{c - x'_{it}\beta_0 - \eta_i}{\sigma_0}$, we have, by the orthogonalization of β, that

$$\frac{\partial \eta_i}{\partial \beta} = -\frac{\sum_{t=1}^{T} \varphi(e^0_{it}) x_{it}}{\sum_{t=1}^{T} \varphi(e^0_{it})}. \qquad [30]$$

Moreover, from [26] we have

$$E\left(d_{\sigma^2 \eta i}(\theta_0, \eta_i) \mid x_i \ \eta_i\right) = -\frac{1}{2\sigma^3} \sum_{t=1}^{T} \xi(e^0_{it}), \qquad [31]$$

and also, by the orthogonalization of σ^2,

$$\frac{\partial \eta_i}{\partial \sigma^2} = -\frac{1}{2\sigma}\frac{\sum_{t=1}^{T}\xi(e_{it}^0)}{\sum_{t=1}^{T}\varphi(e_{it}^0)}. \qquad [32]$$

3.2 Modified likelihood

Cox and Reid (1987) suggest an approximation to the conditional likelihood given a statistic for the individual effect (they considered $\widehat{\lambda}_i(\theta)$). Their approach was motivated by the knowledge that if they were able to obtain a sufficient statistic for the nuisance parameters, then the incidental parameters problem would be solved. This statistic can be regarded as the MLE of nuisance parameters chosen so as to be orthogonal to the parameters of interest in an exponential family model (i.e. $\widehat{\lambda}_i(\theta) = \widehat{\lambda}_i$ is a sufficient statistic for λ_i[4]). But if we are not working with that kind of model, Cox and Reid proved that with the orthogonalization we ensure that $\widehat{\lambda}_i(\theta)$ changes slowly with θ. Hence, the modified profile likelihood was born as an approximation to the conditional likelihood given $\widehat{\lambda}_i(\theta)$.

The modification in question for individual i is[5]

$$l_{Mi}(\theta) = l_i^*\left(\theta, \widehat{\lambda}_i(\theta)\right) - \frac{1}{2}\log\left[-\frac{\partial^2 l_i^*\left(\theta, \widehat{\lambda}_i(\theta)\right)}{\partial \lambda_i^2}\right]. \qquad [33]$$

Thus, the log-likelihood function is

$$L_M(\theta) = \sum_{i=1}^{N} l_{Mi}(\theta). \qquad [34]$$

This approximation is extremely useful when it is proved that its use allows us to center the concentrated score function, that is, its biases are reduced with respect to T (see Arellano, 2003a):

$$E\left(\frac{\partial l_i^*(\theta_0, \widehat{\lambda}_i(\theta_0))}{\partial \theta}\,|x_i, \eta_i\right) = O(1), \qquad [35]$$

$$E\left(\frac{\partial l_{Mi}(\theta_0)}{\partial \theta}\,|x_i, \eta_i\right) = O(\frac{1}{T}). \qquad [36]$$

[4] For instance, in the linear model $\widehat{\lambda}_i = \overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, which is a sufficient statistic for λ_i.

[5] For the linear model we have that $l_{Mi}(\theta) = l_i(\theta \left| \widehat{\lambda}_i = \overline{y}_i \right.)$, because we are working with the exponential family model (see footnote 4).

Thus, we expect an important improvement of the estimators' biases derived from the modified profile likelihood when T becomes large. This statement will be analyzed in the next section with a Monte Carlo study.

The fact that the concentrated likelihood depends on the fixed effects orthogonal to θ could be regarded as a problem. Nevertheless, we can express it in terms of the original parametrization (see Arellano, 2003a):

$$l_{Mi}(\theta) = l_i\left(\theta, \widehat{\eta}_i(\theta)\right) - \frac{1}{2}\log\left[-d_{\eta\eta i}(\theta, \widehat{\eta}_i(\theta))\right] - \log\left(\left|\frac{\partial \eta_i}{\partial \lambda_i}\right| |_{\eta_i = \widehat{\eta}_i(\theta)}\right). \qquad [37]$$

It is important to note that in our case $l_{Mi}(\theta)$ is only defined for individuals who satisfy $\sum_{t=1}^{T} d_{it} > 0$, or in other words, those who are not censored in all periods. If this is not the case for agent i, we would have that $\widehat{\eta}_i(\theta) \to -\infty$ for all θ, and thus $\Phi\left(\frac{c - x_{it}'\beta - \widehat{\eta}_i(\theta)}{\sigma}\right) \to 1$, his contribution to the concentrated log-likelihood being equal to zero. Therefore, individuals who are always censored do not give us relevant information about θ, and, for this reason, we have not taken them into account in the estimation.

Nevertheless, it is possible that we will not be able to determine $\log\left(\left|\frac{\partial \eta_i}{\partial \lambda_i}\right| |_{\eta_i = \widehat{\eta}_i(\theta)}\right)$. However, it is easy to note that (Arellano, 2003a):

$$\frac{\partial}{\partial \theta} log \left|\frac{\partial \eta_i}{\partial \lambda_i}\right| = \frac{\partial}{\partial \eta_i} \left(-E\left(\frac{\partial^2 l_i(\theta_0, \eta_i)}{\partial \theta \partial \eta_i} |x_i, \eta_i\right) \Big/ E\left(\frac{\partial^2 l_i(\theta_0, \eta_i)}{\partial \eta_i^2} |x_i, \eta_i\right)\right). \qquad [38]$$

Thus, we can always obtain the first order condition of the problem $\underset{\theta}{max} L_M(\theta)$, that is:

$$\sum_{i=1}^{N} \Big\{ d_{\theta i}\left(\theta, \widehat{\eta}_i(\theta)\right) + d_{\eta i}\left(\theta, \widehat{\eta}_i(\theta)\right) \widehat{\eta}_{\theta i}(\theta) - \tfrac{1}{2 d_{\eta\eta i}(\theta, \widehat{\eta}_i(\theta))}\left[d_{\eta\eta\theta i}\left(\theta, \widehat{\eta}_i(\theta)\right) + d_{\eta\eta\eta i}\left(\theta, \widehat{\eta}_i(\theta)\right)\ \widehat{\eta}_{\theta i}(\theta)\right] - \tfrac{\partial}{\partial \theta}\log\left|\tfrac{\partial \eta_i}{\partial \lambda_i}\right| \Big\} = 0 \qquad [39]$$

- *Modified profile likelihood in the Tobit Model*

For the Tobit model we have at true values:

$$\frac{\partial}{\partial\beta}\log\left|\frac{\partial\eta_i}{\partial\lambda_i}\right|$$
$$=\frac{1}{\sigma}\frac{\left(\sum_{t=1}^T\varphi'(e_{it})x_{it}\right)\left(\sum_{t=1}^T\varphi(e_{it})\right)-\left(\sum_{t=1}^T\varphi(e_{it})x_{it}\right)\left(\sum_{t=1}^T\varphi'(e_{it})\right)}{\left(\sum_{t=1}^T\varphi(e_{it})\right)^2},\qquad [40]$$

$$\frac{\partial}{\partial\sigma^2}\log\left|\frac{\partial\eta_i}{\partial\lambda_i}\right|=$$
$$\frac{1}{2\sigma^2}\frac{\left(\sum_{t=1}^T\xi'(e_{it})\right)\left(\sum_{t=1}^T\varphi(e_{it})\right)-\left(\sum_{t=1}^T\xi(e_{it})\right)\left(\sum_{t=1}^T\varphi'(e_{it})\right)}{\left(\sum_{t=1}^T\varphi(e_{it})\right)^2}.\qquad [41]$$

Moreover, given σ, from these expressions we can achive an exact formula for λ_i.

PROPOSITION 1 *For the Tobit model with fixed effects and standard and independently distributed errors, an effect orthogonal to η_i, λ_i, for fixed σ is given by*

$$\lambda_i=\overline{x}_i'\beta+\eta_i+\frac{\sigma}{T}\sum_{t=1}^T\int_{-\infty}^{e_{it}}F(e)de \qquad [42]$$

where $F(e)=1-\varphi(e)$ is a distribution function[6] .

Proof: see Appendix.

On the other hand, when all the parameters are unknown, we will work with local orthogonality, as we show in the next proposition.

PROPOSITION 2 *When σ is unknown, global orthogonality can be achieved only for the linear case.*

Proof: see Appendix.

[6] Note that for the linear case we have $\lambda_i=\overline{x}_i'\beta+\eta_i$ (see Lancaster, 1997).

3.3 Asymptotic results

For some time now an important part of the investigation with panel data has had the objective of searching for fixed-T, large-N consistent estimators. In some cases T is higher than 2, which shows the relevance of trying to work with estimators that reduce the order of the bias in T. In this sense, given that the modified profile likelihood reduces the bias of the scores of the log-likelihood, we expect that these estimators will improve with T (in terms of biases) more quickly than the MLE. From this point of view, Arellano (2003a) has analyzed the asymptotic properties of the Modified Maximum Likelihood Estimator (MMLE) when $T/N \to c$, with $0 < c < \infty$. For the Tobit model, I will study its asymptotic properties with a Monte Carlo study.

Regarding consistency, the ML estimator of θ can be shown to be consistent as $T \to \infty$, regardless of N, because it satisfies the properties needed for the consistency theorem in Amemiya (1985, pp. 270-72). For σ known, it is easy to prove the consistency of the MMLE, because it is sufficient to realize that this follows from the previous formulae. We have that

$$p\lim_{T\to\infty}\frac{1}{T}l_{Mi}(\theta) = p\lim_{T\to\infty}\frac{1}{T}l_i\left(\theta,\widehat{\eta}_i(\theta)\right) -$$

$$p\lim_{T\to\infty}\frac{1}{T}\log\left[\frac{1}{T\sigma^2}\sum_{t=1}^{T}\left\{d_{it}+(1-d_{it})\widehat{M}(\theta)\right\}\right]^{1/2}$$

$$+p\lim_{T\to\infty}\frac{1}{T}\log\left[\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\varphi}(\theta)\right)\right] = p\lim_{T\to\infty}\frac{1}{T}l_i\left(\theta,\widehat{\eta}_i(\theta)\right), \qquad [43]$$

where we have used that $\widehat{\varphi}(\theta) = \varphi(\widehat{e}_{it})$ and that $\widehat{M}(\theta) = M(\widehat{e}_{it})$. This convergence is uniform in θ in a neighborhood of θ_0. When σ is unknown, I will check this result by using a simulation.

4. Monte Carlo investigation

In this section I summarize a Monte Carlo study designed to illustrate the sample properties of a Tobit model with fixed effects. I have carried out this study both for the ML and the MML estimators.

All the results presented here are based on 1000 replications from the model:

$$y_{it} = \max\left\{\sum_{k=1}^{K} x_{itk}\beta_k + \eta_i + v_{it}, 0\right\} \qquad i = 1, ..., N;\ t = 1, ..., T \tag{44}$$

where $v_i \,|x_i, \eta_i \overset{d}{\sim} N(0, \sigma^2 I_T)$.

In the specification I have worked with $N = 100$ and $K = 2$, while T varies from 2 to 20 so as to analyze the behaviour of both estimators for different sample sizes of the time-series. Similarly, I have assumed that $x_{it1} = \eta_i + \varepsilon_{it}$, to introduce linear correlation between the fixed effects and the individual variables. The true values of the generator model are always $\beta = (1\ 1)'$ and $\sigma = 0.7$,[7] whereas the random variables x_{it2} , η_i and ε_{it} are independent and distributed as $N(0, 1)$, following Honoré (1992).

4.1 Solution

All the calculations reported in this paper were performed in GAUSS 5.3.1 (R11.1) and its pseudo-random generators were used to generate the samples.

We can summarize the algorithm used for the MML estimations for a fixed T in the following steps:

- Step 1: Repeat 1,000 times. Construct the generator model from the values $\beta = (1\ 1)'$ and $\sigma = 0.7$, where the random variables x_{it2} , η_i and ε_{it} are independent and normally distributed as $N(0, 1)$.

- Step 2: Fix initial values for η and θ, say η^0 and θ^0.

- Step 3: Set $r = 0$, $\widetilde{\eta}^r = \eta^0$ and $\widetilde{\theta}^r = \theta^0$.

  Step 3.1: Set $r = r + 1$. For $\widetilde{\theta}^{r-1}$ fixed, determine the $\widetilde{\eta}^r$ that satisfies for every $i = 1, ..., N$:

$$\widetilde{\eta}_i^r(\widetilde{\theta}^{r-1}) = \arg\max_{\eta i} l_i(\widetilde{\theta}^{r-1}, \eta_i) \tag{45}$$

[7] These specific values do not affect the result.

where

$$l_i(\widetilde{\theta}^{r-1}, \eta_i) = \sum_{t=1}^{T} \left\{ -\frac{1}{2} d_{it} \left[\log\left(\widetilde{\sigma}^{2(r-1)}\right) + \left(\frac{y_{it} - x'_{it}\widetilde{\beta}^{r-1} - \eta_i}{\widetilde{\sigma}^{r-1}} \right)^2 \right] + (1 - d_{it}) \log \Phi \left(\frac{c - x'_{it}\widetilde{\beta}^{r-1} - \eta_i}{\widetilde{\sigma}^{r-1}} \right) \right\}. \quad [46]$$

Step 3.2: For $\widetilde{\eta}^r$ fixed, determine the $\widetilde{\theta}^r$ that satisfies

$$\sum_{i=1}^{N} \left\{ d_{\theta i}\left(\widetilde{\theta}^r, \widetilde{\eta}^r\right) + d_{\eta i}\left(\widetilde{\theta}^r, \widetilde{\eta}^r\right) \widetilde{\eta}^r - \frac{1}{2 d_{\eta\eta i}\left(\widetilde{\theta}^r, \widetilde{\eta}^r\right)} \left[d_{\eta\eta\theta i}\left(\widetilde{\theta}^r, \widetilde{\eta}^r\right) + d_{\eta\eta\eta i}\left(\widetilde{\theta}^r, \widetilde{\eta}^r\right) \widetilde{\eta}^r \right] - \frac{\partial}{\partial \theta} \log \left| \frac{\partial \eta_i}{\partial \lambda_i} \right| \right\} = 0 \quad . \quad [47]$$

Step 3.3: Repeat from Step 3.1 if

$$max \left\{ \|\widetilde{\eta}^r - \widetilde{\eta}^{r-1}\|_\infty, \|\widetilde{\theta}^r - \widetilde{\theta}^{r-1}\|_\infty \right\} \geq tolerance.^8 \quad [48]$$

In practice, one of the main problems is to choose some initial values which are sufficiently close to the true ones (which are unknown) to ensure a quick convergence. In particular, it is useful to get a quick approximation of $\widetilde{\eta}^1(\theta^0)$ given any initial value η^0.

If we solve the system resulting from the first order conditions of the maximization of the η's given θ, we face a system of N equations which are not related. For each of them we have

$$d_{\eta_i}(\theta, \widehat{\eta}_i(\theta)) = \frac{1}{\sigma} \sum_{t=1}^{T} \left\{ d_{it} \left(\frac{y_{it} - x'_{it}\beta - \widehat{\eta}_i}{\sigma} \right) - (1 - d_{it}) \Lambda(\widehat{e}_{it}) \right\} = 0. \quad [49]$$

The heuristic strategy consists of viewing this as a fixed point problem, because after solving it we get[9]

$$\widehat{\eta}_i = P(\widehat{\eta}_i) = \frac{\sum_{t=1}^{T} \left\{ d_{it} \left(y_{it} - x'_{it}\beta \right) - \sigma(1 - d_{it}) \Lambda(\widehat{e}_{it}) \right\}}{\sum_{t=1}^{T} d_{it}}, \quad [50]$$

[8] $\|.\|_\infty$ denotes the infinity norm in R^N. That is, $\|x\|_\infty = \max_{i=1,\ldots,N} |x_i|$, for $x \in R^N$.

[9] Note that for the linear case $d_{it} = 1$. Thus, $\widehat{\eta}_i = \frac{1}{T} \sum_{t=1}^{T} \{y_{it} - x'_{it}\beta\} = \overline{y}_i - \overline{x}'_i\beta$.

which is well defined because we are working with the individuals such that $\sum_{t=1}^{T} d_{it} > 0$.

Now we would only have to choose an initial value and, assuming that the function $P = (P_1, ..., P_N)$ is a contraction, to calculate P^2, P^3, ..., P^n until we converge to the fixed point[10]. We have done the simulations for the ML estimators analogously.

TABLE 1
Results of the simulations for the MMLE with $N = 100$, 1000 replications and censoring 50%

		True	Mean	Bias	SD	RMSE	LQ	Median	UQ	MAE	PMAE
T=2	β_1	1.0000	0.9920	-0.0080	0.0969	0.0971	0.9255	0.9936	1.0567	0.0648	6.48%
	β_2	1.0000	0.9984	-0.0016	0.1090	1.089	0.9237	0.9965	1.0661	0.0711	5.97%
	σ	0.7000	0.6532	-0.0468	0.0703	0.0844	0.6039	0.6499	0.7025	0.0605	8.64%
T=4	β_1	1.0000	1.0050	0.0050	0.0788	0.0789	0.9525	1.0026	1.0613	0.0537	5.37%
	β_2	1.0000	0.9992	-0.0008	0.0817	0.0816	0.9424	0.9983	1.0525	0.0552	6.28%
	σ	0.7000	0.6776	-0.0224	0.0542	0.0586	0.6421	0.6753	0.7125	0.0414	5.91%
T=6	β_1	1.0000	1.0003	0.0003	0.0706	0.0706	0.9509	0.9989	1.0458	0.0467	4.67%
	β_2	1.0000	0.9961	-0.0039	0.0714	0.0715	0.9464	0.9919	1.0420	0.0480	4.70%
	σ	0.7000	0.6864	-0.0136	0.0492	0.0510	0.6524	0.6862	0.7190	0.0359	5.13%
T=8	β_1	1.0000	1.0006	0.0006	0.0391	0.0391	0.9755	0.9986	1.0264	0.0253	2.53%
	β_2	1.0000	0.9992	-0.0008	0.0412	0.0412	0.9704	0.9981	1.0268	0.0279	2.68%
	σ	0.7000	0.6960	-0.0040	0.0275	0.0278	0.6778	0.6955	0.7142	0.0188	2.69%
T=10	β_1	1.0000	1.0026	0.0026	0.0353	0.0353	0.9790	1.0027	1.0259	0.0233	2.33%
	β_2	1.0000	1.0001	0.0001	0.0343	0.0343	0.9783	1.0017	1.0229	0.0222	2.29%
	σ	0.7000	0.6954	-0.0046	0.0244	0.0248	0.6786	0.6964	0.7118	0.0164	2.34%
T=12	β_1	1.0000	1.0021	0.0021	0.0345	0.0346	0.9791	1.0020	1.0257	0.0237	2.37%
	β_2	1.0000	1.0011	0.0011	0.0352	0.0352	0.9785	1.0007	1.0250	0.0236	2.16%
	σ	0.7000	0.6978	-0.0022	0.0232	0.0233	0.6826	0.6972	0.7123	0.0150	2.14%
T=14	β_1	1.0000	0.9992	-0.0008	0.0281	0.0281	0.9801	0.9991	1.0172	0.0188	1.88%
	β_2	1.0000	1.0008	0.0008	0.0288	0.0288	0.9817	1.0009	1.0207	0.0190	1.93%
	σ	0.7000	0.6971	-0.0029	0.0193	0.0195	0.6830	0.6976	0.7096	0.0131	1.87%
T=16	β_1	1.0000	0.9995	-0.0005	0.0274	0.0274	0.9816	0.9994	1.0171	0.0177	1.77%
	β_2	1.0000	0.9999	-0.0001	0.0281	0.0281	0.9806	0.9993	1.0204	0.0198	1.86%
	σ	0.7000	0.6974	-0.0026	0.0189	0.0191	0.6850	0.6972	0.7098	0.0125	1.79%
T=18	β_1	1.0000	1.0010	0.0010	0.0238	0.0238	0.9841	1.0013	1.0175	0.0170	1.70%
	β_2	1.0000	1.0016	0.0016	0.0254	0.0254	0.9835	1.0020	1.0183	0.0174	1.70%
	σ	0.7000	0.6986	-0.0014	0.0175	0.0176	0.6866	0.6984	0.7102	0.0119	1.70%
T=20	β_1	1.0000	0.9995	-0.0005	0.0234	0.0234	0.9852	0.9992	1.0151	0.0150	1.50%
	β_2	1.0000	0.9997	-0.0003	0.0240	0.0240	0.9837	1.0004	1.0152	0.0155	1.53%
	σ	0.7000	0.6979	-0.0021	0.0164	0.0165	0.6864	0.6976	0.7090	0.0115	1.64%

Note: The results exclude 6.20% of the cases for T=2 and 0.30% for T=4 in which the objective function did not have an unique minimum. SD: standard deviation. RMSE: root of the mean squared error. LQ: first quartil. UQ: third quartil. MAE: median absolute error. PMAE: percentage median absolute error.

[10] This method needs to be more robust to avoid cycle problems, but in practice it works very well and it is very quick.

TABLE 2
Results of the simulations for the MLE with $N = 100$, 1000 replications and censoring 50%

		True	Mean	Bias	SD	RMSE	LQ	Median	UQ	MAE	PMAE
T=2	β_1	1.0000	1.0083	0.0083	0.1149	0.1152	0.9307	1.0059	1.0833	0.0757	7.57%
	β_2	1.0000	1.0055	0.0055	0.1127	0.1128	0.9285	1.0040	1.0779	0.0751	7.48%
	σ	0.7000	0.4308	-0.2692	0.0501	0.2738	0.3963	0.4295	0.4640	0.2706	38.65%
T=4	β_1	1.0000	1.0017	0.0017	0.0784	0.0784	0.9477	1.0005	1.0531	0.0528	5.28%
	β_2	1.0000	1.0032	0.0032	0.0753	0.0753	0.9543	0.9992	1.0516	0.0487	5.07%
	σ	0.7000	0.5630	-0.1370	0.0479	0.1451	0.5318	0.5613	0.5935	0.1387	19.81%
T=6	β_1	1.0000	1.0063	0.0063	0.0623	0.0626	0.9670	1.0050	1.0445	0.0397	3.97%
	β_2	1.0000	1.0044	0.0044	0.0616	0.0617	0.9623	1.0048	1.0424	0.0402	4.12%
	σ	0.7000	0.6078	-0.0922	0.0382	0.0998	0.5825	0.6077	0.6319	0.0923	13.19%
T=8	β_1	1.0000	1.0053	0.0053	0.0379	0.0382	0.9803	1.0041	1.0289	0.0240	2.40%
	β_2	1.0000	1.0030	0.0030	0.0391	0.0392	0.9759	1.0039	1.0306	0.0275	2.79%
	σ	0.7000	0.6316	-0.0684	0.0251	0.0728	0.6158	0.6311	0.6474	0.0689	9.84%
T=10	β_1	1.0000	1.0070	0.0070	0.0363	0.0370	0.9825	1.0077	1.0317	0.0249	2.49%
	β_2	1.0000	1.0031	0.0031	0.0350	0.0351	0.9811	1.0036	1.0277	0.0232	2.35%
	σ	0.7000	0.6453	-0.0547	0.0230	0.0594	0.6294	0.6463	0.6605	0.0537	7.67%
T=12	β_1	1.0000	1.0011	0.0011	0.0305	0.0305	0.9818	1.0009	1.0220	0.0198	1.98%
	β_2	1.0000	1.0019	0.0019	0.0301	0.0302	0.9829	1.0011	1.0238	0.0210	2.05%
	σ	0.7000	0.6540	-0.0460	0.0198	0.0501	0.6410	0.6535	0.6677	0.0465	6.64%
T=14	β_1	1.0000	1.0012	0.0012	0.0281	0.0281	0.9817	1.0009	1.0196	0.0192	1.92%
	β_2	1.0000	1.0028	0.0028	0.0288	0.0290	0.9839	1.0029	1.0230	0.0191	1.94%
	σ	0.7000	0.6911	-0.0389	0.0186	0.0431	0.6478	0.6617	0.6732	0.0383	5.47%
T=16	β_1	1.0000	1.0017	0.0017	0.0264	0.0264	0.9840	1.0011	1.0187	0.0177	1.77%
	β_2	1.0000	1.0018	0.0018	0.0266	0.0266	0.9838	1.0004	1.0196	0.0182	1.71%
	σ	0.7000	0.6654	-0.0346	0.0179	0.0389	0.6535	0.6653	0.6777	0.0347	4.96%
T=18	β_1	1.0000	1.0015	0.0015	0.0250	0.0250	0.9849	1.0009	1.0178	0.0170	1.70%
	β_2	1.0000	1.0019	0.0019	0.0254	0.0255	0.9843	1.0023	1.0183	0.0170	1.73%
	σ	0.7000	0.6703	-0.0297	0.0169	0.0342	0.6593	0.6704	0.6816	0.0296	4.23%
T=20	β_1	1.0000	1.0011	0.0011	0.0221	0.0221	0.9868	1.0003	1.0164	0.0145	1.45%
	β_2	1.0000	1.0007	0.0007	0.0240	0.0240	0.9850	1.0010	1.0157	0.0150	1.46%
	σ	0.7000	0.6720	-0.0280	0.0160	0.0322	0.6604	0.6718	0.6830	0.0282	4.03%

Note: SD: standard deviation. RMSE: root of the mean squared error. LQ: first quartil. UQ: third quartil. MAE: median absolute error. PMAE: percentage median absolute error.

4.2 Results

In Tables 1 to 3 I report for each simulation the estimator, the bias, the standard deviation (SD) and the root of the mean squared error (RMSE), as well as quartiles, the median absolute error (MAE) of the estimator and the percentage median absolute error (PMAE). As for the censoring, it is always about 50% by construction. The behaviour of both kinds of estimators is shown in Figures 1 to 3.

We observe that $\widehat{\beta}_{1MML}$ and $\widehat{\beta}_{1ML}$ have similar biases when T is less than 4, but when T becomes large the improvement (in terms of bias) is greater for the former. Similarly, the bias of both estimators goes to zero when T tends to infinity. On the other hand, in the case of the estimators for β_2, the reduction of the bias is relevant also for low T's.

It is important to note that when we observe the RMSE or the MAE of the β's there are no improvements related to the use of the MML estimators. The RMSE is similar for both kinds of estimators. I will discuss this result below.

TABLE 3
Summary of the simulations with $N = 100$, 1000 replications and censoring of 50%

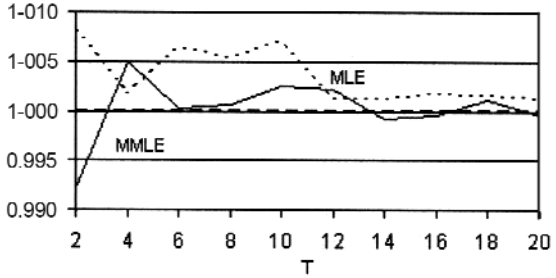
	β_1 1.0000				β_2 1.0000				σ 1.0000			
	ML		MML		ML		MML		ML		MML	
	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
T=2	1.0083	0.1152	0.9920	0.0971	1.0055	0.1128	0.9984	0.1089	0.4308	0.2738	0.6532	0.0844
T=4	1.0017	0.0784	1.0050	0.0789	1.0032	0.0753	0.9992	0.0816	0.5630	0.1451	0.6776	0.0586
T=6	1.0063	0.0626	1.0003	0.0706	1.0044	0.0617	0.9961	0.0715	0.6078	0.0998	0.6864	0.0510
T=8	1.0053	0.0382	1.0006	0.0391	1.0030	0.0392	0.9992	0.0412	0.6316	0.0728	0.6960	0.0188
T=10	1.0070	0.0370	1.0026	0.0353	1.0031	0.0351	1.0001	0.0343	0.6453	0.0594	0.6954	0.0248
T=12	1.0011	0.0305	1.0021	0.0346	1.0019	0.0302	1.0011	0.0352	0.6540	0.0501	0.6978	0.0233
T=14	1.0012	0.0281	0.9992	0.0281	1.0028	0.0290	1.0008	0.0288	0.6611	0.0431	0.6971	0.0195
T=16	1.0017	0.0264	0.9995	0.0274	1.0018	0.0266	0.9999	0.0281	0.6654	0.0389	0.6974	0.0191
T=18	1.0015	0.0250	1.0010	0.0238	1.0019	0.0255	1.0016	0.0254	0.6703	0.0342	0.6986	0.0176
T=20	1.0011	0.0221	0.9995	0.0234	1.0007	0.0240	0.9997	0.0240	0.6720	0.0322	0.6979	0.0165

Note: RMSE: root of the mean squared error.

Regarding the simulated estimators of σ, there is a very important improvement of the MML estimators as opposed to the ML ones for all values of T, both in terms of the bias and of the RMSE or the MAE. The negative sign of the bias for both estimators is also surprising, but this fact has an explanation if we look to the linear case. We have just said that the MLE of σ^2 for the linear model is biased due to the incidental parameter problem. Recall that $p \lim_{N\to\infty} \widehat{\sigma}^2_{ML} = \sigma^2 - \frac{\sigma^2}{T}$, that is, the bias is negative and of order $O(1/T)$. Therefore, as can be seen in Figure 3, the proposed method reduces the bias but maintains its sign.

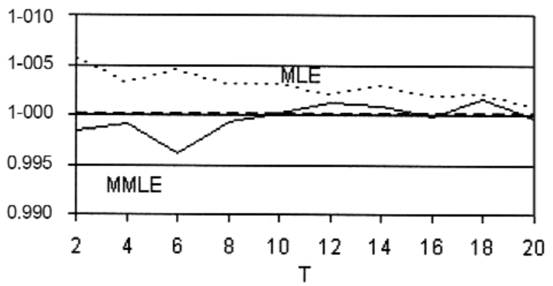
The linear case is also useful for understanding what has happened with the estimators of the β's. For this model it can be shown that both estimators are the same, i.e. $\widehat{\beta}^{linear}_{ML} = \widehat{\beta}^{linear}_{MML}$. By a continuity argument, we could think that the improvement for β will not be as important as that obtained for σ (that was biased for the MMLE). This fact is confirmed with the study.

FIGURE 1
Comparison of estimators of β_1 (means)



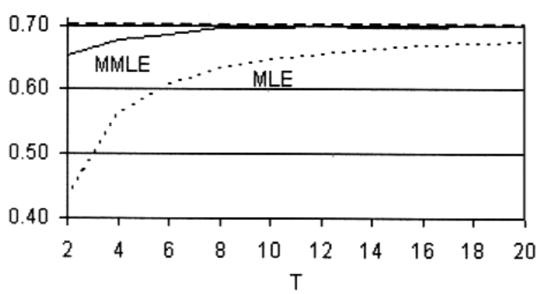
Note: MLE: Maximum Likelihood Estimator. MMLE: Modified Maximum Likelihood Estimator.

FIGURE 2
Comparison of estimators of β_2 (means)



Note: MLE: Maximum Likelihood Estimator. MMLE: Modified Maximum Likelihood Estimator.

FIGURE 3
Comparison of estimators of σ (means)



Note: MLE: Maximum Likelihood Estimator. MMLE: Modified Maximum Likelihood Estimator.

4.3 Other comparisons

In this paper I have discussed the usual search of estimators carried out by econometricians: insisting on fixed-T consistency. In this context Honoré (1992, 2000) contributes a class of estimators based on moment conditions that arise after making some assumptions about the distribution of the errors (but not about their exact form).

Therefore, assuming that the v_{it} are independent and identically distributed conditional on (x_i, η_i) we define

$$v_{ist}(b) = \max\{y_{is}, c + (x_{is} - x_{it})b\} - \max\{c, c + (x_{is} - x_{it})b\} \geq 0. \tag{51}$$

At $b = \beta$, we have

$$\begin{aligned} v_{ist}(\beta) &= \max\{y_{is}, c + (x_{is} - x_{it})\beta\} - \max\{c, c + (x_{is} - x_{it})\beta\} = \\ &= \max\{\eta_i + v_{is}, c - x_{is}\beta, c - x_{it}\beta\} - \max\{c - x_{is}\beta, c - x_{it}\beta\}. \end{aligned} \tag{52}$$

This is a symmetric function in s and t. Then, any function of $v_{ist}(\beta)$ minus the same function $v_{its}(\beta)$ will be symmetrically distributed around 0. Therefore,

$$E\left[\xi(\psi(v_{its}(\beta)) - \psi(v_{ist}(\beta)))\,|x_i\ \eta_i\right] = 0 \tag{53}$$

for any increasing function $\psi(\cdot)$ and any increasing and odd function $\xi(\cdot)$. From which it can be infered that

$$E\left[(\xi(\psi(v_{its}(\beta)) - \psi(v_{ist}(\beta))))\,(x_{is} - x_{it})\right] = 0. \tag{54}$$

This orthogonality condition can be expressed as

$$E\left[r(y_{is}, y_{it}, (x_{is} - x_{it})\,\beta)\,(x_{is} - x_{it})\right] = 0 \tag{55}$$

where $r(\cdot,\cdot,\cdot)$ is a monotone function of its third argument, because of the assumptions about $\psi(\cdot)$ and $\xi(\cdot)$[11] . At the same time, we can turn this moment condition into the first order condition for a convex minimization problem of the form

$$\min_b E\left[R(y_{is}, y_{it}, (x_{is} - x_{it})\,b)\right]. \tag{56}$$

[11] And because $v_{ist}(\beta)$ is monotone decreasing in $(x_{is} - x_{it})\beta$.

Honoré studies the case with $\psi(d) = \xi(d) = d$, $c = 0$ and $T = 2$, and he proposes the estimator $\widehat{\beta}_4$ that solves

$$\widehat{\beta}_4 = \arg\min_b \sum_{i=1}^{N} \chi\left(y_{i1}, y_{i2}, (x_{i1} - x_{i2})b\right) \qquad [57]$$

where

$$\chi(x, y, z) = \begin{cases} x^2 - 2x(y+z) & \text{if } z \leq -y \\ (x-y-z)^2 & \text{if } -y < z < x \\ y^2 + 2y(z-x) & \text{if } x \leq z \end{cases} \qquad [58]$$

For the case $T > 2$, he suggests solving

$$\widehat{\beta}_4 = \arg\min_b \sum_{i=1}^{N} \sum_{s<t} \chi\left(y_{is}, y_{it}, (x_{is} - x_{it})b\right). \qquad [59]$$

Table 4 shows the estimations obtained by simulation for $N = 100$ and $N = 1,000$ replications following the previous methodology. We plot their biases in Figures 4 and 5.

TABLE 4
Results of the simulations for the estimators suggested by Honoré with $N = 100$, 1000 replications and censoring of 50%

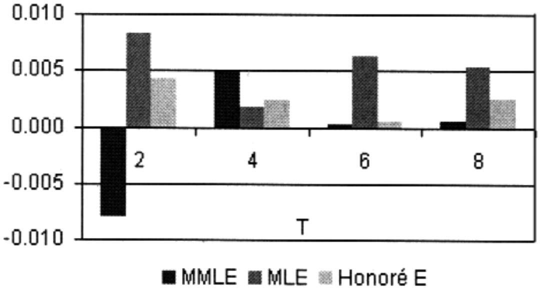
		True	Mean	Bias	SD	RMSE	LQ	Median	UQ	MAE	PMAE
T=2	β_1	1.0000	1.0043	0.0043	0.1244	0.1244	0.9219	1.0015	1.0851	0.0814	8.14%
	β_2	1.0000	1.0001	0.0001	0.1250	0.1250	0.9155	1.0011	1.0797	0.0825	8.15%
T=4	β_1	1.0000	1.0024	0.0024	0.0667	0.0667	0.9583	0.9979	1.0489	0.0455	4.55%
	β_2	1.0000	1.0024	0.0024	0.0716	0.0716	0.9515	1.0003	1.0497	0.0488	4.57%
T=6	β_1	1.0000	1.0005	0.0005	0.0516	0.0516	0.9651	0.9998	1.0343	0.0348	3.48%
	β_2	1.0000	1.0018	0.0018	0.0516	0.0516	0.9680	0.9996	1.0359	0.0343	3.29%
T=8	β_1	1.0000	1.0025	0.0025	0.0431	0.0431	0.9736	0.9994	1.0313	0.0286	2.86%
	β_2	1.0000	1.0015	0.0015	0.0451	0.0451	0.9707	1.0023	1.0314	0.0308	3.10%

Note: SD: standard deviation. RMSE: root of the mean squared error. LQ: first quartil. UQ: third quartil. MAE: median absolute error. PMAE: percentage median absolute error.

These figures show that the methodology suggested by Honoré is always better than Maximum Likelihood (in terms of bias). It is also the one which has the lowest bias for $T = 2$, as it is a method specifically created for a very small time dimension. In general, MML estimators have lower biases than those suggested by Honoré when T increases.

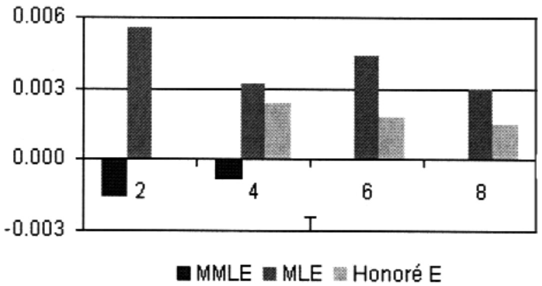
It is relevant to note that, surprisingly, in terms of the RMSE or the MAE Honoré's estimators are best for $T = 4$ or $T = 6$, but this is not the case for $T = 2$ or when T is greater than 6.

FIGURE 4
Comparison of biases for β_1



Note: MLE: Maximum Likelihood Estimator. MMLE: Modified Maximum Likelihood Estimator.

FIGURE 5
Comparison of biases for β_2



Note: MLE: Maximum Likelihood Estimator. MMLE: Modified Maximum Likelihood Estimator.

5. Illustration: An application to earnings equations

In this section I use the model presented previously to estimate the (log)returns to some individual characteristics with earnings in Spain during the 1980s. The aim of this section is to compute a numerical example and to compare the results with those obtained when the fixed effects have not been taken into account. In this sense, I have started from the paper by Bover *et al.* (2002) concerning the distributions of earnings in Spain during the decade of the 1980s and we have examined whether there is any change when we apply this new methodology.

I have worked with a database that contains the information provided monthly by firms when they pay contributions for their employees to the Spanish Social Security System. The matched data set contains information on workers' characteristics, i.e. sex, age, occupation and skill, and on the firm they work for, i.e. its sector, region of location, and size (number of workers).

The main characteristic of this database, which makes it useful for our purpose, is the censoring of earnings. In particular, what we observe is the taxable earnings base rather than actual earnings. This base is subject to floors and ceilings which depend on the worker's occupation and which vary over time. This causes the censoring of earnings for a fraction of the observations, most of them at the top[12] .

The data used are a subsample of those used by Bover *et al.* (2002) which has been previously filtered. Firstly, we start from a balanced panel data of earnings paid in December from 1980 to 1987 ($T = 8$). In this way, we have a total of $9,579$ individuals observed during eight years. From this sample I have eliminated those who were censored at the bottom (0.50%), and I am left with $9,531$ individuals. Moreover, since the subsequent study will be carried out by sector, I have also eliminated those workers who changed sector during the eight-year period (4.39% of the rest). In the end $9,113$ workers are observed during eight years, top coded and remaining in the same sector.

It is important to note the different groups we have considered. For skill we have the following groups: *College graduates* (ingenieros y licenciados), *Junior college* (ingenieros técnicos, peritos, ayudantes titulados y asimilados), *Unskilled* (peones) and *Medium skilled* (residual) (jefes administrativos y de taller, ayudantes no titulados, oficiales administrativos, subalternos, auxiliares administrativos, oficiales de 1^a y 2^a, oficiales de 3^a y especiales).

Regarding firm size I distinguish 3 classes: *small* (up to 100 employees), *medium-sized* (between 101 and 1,000 employees) and *large* firms (above 1,000 employees).

The economic sectors are grouped in the following eight: *Mining* (8), *Construction* (1), *Manufacturing* (2), *Transportation and public utilities* (3), *Wholesale and retail trade* (4), *Finance, insurance, and real estate* (5), *Hotels and Catering* (6) and *Other services* (7).

[12] For a more detailed description of this database see Bover *et al.* (2002).

5.1 Controlling for unobserved heterogeneity

In this section I have controlled for unobserved heterogeneity beyond the cross-sectional analysis performed by Bover *et al.* (2002). For this reason we need seven balanced data panels, one for each sector.

Denoting by w_{its} the observed censored log earnings variable for individual i of industry s in year t, and by w^*_{its} the underlying log earnings, we have

$$w_{its} = \min\{w^*_{its}, c_{it}\} \qquad [60]$$

where c_{it} represents the (log) top codes.

On the one hand, the cross-section analysis presupposes that

$$w^*_{its}\,|x_{it} \overset{d}{\sim} N[\mu_{ts}(x_{it}), \sigma^2_{ts}]. \qquad [61]$$

where x_{it} denotes the individual and firm characteristics. I have used a linear representation for μ_{its}, so that

$$\begin{aligned}
\mu_{its} \;=\;& z'_{it}\beta_{ts} = \beta_{1ts}Small + \beta_{2ts}Medium - sized + \beta_{3ts}Large \\
&+\beta_{4ts}College \times Small \\
&+\beta_{5ts}College \times (Medium - sized + Large) \\
&+\beta_{6ts}Junior\ college \times Small \\
&+\beta_{7ts}Junior\ college \times (Medium - sized + Large) \\
&+\beta_{8ts}Medium\ skilled \times Small \\
&+\beta_{9ts}Medium\ skilled \times (Medium - sized + Large). \qquad [62]
\end{aligned}$$

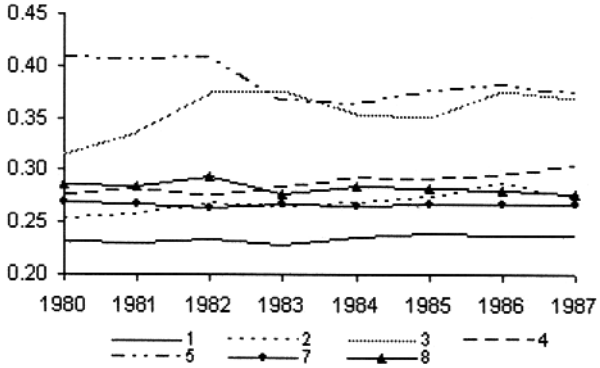
On the other hand, the model with fixed effects for each sector is

$$w^*_{its} = \alpha_{st} + z'_i\beta_{ts} + \eta_i + v_{ist}, \text{ where } v_i\,|x_i, \eta_i \overset{d}{\sim} N(0, \sigma^2 I_T). \qquad [63]$$

For this reason, what we really estimate is the change in the returns with respect to the previous year.

It is important to note that for each sector we have to estimate 64 parameters related to individual and firm characteristics, plus the estimation of σ, which we assume constant over the years. This is not a strong assumption, as we can observe in Figure 6, where we show the estimates of σ obtained with the cross-sectional analysis.

FIGURE 6
Estimations of σ for each sector (without fixed effects)



5.2 Results

Firstly, to compare the estimates obtained controlling for unobserved heterogeneity with those obtained without controlling for it, I have taken differences in the latter with respect to the previous year. I show only a graphical comparison for sector 1 (construction) in Figure 7. Tables 5 to 7 display the estimates of equation (63) for sectors 1, 7 and 8, where I have included the time series correlation with the series without controlling for unobserved heterogeneity to get an idea of their temporal evolution.

TABLE 5
Estimation of the differences of the returns for Sector 1 controlling for unobserved heterogeneity. $N = 679$, averaged censoring of 12.21%

Sector 1	1981	1982	1983	1984	1985	1986	1987	S.C.
Small	-0.0186 (-139.4)	0.0446 (32.7)	0.0619 (109.9)	0.0209 (27.7)	-0.0551 (-40.6)	0.0687 (67.9)	0.0054 (15.7)	0.9599
Medium-sized *(Mz)*	-0.0347 (-30.4)	0.0395 (36.0)	0.0423 (69.0)	-0.0050 (-52.1)	0.0054 (3.4)	0.1006 (67.2)	0.0288 (140.1)	0.8196
Large	-0.0671 (-148.5)	0.0751 (51.4)	0.0444 (52.7)	0.0286 (53.7)	0.0314 (17.3)	0.0917 (51.6)	0.0629 (56.8)	0.7432
College x Small	-0.0120 (1.8)	-0.0132 (-25.5)	-0.0233 (-96.8)	-0.0025 (-3.0)	0.0095 (24.0)	0.0390 (91.4)	0.0340 (147.2)	0.6444
College x *(Mz+Large)*	-0.0073 (-1.6)	0.0156 (15.1)	0.0296 (102.9)	-0.0492 (-17.6)	-0.0268 (-9.9)	-0.0069 (-4.6)	0.0419 (6.2)	0.8327
Junior college x Small	-0.0202 (-3.0)	-0.0314 (-33.5)	0.0489 (25.4)	-0.0504 (-12.9)	-0.0969 (-33.8)	-0.0124 (-1.8)	0.0896 (218.8)	0.7343
Junior college x *(Mz +Large)*	-0.0159 (-7.8)	0.0050 (1.5)	-0.0027 (-1.6)	0.0264 (28.7)	-0.0200 (-4.1)	0.0351 (3.4)	-0.0652 (-27.2)	0.3954
Medium skilled x Small	-0.0096 (-15.9)	-0.0069 (-17.3)	-0.0226 (-149.1)	-0.0124 (-66.3)	0.0277 (105.8)	0.0013 (15.4)	0.0082 (23.5)	0.6725
Medium skilled x *(Mz +Large)*	0.0066 (7.3)	0.0206 (45.7)	-0.0093 (-8.7)	0.0006 (1.2)	-0.0109 (-30.3)	-0.0216 (-15.1)	-0.0286 (-63.3)	0.4946
σ	0.1015	(13.5)						

Note: t-values in brackets. Serial correlation (S.C.) with the differences calculated in 5.1.

TABLE 6

Estimation of the differences of the returns for Sector 7 controlling for unobserved heterogeneity. $N = 927$, averaged censoring of 16.17%

Sector 7	1981	1982	1983	1984	1985	1986	1987	S.C.
Small	-0.0128	-0.0204	0.0606	0.0086	-0.0053	0.0249	0.0367	0.6323
	(-7.4)	(-44.9)	(105.4)	(355.6)	(-44.6)	(610.1)	(425.9)	
Medium-sized *(Mz)*	0.0255	-0.0281	0.0373	0.0503	-0.0280	0.0225	0.0634	0.6378
	(9.0)	(-131.5)	(115.5)	(238.1)	(-106.3)	(197.6)	(136.0)	
Large	0.0214	-0.0478	0.0497	0.0478	-0.0385	0.0317	0.0807	0.7249
	(4.3)	(-482.2)	(143)	(89.7)	(-56.3)	(307.8)	(97.9)	
College x Small	0.0669	-0.0098	0.0055	0.0622	0.0286	-0.0135	-0.0399	0.8955
	(41.3)	(-62.2)	(31.6)	(176.4)	(1366.1)	(-9.2)	(-256.6)	
College x *(Mz+Large)*	0.0052	0.0385	0.0144	-0.0433	0.0585	-0.0064	0.0045	0.9587
	(1.3)	(115.3)	(291.3)	(-27.8)	(83.8)	(-13.9)	(1.7)	
Junior college x Small	0.0237	0.0379	-0.0222	0.0278	-0.0117	-0.0055	-0.0135	0.8532
	(12.1)	(160.7)	(-413.6)	(445.2)	(-10.6)	(44.5)	(-21.0)	
Junior college x *(Mz +Large)*	-0.0335	0.0390	0.1027	-0.0394	0.0522	-0.0218	0.0103	0.9722
	(-9.4)	(54.0)	(968.4)	(-100.8)	(88.3)	(-87.7)	(14.2)	
Medium skilled x Small	0.0063	0.0216	-0.0204	-0.0058	0.0217	-0.0064	0.0101	0.9025
	(31.2)	(51.2)	(-94.7)	(-12.7)	(56.1)	(-399.6)	(208.8)	
Medium skilled x *(Mz +Large)*	-0.0101	0.0216	-0.0142	-0.0344	0.0430	-0.0072	-0.0401	0.9556
	(-9.9)	(32.8)	(-22.4)	(-50.0)	(572.5)	(-23.3)	(-395.4)	
σ	0.1078	(13.5)						

Note: t-values in brackets. Serial correlation (S.C.) with the differences calculated in 5.1.

TABLE 7

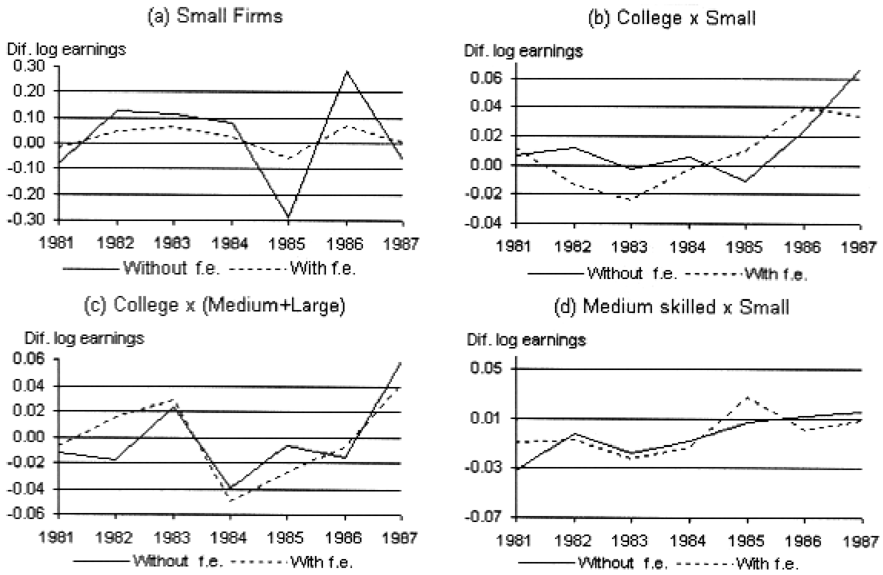
Estimation of the differences of the returns for Sector 8 controlling for unobserved heterogeneity. $N = 873$, averaged censoring of 30.67%

Sector 8	1981	1982	1983	1984	1985	1986	1987	S.C.
Small	-0.0093	-0.0154	0.0537	-0.0076	0.0034	0.0013	0.0491	0.8221
	(-2.8)	(-41.4)	(86.3)	(-22.2)	(27.0)	(1.9)	(37.4)	
Medium-sized *(Mz)*	-0.0081	0.0111	0.0551	0.0156	0.0291	0.0021	0.0268	0.3051
	(-0.6)	(12.1)	(18.5)	(1.5)	(3.2)	(0.7)	(7.1)	
Large	-0.0372	0.0185	0.0754	0.0232	0.0425	-0.0056	-0.0531	-0.1196
	(-3.0)	(42.4)	(50.8)	(2.9)	(5.0)	(-2.6)	(-5.6)	
College x Small	-0.0225	0.0381	0.0259	0.0206	0.0385	0.0055	0.0011	0.5554
	(-5.4)	(12.0)	(3.8)	(38.2)	(31.4)	(0.8)	(4.1)	
College x *(Mz+Large)*	0.0328	0.0187	-0.0144	0.0714	-0.0282	0.0901	0.0509	0.6984
	(3.3)	(23.1)	(-2.9)	(79.5)	(-7.2)	(15.4)	(12.2)	
Junior college x Small	-0.0038	0.2357	-0.1355	0.1242	0.0315	0.0012	-0.0034	0.7534
	(-1.4)	(117.1)	(-18.4)	(228.4)	(7.2)	(11.3)	(-4.0)	
Junior college x *(Mz +Large)*	0.0260	-0.0044	-0.0193	-0.0706	0.1207	-0.0030	0.0415	0.7546
	(-1.8)	(-1.9)	(-10.8)	(-7.9)	(9.2)	(-0.5)	(5.3)	
Medium skilled x Small	0.0101	0.0167	-0.0154	0.0247	-0.0151	0.0235	-0.0037	0.7472
	(6.3)	(10.3)	(-10.4)	(274.0)	(-121.8)	(633.3)	(-2.8)	
Medium skilled x *(Mz +Large)*	0.0284	-0.0217	-0.0288	-0.0220	-0.0319	0.0290	0.0404	0.2530
	(4.2)	(-28.8)	(-65.1)	(-1.9)	(-3.6)	(10.8)	(7.2)	
σ	0.1064	(12.6)						

Note: t-values in brackets. Serial correlation (S.C.) with the differences calculated in 5.1.

The main facts to be emphasized are, on the one hand, the low variability of the differences of the estimated returns over time when we control for unobserved heterogeneity, and, on the other hand, the high correlation between the two estimates. Since this analysis cannot determine the evolution of the returns when we introduce individual effects, we can only carry out a comparison in differences.

FIGURE 7
Comparison between the differences of the returns for Sector 1 (Construction) with fixed effects and without fixed effects



6. Conclusions

In this study I have proposed a method that brings back the likelihood-based analysis of fixed effects with panel data, and I have carried out a specific analysis for the Tobit model with individual effects and under the assumption of normality of the errors. To do so I have developed the *modified profile likelihood*. Two ideas constitute the basis of this function: orthogonalization and an approximation to the conditional log-likelihood by sufficient statistics of the fixed effects. It can be shown that this new function reduces the order of its scores with respect to T. Thus, we expect an important improvement of the biases of the estimators derived from the modified profile likelihood, when T becomes larger. For the Tobit model I have shown when is it possible to recover the orthogonal parameters. Also, carrying out a Monte Carlo study, I have proved that this reduction of the biases is very important for the estimators of σ, although this improvement is smaller for the estimators of β. The reason can be found in the linear model, which is latent in the Tobit model: with this approach, what we obtain for the linear case is the Within-Group estimator (ML estimator). Thus, for the linear model the proposed method does not generate new estimators of β. For this reason, the result obtained for the non-linear model could be seen only as an application of a continuity argument.

Furthermore, I have applied the developed method to obtain estimates of the earnings returns to some characteristics of workers and firms in Spain during the 1980s. I have controlled for unobserved heterogeneity and compared these results to the estimators obtained with a cross-sectional analysis for every year and sector. The results, limited by the fact that I estimate the differences between returns with respect to the previous year, show a strong correlation between both kinds of estimators, and also, a lower variability of those obtained when I control for unobserved heterogeneity.

Appendix: Proofs

Proofs of the following results (Lemma and Corollary 1 and 2) are not shown here. Detailed proofs can be sent on request.

LEMMA *We have that:*

i) $\Lambda(e) = -e - \frac{1}{e} + o(\frac{1}{e})$ *when* $e \to -\infty$.

ii) $0 \leq M(e) \leq 1$ *for all* e.

iii) $\lim_{e \to -\infty} M(e) = 1$.

iv) $\int_{-\infty}^{+\infty} H(e)de = 1$.

v) $H(e) \geq 0$ *for all* e.

Moreover, from this we can obtain two interesting results[13] :

[13] From these results we have some striking inequalities:

$$\begin{aligned} \phi(e) + e\Phi(e) &\geq 0 \\ [\phi(e) + e\Phi(e)]\,[2\phi(e) + e\Phi(e)] &\geq \Phi^2(e) \geq \phi(e)\,[\phi(e) + e\Phi(e)] \\ e\phi(e) + \Phi(e) + e^2\Phi(e) &\geq 0 \end{aligned}$$

COROLLARY 1 *$1 - M(e)$ is a distribution function.*

COROLLARY 2 *$F(e) = 1 - \varphi(e)$ is a distribution function.*

PROPOSITION 1 *For the Tobit model with fixed effects and standard and independently distributed errors, an effect orthogonal to η_i, λ_i, for fixed σ is given by*

$$\lambda_i = \overline{x}_i'\beta + \eta_i + \frac{\sigma}{T}\sum_{t=1}^{T}\int_{-\infty}^{e_{it}} F(e)de$$

where $F(e) = 1 - \varphi(e)$ is a distribution function.

Proof:

From (40) using (30), we have

$$\frac{\partial}{\partial\beta} log\left|\frac{\partial\eta_i}{\partial\lambda_i}\right| = \frac{1}{\sigma}$$

$$\frac{\left(\sum_{t=1}^{T}\varphi'(e_{it})x_{it}\right)\left(\sum_{t=1}^{T}\varphi(e_{it})\right) - \left(\sum_{t=1}^{T}\varphi(e_{it})x_{it}\right)\left(\sum_{t=1}^{T}\varphi'(e_{it})\right)}{\left(\sum_{t=1}^{T}\varphi(e_{it})\right)^2}$$

$$= \frac{1}{\sigma}\left(\frac{\sum_{t=1}^{T}\varphi'(e_{it})x_{it}}{\sum_{t=1}^{T}\varphi(e_{it})} - \frac{\sum_{t=1}^{T}\varphi(e_{it})x_{it}}{\sum_{t=1}^{T}\varphi(e_{it})}\frac{\sum_{t=1}^{T}\varphi'(e_{it})}{\sum_{t=1}^{T}\varphi(e_{it})}\right)$$

$$= \frac{1}{\sigma}\left(\frac{\sum_{t=1}^{T}\varphi'(e_{it})x_{it}}{\sum_{t=1}^{T}\varphi(e_{it})} + \frac{\partial\eta_i}{\partial\beta}\frac{\sum_{t=1}^{T}\varphi'(e_{it})}{\sum_{t=1}^{T}\varphi(e_{it})}\right)$$

$$= \frac{1}{\sigma}\frac{\sum_{t=1}^{T}\varphi'(e_{it})\left(x_{it} + \frac{\partial\eta_i}{\partial\beta}\right)}{\sum_{t=1}^{T}\varphi(e_{it})} = -\frac{\partial}{\partial\beta}\log\sum_{t=1}^{T}\varphi(e_{it})$$

$$= \frac{\partial}{\partial\beta} log\frac{1}{\sum_{t=1}^{T}\varphi(e_{it})} \qquad [A.1]$$

If global orthogonality is possible for σ known, integrating, it turns out that

$$\frac{\partial\eta_i}{\partial\lambda_i} = \frac{1}{\sum_{t=1}^{T}\varphi(e_{it})} \qquad [A.2]$$

This expression is positive because $\varphi(e) = 1 - F(e) = \Phi(e)\left[1 - M(e)\right]$, and we know form Lemma that $M(e) \leq 1$ for all real value e.

Hence, for a certain real value a, we have:

$$\begin{aligned}\lambda_i &= -\sigma \sum_{t=1}^{T} \int_a^{e_{it}} \varphi(e)\ de = -\sigma \sum_{t=1}^{T} \int_a^{e_{it}} [1-F(e)]\, de \\ &= T\sigma a + T\ \overline{x}_i'\beta + T\ \eta_i - Tc + \sigma \sum_{t=1}^{T} \int_a^{e_{it}} F(e)de \qquad \text{[A.3]}\end{aligned}$$

where $\overline{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$. Also, removing constants we have

$$\lambda_i = T\ \overline{x}_i'\beta + T\ \eta_i + \sigma \sum_{t=1}^{T} \int_a^{e_{it}} F(e)de \qquad \text{[A.4]}$$

and, we can choose the orthogonal reparametrization given by

$$\lambda_i = \overline{x}_i'\beta + \eta_i + \frac{\sigma}{T} \sum_{t=1}^{T} \int_a^{e_{it}} F(e)de \qquad \text{[A.5]}$$

And also,[14]

$$\lambda_i = \overline{x}_i'\beta + \eta_i + \frac{\sigma}{T} \sum_{t=1}^{T} \int_{-\infty}^{e_{it}} F(e)de. \qquad \text{[A.6]}$$

■

PROPOSITION 2 *When σ is unknown, global orthogonality can be achieved only for the linear case.*

Proof:

From [41] using [32], we have

$$\begin{aligned}&\frac{\partial}{\partial\sigma^2} log \left|\frac{\partial\eta_i}{\partial\lambda_i}\right| \\ &= \frac{1}{2\sigma^2} \frac{\left(\sum_{t=1}^{T}\xi'(e_{it})\right)\left(\sum_{t=1}^{T}\varphi(e_{it})\right) - \left(\sum_{t=1}^{T}\xi(e_{it})\right)\left(\sum_{t=1}^{T}\varphi'(e_{it})\right)}{\left(\sum_{t=1}^{T}\varphi(e_{it})\right)^2} \\ &= \frac{1}{2\sigma^2}\frac{\sum_{t=1}^{T}\xi'(e_{it})}{\sum_{t=1}^{T}\varphi(e_{it})} - \frac{1}{2\sigma^2}\frac{\sum_{t=1}^{T}\xi(e_{it})}{\sum_{t=1}^{T}\varphi(e_{it})}\frac{\sum_{t=1}^{T}\varphi'(e_{it})}{\sum_{t=1}^{T}\varphi(e_{it})} \\ &= \frac{1}{2\sigma^2}\frac{\sum_{t=1}^{T}\xi'(e_{it})}{\sum_{t=1}^{T}\varphi(e_{it})} + \frac{1}{\sigma}\frac{\partial\eta_i}{\partial\sigma^2}\frac{\sum_{t=1}^{T}\varphi'(e_{it})}{\sum_{t=1}^{T}\varphi(e_{it})}. \qquad \text{[A.7]}\end{aligned}$$

[14] Note that if σ is unknown we have that $\frac{\partial\eta_i}{\partial\lambda_i} = \frac{k(\sigma)}{\sum_{t=1}^{T}\varphi(e_{it})}$. Or, $-log\left|\frac{\partial\eta_i}{\partial\lambda_i}\right| = log\sum_{t=1}^{T}\varphi(e_{it}) + K(\sigma)$.

Conversely,

$$\frac{\partial}{\partial \sigma^2} log \frac{1}{\sum_{t=1}^{T} \varphi(e_{it})} = \frac{1}{2\sigma^2} \frac{\sum_{t=1}^{T} \varphi'(e_{it}) e_{it}}{\sum_{t=1}^{T} \varphi(e_{it})} + \frac{1}{\sigma} \frac{\partial \eta_i}{\partial \sigma^2} \frac{\sum_{t=1}^{T} \varphi'(e_{it})}{\sum_{t=1}^{T} \varphi(e_{it})}. \quad [A.8]$$

Then,

$$\frac{\partial}{\partial \sigma^2} log \left| \frac{\partial \eta_i}{\partial \lambda_i} \right| = \frac{\partial}{\partial \sigma^2} log \frac{1}{\sum_{t=1}^{T} \varphi(e_{it})} + \frac{1}{2\sigma^2} \frac{\sum_{t=1}^{T} \left[\xi'(e_{it}) - \varphi'(e_{it}) e_{it} \right]}{\sum_{t=1}^{T} \varphi(e_{it})}. \quad [A.9]$$

>From [16] and [18], it is easy to check that

$$\xi'(e) - \varphi'(e) e = \Phi(e) M(e).$$

So,

$$\frac{\partial}{\partial \sigma^2} log \left| \frac{\partial \eta_i}{\partial \lambda_i} \right| = \frac{\partial}{\partial \sigma^2} log \frac{1}{\sum_{t=1}^{T} \varphi(e_{it})} + \frac{1}{2\sigma^2} \frac{\sum_{t=1}^{T} \Phi(e_{it}) M(e_{it})}{\sum_{t=1}^{T} \varphi(e_{it})}. \quad [A.10]$$

From [A2.29] $\frac{\partial}{\partial \beta} log \left| \frac{\partial \eta_i}{\partial \lambda_i} \right| = \frac{\partial}{\partial \beta} log \frac{1}{\sum_{t=1}^{T} \varphi(e_{it})}$.

Given these results, global orthogonality can be achived if and only if $\sum_{t=1}^{T} \Phi(e_{it}) M(e_{it}) = 0$, what is only possible for the linear case $(c \to -\infty)$.■

References

Amemiya, T. (1985): "Advanced Econometrics", Cambridge, Harvard University Press.

Andersen, E. (1970): "Asymptotic properties of conditional maximum likelihood estimators", *Journal of the Royal Statistical Society,* Series B, 32, pp. 283-301.

Arellano, M. and Honoré, B. (2001): "Panel data models: Some recent developments", in J. Heckman and E. Leamer (eds.): *Handbook of Econometrics,* Volume 5.

Arellano, M. (2003a): "Discrete choices with panel data", *Investigaciones Económicas* 27, pp. 423-458.

Arellano, M. (2003b): "Panel Data Econometrics", Oxford University Press, Advanced Texts in Econometrics.

Bover, O., Bentolila, S. and Arellano, M. (2002): "The distribution of earnings in Spain during the 1980s: the effects of skill, unemployment, and union power", in D. Cohen, T. Piketty y G. Saint-Paul (eds.), *The Economics of Rising Inequalities,* CEPR and Oxford University Press, pp. 3-53.

Cox, D. R. and Reid, N. (1987): "Parameter orthogonality and approximate conditional inference (with discussion)", *Journal of the Royal Statistical Society,* Series B, 49, pp. 1-39.

Heckman, J. J. and MaCurdy, T. E. (1980): "A life cycle model of female labour supply", *Review of Economic Studies* 47, pp. 47-74.

Honoré, B. E. (1992): "Trimmed LAD and least squares estimation of truncated an censored regression models with fixed effects", *Econometrica* 60, pp. 533-565.

Lancaster, T. (1997): "Orthogonal parameters and panel data", Working Paper 97-32, Department of Economics, Brown University.

Liang, K. (1987): "Estimating functions and approximate conditional likelihood", *Biometrika* 74, pp. 695-702.

Neyman, J. and Scott, E. L. (1948): "Consistent estimates based on partially consistent observations", *Econometrica* 16, pp. 1-32.

Resumen

El presente trabajo parte del método de ortogonalización propuesto por Cox y Reid para aplicarlo al modelo Tobit con datos de panel y efectos fijos. Neyman y Scott mostraron que, en general, el estimador máximo verosímil es inconsistente (el problema de los parámetros incidentales). La metodología aquí expuesta recupera el uso de la función de log-verosimilitud para resolver este problema explotando, para ello, la dimensión temporal de los datos de panel. Para el modelo Tobit se muestra cuándo es posible recuperar los parámetros ortogonales y se estudian las características de los estimadores obtenidos mediante métodos de simulación. Asimismo, se ha realizado una aplicación con ecuaciones de salarios.

Palabras clave: Parámetros ortogonales, verosimilitud modificada, datos de panel, modelo Tobit.

Recepción del original, julio de 2002
Versión final, octubre de 2004