Simulation-based estimation of Tobit model with random effects

Giorgio Calzolari and Laura Magazzini and Fabrizia Mealli

Universita' di Firenze, Italy., S.Anna School of Advanced Studies, Pisa, Italy

2001

Simulation-Based Estimation of Tobit Model with Random Effects[1]

Giorgio Calzolari[a], Laura Magazzini[b] and Fabrizia Mealli[a]

[a] Università di Firenze, Dipartimento di Statistica "G. Parenti", Viale Morgagni 59, 50134 Firenze, Italy

[b] S. Anna School of Advanced Studies, Via Carducci 40, 56127 Pisa, Italy

Abstract. The performance of alternative simulation-based estimators for a panel data Tobit model (censored regression) with random effects is evalueted with Monte Carlo experiments. Bias and efficiency of the methods is discussed. An example of application is provided on a model of female labour supply.

1 Introduction

The estimation of limited dependent variable panel data models usually involves objective functions in which integrals appear without a closed form solution: this is the case of the panel data Tobit model with random effects. Recently, simulation methods have shown to be useful in the inference process, as they offer methods to approximate such integrals (Laroque, Salanie, 1989; Gouriéroux, Monfort, 1991, 1993; Hajivassiliou, McFadden, 1998; Mealli, Rampichini, 1999; Inkmann, 2000). Although the asymptotic performances of such methods are known and their application has been successfully undertaken, more precise ideas on their finite sample performance and computational efficiency is still needed. In this paper we propose to use the method of indirect inference, using different auxiliary models, and the simulated maximum likelihood to estimate these models. We use a panel data Tobit model with a simple correlation structure in the unobservables (i.e. a one-factor structure), but the model could be easily extended. Using both simulated and real data, we show the perfomances of the proposed methods in finite samples.

[1]Financial support from MURST through projects "Stochastic Models for Dependent Data" and "Evaluating Quality, Effectiveness and Efficiency in Individual Services" is gratefully acknowledged. We are also grateful to Marco Barnabani, Fabrizio Cipollini, and Carla Rampichini for helpful comments and suggestions, but retain full responsibility for the contents of this paper.

The application on real data is concerned with a model of female labour supply.

1.1 Tobit Model

The Tobit model[2], or censored regression model, is used when a large number of observations on the dependent variable assumes the value of zero: that is the case, for instance, of the expenditure on durable goods or the number of hours at work for a certain person (the number of hours at work is set to zero when a person is not employed). The data generating process can be described as follows:

$$y_{it} = \begin{cases} 0 & \text{if } y_{it}^* \leq 0 \\ y_{it}^* & \text{if } y_{it}^* > 0 \end{cases}$$

where $y_{it}^* = x_{it}'\beta + \varepsilon_{it}$ is observed only if strictly positive; x_{it} is a vector of exogenous variables and it is assumed that $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$ i.i.d. and independent from $x_{it}, i = 1, ..., N; t = 1, ..., T$.

The probability density function for the observed y is (e.g. Amemiya, 1974):

$$f(y_{it} \mid x_{it}; \theta) = \begin{cases} 0 & \text{if } y_{it} < 0 \\ \Phi(-x_{it}'\beta/\sigma_\varepsilon) & \text{if } y_{it} = 0 \\ \phi((y_{it} - x_{it}'\beta)/\sigma_\varepsilon) & \text{if } y_{it} > 0 \end{cases}$$

Φ and ϕ being the cumulated distribution and the probability density functions of the normal distribution, respectively.

When we consider a model for panel data, the error term ε_{it} can be decomposed into:

$$\varepsilon_{it} = \alpha_i + \lambda_t + u_{it},$$

where α_i is the individual effect (representing all the unobservable characteristics specific to the unit i that are assumed constant over time), λ_t is the time effect (representing all the unobservable characteristics of time period t, constant for all the cross-sectional units in the sample) and u_{it} is a random term that varies over time and individuals. Moreover it is assumed that u_{it} are uncorrelated over time (however it is possible to generalize the model and consider, for instance, random terms which are correlated over time or lagged values of the dependent variable).

In the following sections it will be assumed that $\lambda_t = 0$ for every t. Models that consider also the effects of variables which are constant over the cross-sectional units but vary over time (that is $\lambda_t \neq 0$) are not used in practice.

[2]Tobin (1958) proposed a censored regression model to analyse consumption of durables. For its characteristics, close to those of a probit model, the name "Tobin's probit" was used at first, then converted in "Tobit" by Goldberger (1964).

The loss of degrees of freedom is, in fact, too high and usually a more general model than the formulation with dummy variables is used in case we want to include a time effect in the model.

The terms α_i can be treated as fixed parameters to be estimated or as random variables with a known distribution whose parameters have to be estimated together with the other parameters in the model.

1.1.1 Fixed Effects Tobit Model

The fixed effects Tobit model can be written as:

$$y_{it}^* = \alpha_i + x_{it}'\beta + u_{it},\ u_{it} \sim IN(0, \sigma_u^2)$$
$$y_{it} = \begin{cases} y_{it}^* & \text{if } y_{it}^* > 0 \\ 0 & \text{otherwise.} \end{cases}$$

The log-likelihood function is:

$$\log L = \sum_{y_{it}=0} \log \Phi \left(\frac{-\alpha_i - x_{it}'\beta}{\sigma_u} \right) + \sum_{y_{it}>0} \left\{ -\frac{1}{2} \log \sigma_u^2 - \frac{1}{2\sigma_u^2} \left(y_{it} - \alpha_i - x_{it}'\beta \right)^2 \right\}.$$

When the number of observations on each cross-sectional unit is fixed (time dimension), it is not possible to get a consistent estimate of the fixed effect α_i. This problem does not affect the estimation of the β parameter in a linear model, but if the observation on the dependent variable is censored, it is not possible to device a consistent estimator of β and σ_u^2 (Maddala, 1987).

Heckman and MaCurdy (1980) applied the fixed effects Tobit model for the estimation of a female labour supply model. They argued that a fixed effect specification of the model is the most appropriate, because it is not possible to assume the independence of α_i and x_{it}. They derived maximum likelihood estimates of the parameters in the model applying an iterative method. They were aware of their inconsistency, but they argued that, from a practical point of view, this is not a problem when the model does not contain lagged dependent variables. This statement was based on the results obtained by Heckman (1981), who studied the estimator bias in the fixed effects probit model. Heckman and MaCurdy did not perform a simulation study using the Tobit model, but they supposed that the results obtained on the probit model could be extended to the Tobit one.

Their conjecture was confirmed by Honoré (1993): he performed a simulation study to evaluate the bias of the maximum likelihood estimator of a fixed effects Tobit model with lagged dependent variable. Honoré's results

coincide with Heckman's. Moreover, Honoré (1992, 1993) proposes some orthogonality conditions that can be used to construct a GMM estimator[3] of the parameters in the fixed effect Tobit model. Honoré (1992) considers a static model, and in Honoré (1993) an estimator for the model with lagged dependent variables was developed. In his simulation studies, the proposed estimators seemed to have good asymptotic properties. Honoré and Kyriazidou (2000) propose a new class of estimators for the static censored regression model with fixed effects, that rely on weaker assumptions on the transitory error terms.

1.1.2 Random Effects

Tha data generating process can be described as:

$$y_{it} = \max\left\{y_{it}^{*}, 0\right\},$$

where

$$y_{it}^{*} = x_{it}'\beta + \alpha_i + u_{it}.$$

x_{it} is the vector containing the observations on the exogenous variables, α_i represents the individual effect and it is assumed α_i i.i.d. $N(0, \sigma_\alpha^2)$ and u_{it} i.i.d. $N(0, \sigma_u^2)$ independent of α's ($i = 1, \ldots, n$; $t = 1, \ldots, T$).

The equation above can also be written as:

$$y_{it}^{*} = x_{it}'\beta + \sigma_\alpha\alpha_i + \sigma_u u_{it},\ i = 1, \ldots, n;\ t = 1, \ldots, T, \tag{1}$$

where α_i i.i.d. $N(0,1)$ independent from u_{it} i.i.d. $N(0,1)$.

Due to the individual effect $\sigma_\alpha\alpha_i$, the observations on the dependent variable for each individual ($y_{it}, t = 1, \ldots, T$) are correlated. However, conditional on the individual effect α_i, the conditional joint density function can be written as (see Gouriéroux and Monfort, 1993):

$$\begin{aligned} f(y_i \mid x_i, \alpha_i; \theta) \;=\; & \prod_{t:y_{it}>0} \frac{1}{\sigma_u}\phi\left(\frac{y_{it} - x_{it}'\beta - \sigma_\alpha\alpha_i}{\sigma_u}\right) \times \\ & \prod_{t:y_{it}=0} \Phi\left(\frac{-x_{it}'\beta - \sigma_\alpha\alpha_i}{\sigma_u}\right), \end{aligned} \tag{2}$$

where Φ and ϕ are the cumulative and density function of the $N(0,1)$ distribution. α_i is not observable, so (2) cannot be used for inference. To obtain the unconditional likelihood, we have to integrate out the individual effect α_i:

$$f\left(y_i \mid x_i; \theta\right) = \int f\left(y_i \mid x_i, \alpha_i; \theta\right) dP^{\alpha}\left(\alpha\right). \tag{3}$$

[3]The parameters α_i are treated as nuisance parameters.

A possible solution to the integral in (3) is numerical integration. Alternatively, we can approximate the integral (3) by means of replicated simulations, thus obtaining a simulated likelihood function to be maximised in order to obtain a simulated maximum likelihood estimator.

We also propose a set of different simulation-based estimators that do not try to approximate the likelihood, but rather calibrate parameters with the help of an auxiliary model. Gouriéroux, Monfort and Renault (1993) label these methods "Indirect Inference (I.I.)", Gallant and Tauchen (1996) label these methods "Efficient Method of Moments (EMM)". For the models we use in this paper, the last two methods give the same results (just-identified case), and they will be referred to as "Indirect Inference".

2 Simulation Study

We use the methods of indirect inference (I.I.) and simulated maximum likelihood (SML) as feasible ways to estimate the parameters in a random effect Tobit model for panel data. To study the properties of these methods we perform a simulation study and apply the methods to pseudo-observed data, produced by Monte Carlo simulation of the data generating process.

The data generating process can be expressed by:

$$y_{it}^{*} = \beta_0 + \beta_1 x_{it} + \sigma_\alpha \alpha_i + \sigma_u u_{it}, \tag{4}$$

where α_i and u_{it} are independent random variables that have a normal distribution with zero mean and unit variance and y_{it}^{*} indicates the value of the latent variable, not always observable, that rules the observations y_{it}:

$$y_{it} = \max\left\{y_{it}^{*}, 0\right\}. \tag{5}$$

Therefore the parameters vector is:

$$\theta = \left(\beta_0, \beta_1, \sigma_\alpha^2, \sigma_u^2\right).$$

For the simulation study x_{it} are held constant[4] and we assume $\theta = (-10, 2, 1, 1)$, that corresponds to a percentage of censoring of about 40%. As far as the sample of observations is concerned, we use $n = 560, T = 3$ (the same dimensions of the panel discussed in section 3).

A series of GAUSS programs[5] have been developed, which perform all the steps of the estimation procedure. 1000 Monte Carlo replications have been performed for every estimator.

[4] In the first experiment x_{it} are drawn independently from a normal distribution with mean 6 and variance 4. Then they are saved and held constant over all the experiments.

[5] The programs are available upon request from the authors.

2.1 Simulated Maximum Likelihood

We now consider the simulated maximum likelihood estimator (henceforth SML). The log-likelihood function for the panel data Tobit model with random effects can be written in the following form:

$$\frac{1}{n}\sum_{i=1}^{n}\log f\left(y_i \mid x_i;\theta\right)=\frac{1}{n}\sum_{i=1}^{n}\log\int f\left(y_i \mid x_i,\alpha_i;\theta\right)dP^{\alpha}\left(\alpha\right)$$

where

$$f\left(y_i \mid x_i,\alpha_i;\theta\right) = \prod_{t:y_{it}>0}\frac{1}{\sigma_u}\phi\left(\frac{y_{it}-\beta_0-\beta_1 x_{it}-\sigma_\alpha\alpha_i}{\sigma_u}\right)\times \prod_{t:y_{it}=0}\Phi\left(\frac{-\beta_0-\beta_1 x_{it}-\sigma_\alpha\alpha_i}{\sigma_u}\right). \tag{6}$$

Therefore conditioning on α_i, the likelihood function relative to a single observation has a simple closed form. Therefore the simulator can be based on (6).

It is possible to get an approximation of (6) drawing, for each subscript i, S values $\alpha_i^s, s=1,...,S$ from the standard normal distribution and calculating

$$\frac{1}{S}\sum_{s=1}^{S}f\left(y_i \mid x_i,\alpha_i^s;\theta\right).$$

The SML estimator is obtained from the maximization of:

$$\frac{1}{n}\sum_{i=1}^{n}\log\left[\frac{1}{S}\sum_{s=1}^{S}f\left(y_i \mid x_i,\alpha_i^s;\theta\right)\right].$$

The GAUSS program developed to get the simulated maximum likelihood estimator performs the following steps:

1. Generation of the pseudo-observed data.
2. Drawing, for each index i ($i=1,...,n$), of S values α_i^s from the standard normal distribution and simulation of the likelihood function.
3. Maximization of the simulated likelihood (using Newton–Raphson algorithm).

The results are displayed in Table 1.

The estimates of the coefficients β_0 and β_1 are close to the real value of the parameters whatever the value of S (the consistency property is obtained

Table 1:
Simulated Maximum Likelihood: Mean Estimated Parameters (Variances in Parentheses) 1000 Replications

Parameter	S=10	S=30
$\beta_0 = -10$	$-10.01\,(.25 \times 10^{-1})$	$-9.99\,(.26 \times 10^{-1})$
$\beta_1 = 2$	$2.00\,(.47 \times 10^{-3})$	$2.00\,(.48 \times 10^{-3})$
$\sigma_\alpha^2 = 1$	$0.93\,(.16 \times 10^{-1})$	$1.01\,(.11 \times 10^{-1})$
$\sigma_u^2 = 1$	$1.20\,(.41 \times 10^{-2})$	$1.06\,(.32 \times 10^{-2})$

when $S \to \infty$); however when S increases, the variances of the estimates decrease (the maximum likelihood estimate would be obtained when $S \to \infty$). The estimates of the variances σ_α^2 and σ_u^2 are biased. The bias decreases with S: when $S = 30$, the mean values of the estimates are close to the real values of the parameters. When $S = 1$, the average time of execution for one MC replication would be about 15 seconds, but results are unreliable and are not displayed in the table; when $S = 10$, it takes about 22 seconds and when $S = 30$, 1 minute and 8 seconds[6].

We tried to simulate the likelihood function by means of antithetic variates. For every Monte Carlo replication, we simulated $n\frac{S}{2}$ terms α_i^s, $s = 1, ..., \frac{S}{2}$, $i = 1, ..., n$ from the standard normal and we considered the vector

$$\alpha_i^S = \left(\alpha_i^1, ..., \alpha_i^{S/2}, -\alpha_i^1, ..., -\alpha_i^{S/2}\right)'.$$

The results using $S = 10$ are displayed in the Table 2.

Table 2:
Simulated Maximum Likelihood Estimates Using Antithetic Variates

Parameter	S=10	S=10 and a.v.
$\beta_0 = -10$	$-10.01\,(.25 \times 10^{-1})$	$-9.98\,(.21 \times 10^{-1})$
$\beta_1 = 2$	$2.00\,(.47 \times 10^{-3})$	$2.00\,(.39 \times 10^{-3})$
$\sigma_\alpha^2 = 1$	$0.93\,(.16 \times 10^{-1})$	$0.91\,(.15 \times 10^{-1})$
$\sigma_u^2 = 1$	$1.20\,(.41 \times 10^{-2})$	$1.18\,(.37 \times 10^{-2})$

When using antithetic variates to approximate the likelihood function, but

[6]Simulations have been performed on a Pentium II 600 EB MHz.

with a small number of replications (5+5), we obtain a modest reduction in the variances, but the bias of the estimates of σ_α^2 and σ_u^2 is still not negligible.

2.2 Indirect Inference

This method requires an auxiliary model whose estimation must be feasible (and possibly simple). The GAUSS programs for the estimation of the model using the method of Indirect Inference perform the following steps:

1. Generation of a sample of pseudo observed data, using the data generating process defined in (4) and (5).

2. Estimation of the auxiliary parameter $\hat{\beta}$.

3. Generation of a sample of Sn pseudo random effects α_i^s i.i.d. $N(0,1)$ and of S samples of length nT of i.i.d. $N(0,1)$ pseudo random error terms u_{it}^s $(s = 1, ..., S)$, where S is the number of simulations.

4. Generation of the simulated sample $y_{it}^s(\theta)$ simulating the data generating process (the values α_i and u_{it} in the expression are substituted by the simulated values α_i^s and u_{it}^s obtained at step 3).

5. Estimation of the auxiliary parameter $\beta(\theta)$ using the simulated values.

6. The two vectors of parameters $\hat{\beta}$ and $\beta(\theta)$ are compared.
   Given a value[7] ε, if the following condition holds[8]:

$$\left|\hat{\beta}_j - \beta_j(\theta)\right| < \varepsilon \quad j = 1, ..., 4$$

   the estimation procedure stops and the value of θ is the indirect estimator, otherwise the value θ is modified using the following rule:

$$\theta^{j+1} = \theta^j + \left(\hat{\beta} - \beta(\theta)\right),$$

   and a new iteration starts at step 4. As in the considered auxiliary models, the dimension and interpretation of the parameter θ and β are the same, we chose $\hat{\beta}$ as the starting value to implement the algorithm.

The pseudo random values generated at step 3 are kept constant in the experiment. We considered three auxiliary models, and for each one we performed a different number of simulations: with $S = 1$ and $S = 10$.

[7] We chose $\varepsilon = 10^{-2}$ for the experiments of section 2.2.1, 10^{-5} in sections 2.2.2 and 2.2.3.

[8] $\hat{\beta}_j$ and $\beta_j(\theta)$ represent respectively the j-th element in $\hat{\beta}$ and $\beta(\theta)$.

2.2.1 Auxiliary Model 1: Tobit Estimated by Maximum Likelihood

A first auxiliary model is obtained considering the data as if they are a single cross-section ($560 \times 3 = 1680$ observations):

$$\begin{aligned} y_k^* &= \beta_0 + \beta_1 x_k + \varepsilon_k, \qquad k = 1, ..., 1680 \\ y_k &= \max\{0, y_k^*\} \end{aligned}$$

where ε_k have a normal distribution with zero mean and variance σ^2. There is no complication when estimating the parameters using the maximum likelihood method. We obtain an estimate of the two coefficients β_0 and β_1 and the variance σ^2.

As $\varepsilon_k = \sigma_\alpha \alpha_i + \sigma_u u_{it}, k = (i-1) \times 3 + t$, and α_i and u_{it} are uncorrelated variables, we have:

$$\sigma^2 = \sigma_\alpha^2 + \sigma_u^2.$$

Noting that $\text{cov}(\alpha_i + u_{it}, \alpha_i + u_{it'}) = \sigma_\alpha^2$, we can obtain an estimate of the variance of the random term averaging the covariances (or simply the cross products) of residuals at different times, and therefore we can obtain an estimate of the variance of residuals by the difference:

$$\hat{\sigma}_u^2 = \hat{\sigma}^2 - \hat{\sigma}_\alpha^2.$$

Proceeding in this fashion, the results are reported in Table 3.

Table 3:
"Tobit ML" Estimates of the Auxiliary Model

Parameter	Mean	Variance
$\beta_0 = -10$	-10.01	$(.39 \times 10^{-1})$
$\beta_1 = 2$	2.00	$(.60 \times 10^{-3})$
$\sigma_\alpha^2 = 1$	0.47	$(.30 \times 10^{-2})$
$\sigma_u^2 = 1$	1.53	$(.46 \times 10^{-2})$

Reading the table, we see that the estimates of the coefficients β_0 and β_1 seem to have no problems, but the estimates of the two variance terms are surely biased.

We now apply the indirect inference method using the model that has just been described as auxiliary model. The vector of parameters of the auxiliary model is $\beta = (\beta_0, \beta_1, \sigma_\alpha^2, \sigma_u^2)$, and has the same dimension and interpretation of the parameter θ: in the indirect inference context, we say that the model is

Table 4:
Indirect Inference Using Tobit Auxiliary Model Estimated by Maximum Likelihood

Parameter	S=1	S=10	"Tobit ML"
$\beta_0 = -10$	$-9.98\,(.71 \times 10^{-1})$	$-10.00\,(.38 \times 10^{-1})$	$-10.01\,(.39 \times 10^{-1})$
$\beta_1 = 2$	$2.00\,(.14 \times 10^{-2})$	$2.00\,(.65 \times 10^{-3})$	$2.00\,(.75 \times 10^{-3})$
$\sigma_\alpha^2 = 1$	$1.01\,(.29 \times 10^{-1})$	$0.99\,(.18 \times 10^{-1})$	$0.47\,(.30 \times 10^{-2})$
$\sigma_u^2 = 1$	$1.02\,(.19 \times 10^{-1})$	$1.01\,(.11 \times 10^{-1})$	$1.53\,(.46 \times 10^{-2})$

exactly identified. Results for the indirect inference using $S = 1$ and $S = 10$, and results for the auxiliary model are displayed in Table 4.

The estimates of the auxiliary model are clearly biased (column "Tobit ML"). Using the indirect inference estimation procedure, we are able to reduce the bias introduced by the estimation procedure in the auxiliary model. The mean parameter estimates are quite close to the real value of the parameters.

The estimated variances obtained using $S = 1$ are almost twice as those for $S = 10$. This is not surprising: the simulation effect on the covariance matrix is summarized by the term $\left(1 + \frac{1}{S}\right)$ (see Gouriéroux, Monfort and Renault, 1993), therefore increasing the number of simulations, we obtain an increase in the efficiency of the estimates.

Comparing the β_0 and β_1 estimated variances, we see that in the case that the estimator under the auxiliary model is consistent (Tobit ML), adjusting the bias using the indirect inference method leads to an increase in the variance of the estimates due to the new source of randomness introduced through simulation.

We used a Pentium II 600 EB MHz to perform the estimations: the average time to perform a Monte Carlo replication has been 15 seconds in the case $S = 1$ and 55 seconds in the case $S = 10$. All the replications converged when $S = 10$, but we had to discard 2% of the replications[9] in the case $S = 1$.

2.2.2 Auxiliary Model 2: Linear Model Estimated by GLS

We performed the indirect inference procedure using a simpler auxiliary model. We replaced the Tobit model with a linear regression model.

[9]The total number of Monte Carlo replications performed is 1000.

We consider all the observations (both the null and positive values of the dependent variable):

$$y_{it} = \beta_0 + \beta_1 x_{it} + \sigma_\alpha \alpha_i + \sigma_u u_{it}. \tag{7}$$

Therefore the auxiliary parameter vector β is:

$$\beta = \left(\beta_0, \beta_1, \sigma_\alpha^2, \sigma_u^2\right) \tag{8}$$

so it has the same dimension and interpretation of the structural vector of parameters θ.

To estimate the parameters β_0 and β_1 of the auxiliary model we use the generalized least squares estimator (henceforth GLS). This is a two-step method of estimation: at the first stage we obtain an estimate of the variance terms, at the second stage the estimated variances are used to obtain the covariance matrix that is used in the second step. This estimator would be efficient in a linear panel data model with random effects, but it is inconsistent when the observations are censored.

To estimate the terms σ_α^2 and σ_u^2 we use the following procedure (see Greene, 1997, pp. 626-627). The linear model can be written as:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \sigma_\alpha \alpha_i + \sigma_u u_{it}, \tag{9}$$

hence, considering the group means, we have:

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_i + \sigma_\alpha \alpha_i + \sigma_u \bar{u}_i. \tag{10}$$

Therefore:

$$y_{it} - \bar{y}_i = \beta_1 \left(x_{it} - \bar{x}_i\right) + \left(\sigma_u u_{it} - \sigma_u \bar{u}_i\right). \tag{11}$$

It is possibile to get an estimate $\hat{\sigma}_u^2$ of σ_u^2 using the residuals from regression (11).

Consider now the group mean regression. The residuals are:

$$\bar{y}_i - \beta_0 - \beta_1 \bar{x}_i = \sigma_\alpha \alpha_i + \sigma_u \bar{u}_i = \xi_i. \tag{12}$$

The terms ξ_i are independent and their variance is equal to [10]:

$$var[\xi_i] = \sigma_\xi^2 = \sigma_\alpha^2 + \frac{\sigma_u^2}{3}.$$

Therefore it is possible to get an estimate of σ_ξ^2 using the residuals from regression (10).

[10] There are 3 observations for each unit in the sample.

It is simple to device an estimator of the individual effect variance, which is given by:

$$\hat{\sigma}_\alpha^2 = \hat{\sigma}_\xi^2 - \frac{\hat{\sigma}_u^2}{3}.$$

Once we get the estimators of the variances σ_α^2 and σ_u^2, we can use them to construct the covariance matrix V of the observations, which is:

$$V = I_{560} \otimes \Omega,$$

where I_{560} is the identity matrix with dimensions 560×560 and

$$\Omega = \begin{pmatrix} \sigma_u^2 + \sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_u^2 + \sigma_\alpha^2 & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_u^2 + \sigma_\alpha^2 \end{pmatrix}$$

is the covariance matrix relative to a single individual. We then use the GLS estimator to get $\hat{\beta}_0$ and $\hat{\beta}_1$. The results are displayed in the Table 5, where we indicate in parantheses the variance of the estimates.

Table 5:
Indirect Inference Using GLS Estimation of a Linear Auxiliary Model

Parameter	S=1	S=10	GLS
$\beta_0 = -10$	$-9.99\,(.65 \times 10^{-1})$	$-9.97\,(.36 \times 10^{-1})$	$-5.72\,(.36 \times 10^{-2})$
$\beta_1 = 2$	$2.00\,(.12 \times 10^{-2})$	$2.00\,(.66 \times 10^{-3})$	$1.43\,(.16 \times 10^{-3})$
$\sigma_\alpha^2 = 1$	$1.01\,(.39 \times 10^{-1})$	$0.9\,(.20 \times 10^{-1})$	$0.50\,(.47 \times 10^{-2})$
$\sigma_u^2 = 1$	$1.02\,(.26 \times 10^{-1})$	$1.02\,(.14 \times 10^{-1})$	$1.79\,(.39 \times 10^{-2})$

The GLS estimates of the coefficients are biased, but applying the indirect inference method we get means of the estimates that are very close to the real values of the parameters.

Comparing the results in Table 5 with the ones in Table 4 we see that, surprisingly, the indirect estimator does not perform better when estimation of the auxiliary model is performed with a "good" method (maximum likelihood).

The time of execution of the algorithm is extremely reduced from the previous methods: less than 1 second in the case $S = 1$ and about 1.5 seconds in case $S = 10$. All the replications converged.

2.2.3 Auxiliary Model 3: Linear Model Estimated by OLS

We tried to substitute the GLS estimator with an even simpler ordinary least squares estimator (OLS) to get an estimate of the two coefficients β_0 and β_1. To estimate the variances we used the procedure described in section 2.2.1. The results are displayed in Table 6.

Table 6:
Indirect Inference Using OLS Estimation of a Linear Auxiliary Model

Parameter	S=1	S=10	OLS
$\beta_0 = -10$	$-9.98\,(.75 \times 10^{-1})$	$-9.99\,(.37 \times 10^{-1})$	$-5.74\,(.42 \times 10^{-2})$
$\beta_1 = 2$	$2.00\,(.12 \times 10^{-2})$	$2.00\,(.71 \times 10^{-3})$	$1.44\,(.19 \times 10^{-3})$
$\sigma_a^2 = 1$	$1.01\,(.36 \times 10^{-1})$	$1.00\,(.20 \times 10^{-1})$	$0.51\,(.44 \times 10^{-2})$
$\sigma_u^2 = 1$	$1.02\,(.24 \times 10^{-1})$	$1.02\,(.13 \times 10^{-1})$	$1.79\,(.35 \times 10^{-2})$

Let us compare the OLS and GLS estimators. The bias is about the same in estimating the two coefficients β_0 and β_1. The variance of the OLS estimator is higher than the GLS one, for the parameters of the auxiliary model; nevertheless GLS does not increase efficiency for the parameters of the structural model, when the indirect estimator has come to convergence.

The execution time of the algorithm are lower: less than half a second in the $S = 1$ case[11], 1 second for $S = 10$. No replication has been discarded.

2.3 Comparing the Methods

In this section we compare the results obtained with the Indirect Inference and Simulated Maximum Likelihood methods. We consider $S = 30$ for the SML, because the other estimators are not acceptable, and $S = 10$ for the I.I. method (if we consider higher S, we obtain variances that are not significantly lower).

The SML method produces estimates that are more efficient than the Indirect estimates. For example, the variance of the first coefficient's estimate is, for SML, about 30% smaller than for I.I. using the OLS auxiliary model (we consider here $S = 10$). The drawback is that SML is about 70 times slower.

[11] A Fortran 77 program has also been developed for this algorithm: execution time is reduced approximately by a factor 10.

I.I. with the Tobit auxiliary model estimated by maximum likelihood is not worthwhile, because it is less efficient than SML, and there is no significant reduction of computational time. If computational time must be saved, I.I. using OLS estimation of a linear auxiliary model should be recommended; it is almost as efficient as I.I. with Tobit ML, but is about 50 times faster.

3 An Application to Women's Labour Market

We applied the methods described in the previous section to a sub-sample of the data drawn from a survey performed by Banca d'Italia named "Indagine sui Bilanci delle Famiglie Italiane" (Survey of Italian Households). The units in the sample are observed at three different points in time, i.e. 1989, 1991 and 1993.

The sub-sample we consider consists of married women aged between 14 and 55 years in 1989. We included in our sample only those women with observation in all the three periods of time[12]. As a result, our sample contains 560 units observed over three years (1989, 1991, 1993).

For each woman in the sample, we observe the average number of hours worked per week (dependent variable in the model) and a series of socio-demographic characteristics like the age of the woman, the place of residence, the educational background, the number of children in the family and the number of children aged less than 6, the family income and the woman's income.

Using these data, it has been possible to construct the following variables, that have been included in the model as explanatory variables:

- SOUTH and MIDDLE: dummy variables indicating the residence zone; the reference category is living in the Northern Italy (SOUTH = 1 if the woman lives in Southern Italy, and 0 otherwise; MIDDLE = 1 if the woman lives in Central Italy, and 0 otherwise);
- AGE: age of the woman in years;
- SECONDARY, HIGH and UNIVERSITY: dummy variables indicating the highest degree of education attained; the reference category is primary school or no degree;

[12]The data had already been used for a panel data analysis: the data set was built by Zammarano (1995) and used for her thesis.

- CHILDREN: dummy variable indicating the presence of children in the family (CHILDREN = 0 if no child is present, CHILDREN = 1 otherwise);

- KIDS: dummy variable indicating the presence of children aged less than 6 years (KIDS = 1 if there is at least a child aged less than 6 years, and 0 otherwise);

- INCOME: this variable contains the family income, measured in millions of Italian lire, minus the salary earned by the woman (since the salary was measured at current prices, we used the consumer price index to deflate it).

3.1 Description of the Sample

The average age of the women in the sample is 39.5 years; 39.46% of the women in the sample live in the North of Italy, 17.68% in the Middle and 42.86% in the South. The percentage of female unemployment in the South of Italy is higher than the one in the Middle or in the North as it is shown by Table 7.

Table 7:
Percentage of Female Employment

	0 hours	**1 to 29**	**30 or more**
		1989	
North	54.75%	8.14%	37.10%
Middle	55.56%	7.07%	37.37%
South	71.25%	7.08%	21.67%
		1991	
North	54.30%	9.95%	35.75%
Middle	54.55%	7.07%	38.38%
South	68.75%	7.92%	23.33%
		1993	
North	50.23%	13.57%	36.20%
Middle	59.60%	11.11%	29.29%
South	65.42%	9.58%	25.00%

The percentage of censored data is more than 50%. The out of labour force rate in the sample is 61.96% in 1989, it decreases to 60.54% in 1991 and to 58.39% in 1993.

Table 8 shows the composition of the sample with regard to the educational background. The highest percentage is represented by the women who finished only Primary School or who didn't get any degree. The percentage of the women who got a University degree or a higher degree is much lower than the other categories.

Table 8:
Educational Background in the Sample

Educational Background	1989	1991	1993
Primary School or no degree	36.96%	36.25%	33.93%
Secondary School	29.29%	28.93%	30.18%
High School	27.14%	27.68%	28.39%
University Degree or more	6.61%	7.14%	7.50%
Total	100%	100%	100%

Table 9 displays the average number of hours at work per week within each category.

Table 9:
Average Number of Hours at Work per Week

Educational Background	1989	1991	1993
Primary School or no degree	8.04	8.75	8.27
Secondary School	12.56	12.73	14.34
High School	18.57	19.30	20.44
University degree or more	29.91	26.80	24.05

Consider now the number of children in the family. The families with at least one child are 89.35% of the families in the sample. The percentage of families with at least one child aged less than six is decreasing every year: 30.18% in 1989, 24.11% in 1991 and 20.36% in 1993.

The average net family income, measured at costant prices, is about 29 millions lire in 1989, about 33 millions lire in 1991, and it raised to 34 millions in 1993: the raise in the average salary is probably due to the increased

lenght of service of the women and to the fact that the out of labour force rate decreases during the time of observation.

3.2 Estimates

In this section we display the estimates obtained from the data described in the previous section. The model can be described by the following equations:

$$\begin{cases} y_{it}^* = & \beta_0 + \beta_1 SOUTH_{it} + \beta_2 MIDDLE_{it} + \beta_3 AGE_{it} + \beta_4 SECONDARY_{it} + \\ & + \beta_5 HIGH_{it} + \beta_6 UNIVERSITY_{it} + \beta_7 CHILDREN_{it} + \beta_8 KIDS_{it} \\ & + \beta_9 INCOME_{it} + \sigma_\alpha \alpha_i + \sigma_u u_{it}, \\ y_{it} = & \max\{y_{it}^*, 0\} \end{cases}$$

The dependent variable is the average number of hours worked per week. We assume that the variables included in the model are strictly exogenous.

When using Indirect Inference to estimate the model parameters, we chose $S = 10$. When using SML we chose $S = 200$ (100 replications +100 with antithetic variates[13]). Table 10 displays the estimates[14] obtained using the methods of Indirect Inference using OLS as auxiliary model (I.I.-OLS) and of the Simulated Maximum Likelihood (SML).

We have to pay attention for their interpretation. Due to the censoring in the data, the Tobit model estimated coefficients do not represent the increase or decrease in the dependent variable corresponding to a unit increase in the value of the explanatory variables. However the coefficients can be interpreted in terms of their effect on the variable y^*. Actually the latent variable marginal effects are given by:

$$\frac{\partial E\left[y_{it}^* \mid x_{it}\right]}{\partial x_{it}} = \beta$$

(here β is the vector containing the intercept and all the coefficients).

The marginal effects on the observed variable y are indeed obtained as:

$$\frac{\partial E\left[y_{it} \mid x_{it}\right]}{\partial x_{it}} = \beta \Phi\left(\frac{x_{it}'\beta}{\sqrt{\sigma_\alpha^2 + \sigma_u^2}}\right).$$

[13]For this computation we used Fortran77. Execution time was about 15 minutes on a PC Pentium II 600 EB MHz.

[14]To calculate the standard errors of the estimates we used the procedure described in Calzolari, Di Iorio and Fiorentini (1999). To implement this procedure we had to calculate the covariance matrix of the auxiliary model estimates: we used the properties of the GMM estimators (Greene, 1997, pp. 519-526).

Table 10:
Parameter Estimates and Standard Errors

	I.I.–OLS		SML	
Variable	**Estimates**	**Std. Errors**	**Estimates**	**Std. Errors**
Constant	31.19	6.64	5.71	10.88
SOUTH	-11.94	2.46	-14.72	4.00
MIDDLE	-3.34	2.55	-2.49	4.64
AGE	-0.71	0.16	-0.52	0.22
SECONDARY	6.67	3.05	10.39	3.69
HIGH	16.35	3.00	21.74	3.94
UNIVERSITY	26.58	3.43	40.51	5.56
CHILDREN	3.21	3.56	1.29	3.55
KIDS	-7.14	2.66	-0.03	2.97
INCOME	-0.04	0.07	-0.04	0.03
σ_a^2	589.9	92.1	1290	133.1
σ_u^2	168.5	20.4	252.2	17.2

Obviously it makes sense to talk about marginal effects only for the variables AGE and INCOME, which are the continuous explanatory variables in the model. When considering a dummy variable, its coefficient measures the variation in the latent variable when switching from the reference category to the category described by the dummy variable considered.

In order to study the dummy variable effect, we calculated the expected number of hours at work per week of people in different categories (keeping fixed the value of the other variables). Since both the expected value of the variable "average number of hours at work per week" and its marginal effect depend on the value of the explanatory variables, we choose as a benchmark the "average woman" in the sample. This is the woman that presents the mean value of the continuous explanatory variables and the mode for all the other variables. The "average woman" lives in the South of Italy, she is 41.5 years old, finished primary school, has at least one child, but no one aged less than six years and has a net family income of about 32 millions lire. The marginal effects, calculated on the basis of I.I.-OLS estimates, are displayed in Table 11. Increasing the age, the number of hours worked per week decreases and the same is true when considering the net family income.

The effects of the dummy variables are displayed in Table 12. It can be

Table 11:
Marginal Effects

Variable	Marginal Effect
AGE	-0.27
INCOME	-0.02

Table 12:
Expected Value of the Average Number of Hrs. Worked per Week

Category	$E\left[y_{it}^{*} \mid x_{it}\right]$	$E\left[y_{it} \mid x_{it}\right]$
"Average woman"	-8.52	7.25
North	3.42	12.78
Middle	0.08	11.02
High	-1.85	10.09
Secondary	7.83	15.34
University	18.06	22.30
No Children	-11.73	6.10
With Kids	-15.66	4.89

seen from the table that the average number of hours at work decreases if we move from the North of Italy to the South: women living in the center of Italy work on average 1.76 less than the women living in the North (keeping constant the value of the other variables), and for women living in the South the difference is 5.53 hours.

Holding constant the value of the other variable, a higher degree gets an increase in the average number of hours at work. The presence of children gets an increase in the number of hours at work, while when children aged less than six years are present we observe a reduction in the number of hours worked per week.

Using the random effect specification, it is possible to get an estimate of the two variance terms σ_{α}^{2} and σ_{u}^{2}. This is not the case when using a fixed effects specification (only the variance of the random term u can be estimated).

When using I.I.-OLS, the variance of the individual effects and of the random terms are 589.9 and 168.5 respectively. The variability in the sample is mostly due to the variability in the individual effects, in other words to the fact that every woman in the sample has her own unobserved characteristics.

References

Amemiya, T. (1984), Tobit models: A survey, *Journal of Econometrics*, 24, 3–61.

Calzolari, G., F. Di Iorio, and G. Fiorentini (1999), Indirect estimation of just-identified models with control variates, Università di Firenze, *Quaderni del Dipartimento di Statistica* N. 46.

Gallant, R., and G. Tauchen (1996), Which moments to match?, *Econometric Theory*, 12, 657-681.

Goldberger, A.S. (1964), *Econometric theory*, New York: John Wiley & Sons.

Gouriéroux, C., and A. Monfort (1991), Simulation-based inference in models with heterogeneity, *Annales d'Economie et de Statistique*, 20/21, 69-107.

Gouriéroux, C., and A. Monfort (1993), Simulation-based inference: A survey with special reference to panel data models, *Journal of Econometrics*, 59, 5-33.

Gouriéroux, C., A. Monfort, and E. Renault (1993), Indirect inference, *Journal of Applied Econometrics*, 8, S85-S118.

Greene, W.H. (1997), *Econometric analysis*, 3rd ed., London: Prentice-Hall International.

Hajivassiliou, V.A., and D. McFadden (1998), The method of simulated scores for the estimation of LDV models, *Econometrica*, 66, 63-96.

Heckman, J.J. (1981), The incidental parameter problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process, in *Structural analysis of discrete data with econometric applications*, C.F. Manski and D. McFadden (eds.), Cambridge, Mass: MIT Press, 179-195.

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

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

Honoré, B.E. (1993), Orthogonality conditions for Tobit models with fixed effects and lagged dependent variables, *Journal of Econometrics*, 59, 35-61.

Honoré, B.E. and E. Kyriazidou (2000), Estimation of Tobit-type models with individual specific effects, *Econometric Reviews*,19, 341–366.

Inkmann, J. (2000), Misspecified heteroskedasticity in the panel Probit model: A small sample comparison of GMM and SML estimators, *Journal of Econometrics*, 97, 227-260.

Laroque, G., and B. Salanie (1989), Estimation of multi-market fix-price models: An application of pseudo maximum likelihood methods, *Econometrica*, 57, 831–860.

Maddala, G. (1987), Limited dependent variable models using panel data, *Journal of Human Resources*, XXII (3), 307-338.

Mealli, F. and C. Rampichini (1999), Estimating binary multilevel models through indirect inference, *Computational Statistics and Data Analysis*, 29, 313-324.

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

Zammarano, S. (1995), Mercato del lavoro e scelta dello stato occupazionale: Aspetti teorici e valutazioni empiriche, Università di Firenze, Facoltà di Economia, Tesi di Laurea.