Working Paper 97-93
Statistics and Econometrics Series 30
Economics Series 45
December 1997

Departamento de Estadística y Econometría
Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

ESTIMATING BINARY CHOICE MODELS FROM COHORT DATA.

M. Dolores Collado*

Abstract

In this paper we discuss the estimation of binary choice models with individual effects, when the data available are time series of independent cross-sections. We specify a random effects model assuming that the conditional expectation of the individual effects is a linear function of the explanatory variables, and we show how to obtain a consistent estimator of the reduced form parameters. Then, we consider a minimum distance estimator and a within-groups estimator of the structural parameters, and we derive their asymptotic distributions. Finally, we carry out some Monte Carlo simulations to analyze the small sample performance of our estimators.

Keywords:
Time series of cross-sections, cohorts, measurement errors, binary choice models.

*Departamento de Estadística y Econometría, Departamento de Economía, Universidad Carlos III de Madrid. C/ Madrid, 126 28903 Madrid. Spain. Tel: 34-1-624.98.61 Fax: 34-1-624.98.49, e-mail: collad@eco.uc3m.es. I would like to thank Manuel Arellano for his suggestions and comments, I am also grateful to César Alonso for his comments. This research was supported by the Spanish Dirección General de Enseñanza Superior, Project No. PB95-0292.

1 Introduction

In this paper we discuss the estimation of binary choice models with individual effects when the data are time series of independent cross-sections, that is, when we observe independent samples of individuals over time. This kind of models are relevant for empirical applications. For example, when we estimate demand systems using household data, we find that, for some goods, a substantial percentage of households do not buy the good (alcohol and tobacco are clear examples). If this is the case, we will need to consider the household decision of whether to buy the good, and therefore, we will need to estimate a binary choice model. An important issue in this kind of models is that in many cases, the household purchasing decision is influenced by some household characteristics that are unobservable to the econometrician. If these unobservable effects are correlated with the explanatory variables, the model cannot be identified using a single cross-section. Nevertheless, if the unobservable effects are constant over time, the model can be identified using panel data.

The problem that arises at this stage is that for some countries, as for the UK, there are no panel data on household expenditures. However, in most of those countries a large survey on household consumption is carried out with a regular periodicity (in the UK the Family Expenditure Survey provides detailed information on annual expenditures). The difference of this type of data, compared to panel data, is that we observe different individuals for different periods of time. Therefore, the estimation methods for panel data can no longer be used with the individual observations.

Deaton (1985) proposed an alternative approach to estimate linear models of individual behaviour using micro data. If we have time series of independent cross-sections, we can divide the population into groups (cohorts) on the basis of a certain characteristic. This characteristic has to be constant over time for each individual in the population. The variable most widely used to define the cohorts in applied research is the year of birth. The key idea of this approach is that at the population level, the groups contain the same individuals over time, and therefore, the cohort population means have a panel structure. Although we do not observe the cohort population means, we can nevertheless consider the cohort sample means as estimates of the cohort population means, what will provide a panel with measurement errors. The advantage compared to the usual errors in variables problem is that, in this case, we can estimate the variances of the measurement errors using the individual data. For linear models, Deaton (1985) and Collado (1997) showed that the estimated variances can be used to correct the classical estimators for panel data.

The researcher has to decide how to define the cohorts. Notice that the larger the cohorts are, the less important the measurement error problem will be. However, since the cross-section dimension of the data set is finite, if the cohorts contain a large number of individuals, the number of groups will be small, and therefore, the cross-section dimension of the cohort panel will not be very large. In most applied research using cohort data (see Browning, Deaton and Irish (1985), Attanasio and Weber (1993) and Blundell, Browning and Meghir (1994)), the sample is divided in a small number of groups with a large number of observations to avoid the measurement error problem. Verbeek and Nijman (1992) analyze the conditions for this approach to be valid for linear models. The definition of the cohorts is an interesting question for applied research, nevertheless, in this article, we will not discuss the cohort design.

When we estimate linear models with individual effects using panel data, the usual approach is to transform the model into first differences or deviations from time means to eliminate the individual effects (see Hsiao (1986), and Arellano and Bover (1990) among others). Unfortunately, this approach is no longer possible in the case of binary choice models and we need additional assumptions for identification. Chamberlain (1984) proposed to parameterize the conditional expectation of the individual effects as a linear function of the explanatory variables. Then, the latent variable on each period is a function of all leads and lags of the explanatory variables, and the reduced form parameters can be estimated using panel data. Once we have the reduced form estimates, the structural parameters can be estimated by minimum distance. The problem of this approach, when we have time series of independent cross-sections, is that any individual is just observed one period, and therefore, we do not observe any lead or lag of the explanatory variables. In this article, we propose to use the cohort sample means instead of the individual observations as explanatory variables. Consequently, we will have a model where the explanatory variables are correlated with the disturbances. However, under normality, the covariance between the explanatory variables and the disturbances is a known function of the variances of the measurement errors. As mentioned above, the measurement error variances can be estimated using the individual observations. Therefore, we can correct the classical estimators for binary choice models and we will show how to obtain consistent estimates of the reduced form parameters. Using the reduced form estimates, we will derive an optimal minimum distance estimator of the structural parameters.

In the context of panel data, Bover and Arellano (1997) proposed a within-groups estimator of the structural parameters. This estimator is also based on the reduced form estimates, and it is easier to calculate than the minimum distance estimator. In this article, we show that it is also possible to obtain a consistent within-groups estimator using cohort data.

The paper is organized as follows. In section 2, we present a binary choice model with individual effects. We obtain a consistent estimator of the reduced form parameters and we derive its asymptotic distribution. Using this estimator, we can obtain a consistent estimator of the structural parameters by minimum distance. In section 3, we consider a within-groups estimator of the structural parameters, and we calculate its asymptotic distribution. The finite sample performance of the estimators is analyzed in section 4. We carry out some Monte Carlo Simulations and we compare the results for different values of the parameters. Section 5 concludes.

2 A Binary Choice Model with Individual Effects

We consider the following linear model with individual effects

$$y^*_{it} = x'_{it}\beta + \eta_i + v_{it}, \qquad (t = 1, \ldots, T,\ i = 1, \ldots, N) \tag{1}$$

where x_{it} is a $k \times 1$ vector of exogenous variables such that

$$E(v_{it} \mid x_{i1}, \ldots, x_{iT}, \eta_i) = 0$$

η_i is the unobservable individual effect and it is potentially correlated with the explanatory variables. The dependent variable y^*_{it} is not observed. What we observe is the binary variable y_{it} defined by $y_{it} = 1(y^*_{it} > 0)$, where $1(A)$ is the indicator function. Given that y^*_{it} is not observable, we do need additional assumptions on the distribution of η_i to identify β. Chamberlain (1984) proposed to parameterize the conditional expectation of η_i given the exogenous variables as a linear function of the x_{it}'s:

$$\eta_i = x'_{i1}\lambda_1 + \ldots + x'_{iT}\lambda_T + \theta_i \tag{2}$$

$$E(\theta_i \mid x_{i1}, \ldots, x_{iT}) = 0$$

We can then substitute (2) in (1) to get the reduced form model

$$y^*_{it} = x'_{i1}\pi_{t1} + \ldots + x'_{iT}\pi_{tT} + \varepsilon_{it}, \qquad (t = 1, \ldots, T,\ i = 1, \ldots, N) \tag{3}$$

where $\pi_{ts} = \lambda_s$ if $t \neq s$, $\pi_{tt} = \beta + \lambda_t$, and $\varepsilon_{it} = v_{it} + \theta_i$ is the error term, which is uncorrelated with the x_{it}'s. If we observed the same individuals over time (i.e. if we had panel data), the reduced form parameters $(\pi_t,\ t = 1, \ldots, T)$ could be estimated using the classical estimators for binary choice models. Thus, once we had the reduced form estimates, β could be estimated by minimum distance, or alternatively, a within-groups estimator of β could be obtained as proposed by Bover and Arellano (1997).

The purpose of this paper is to obtain a consistent estimator of β when the data available are time series of independent cross-sections. We will begin with the estimation of the reduced form parameters in (3). The problem in this case is that the x_{is}'s $s \neq t$ in (3) are not observed since the individuals are different from period to period. However, as explained above, the population can be divided in groups with fixed membership over time (cohorts) on the basis of a certain characteristic (see Deaton (1985), Collado (1997)). Let g be the random variable determining the cohort membership for each individual (i.e. individual i belongs to cohort c if and only if $g_i \in I_c$). We define the cohort population means as $x^*_{ct} = E(x_{it} \mid g_i \in I_c)$, and we can assume that for any individual in a given cohort c

$$x_{it} = x^*_{ct} + \varsigma_{it}, \qquad \varsigma_{it} \backsim iid(0, \Sigma)$$

Then, we can calculate the cohort sample means as

$$x_{ct} = \frac{1}{nc}\sum_{g_i\ \epsilon I_c} x_{it} = x^*_{ct} + \varsigma_{ct}, \qquad \text{where } \varsigma_{ct} = \frac{1}{nc}\sum_{g_i\ \epsilon I_c} \varsigma_{it} \backsim iid\left(0, \frac{1}{nc}\Sigma\right) \tag{4}$$

where nc is the number of individuals per cohort[1].

We can write (3) in terms of the cohort sample means as

$$y^*_{it} = x'_{c1}\pi_{t1} + \ldots + x'_{cT}\pi_{tT} + \varepsilon^*_{it}, \qquad (t = 1, \ldots, T,\ i = 1, \ldots, N) \tag{5}$$

where $\varepsilon^*_{it} = \varepsilon_{it} + \pi'_t(x_i - x_c) = \varepsilon_{it} + \pi'_t(\varsigma_i - \varsigma_c)$, $\pi_t = (\pi'_{t1}, \ldots, \pi'_{tT})'$, $x_i = (x'_{i1}, \ldots, x'_{iT})'$, $x_c = (x'_{c1}, \ldots, x'_{cT})'$, $\varsigma_i = (\varsigma'_{i1}, \ldots, \varsigma'_{iT})'$, $\varsigma_c = (\varsigma'_{c1}, \ldots, \varsigma'_{cT})'$ [2]. The problem in (5) is that the disturbances are correlated with the x_c's. However, assuming that the joint distribution of the disturbances and the x_c's is normal, the conditional expectation can be estimated using the individual data. We will propose a measurement error corrected probit estimator of the reduced form coefficients in (5).

Under normality, the conditional expectation of the disturbances given x_c is given by

$$E(\varepsilon^*_{it} \mid x_c) = \pi'_t E(\varsigma_i - \varsigma_c \mid x_c) = \pi'_t \Sigma_{12}\Sigma_{22}^{-1} x_c$$

where $\Sigma_{22} = \text{var}(x_c) = \Sigma_x$, and $\Sigma_{12} = \text{cov}(\varsigma_i - \varsigma_c, x_c)$ is a block-diagonal matrix with T blocks $k \times k$.

The s-th block is

$$\Sigma^s_{12} = \text{cov}(\varsigma_{is} - \varsigma_{cs}, x_{cs}) = \text{cov}(\varsigma_{is} - \varsigma_{cs}, x^*_{cs} + \varsigma_{cs}) = -\frac{1}{nc}\Sigma \tag{6}$$

and the t-th block is

$$\Sigma^t_{12} = \text{cov}(\varsigma_{it} - \varsigma_{ct}, x_{ct}) = \text{cov}(\varsigma_{it} - \varsigma_{ct}, x^*_{ct} + \varsigma_{ct}) = \frac{1}{nc}\Sigma - \frac{1}{nc}\Sigma = 0 \tag{7}$$

The reason why the covariances (6) and (7) are different is the following: given that the individuals are different from period to period, the i-th individual in period t is

[1] We are assuming that nc is constant across cohorts and over time to simplify notation, but this assumption can be easily relaxed.

[2] Notice that the individuals are different from period to period, and therefore, in model (5), the i index in period t does not correspond to the same individual than the i index in period s.

not observed in any other period. Therefore, ς_{it} is included in the average ς_{ct} but ς_{is} is not included in ς_{cs}[3].

We can also calculate the conditional variance of the disturbances given x_c

$$\text{var}(\varepsilon^*_{it} \mid x_c) = \sigma^2_\varepsilon + \pi'_t \text{var}(\varsigma_i - \varsigma_c \mid x_c)\pi_t$$

and

$$\text{var}(\varsigma_i - \varsigma_c \mid x_c) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma'_{12} = \Omega$$

where $\Sigma_{11} = \text{var}(\varsigma_i - \varsigma_c)$ is a block-diagonal matrix with T blocks $k \times k$.

The t-th block is

$$\Sigma^t_{11} = \text{var}(\varsigma_{it} - \varsigma_{ct}) = \frac{nc - 1}{nc}\Sigma \tag{8}$$

and the s-th block $s \neq t$ is

$$\Sigma^s_{11} = \text{var}(\varsigma_{is} - \varsigma_{cs}) = \frac{nc + 1}{nc}\Sigma \tag{9}$$

The reason why the variances (8) and (9) are different is the same as explained above for the covariances (6) and (7).

If we define

$$\varepsilon^+_{it} = \frac{\varepsilon^*_{it} - E(\varepsilon^*_{it} \mid x_c)}{\sqrt{\text{var}(\varepsilon^*_{it} \mid x_c)}}$$

then $\varepsilon^+_{it} \mid x_c \backsim iidN(0,1)$. Thus, we have that

$$\text{Prob}(y_{it} = 1 \mid x_c) = \text{Prob}(\varepsilon^*_{it} \geq -\pi'_t x_c \mid x_c) =$$

$$= \text{Prob}\left(\varepsilon^+_{it} \geq \frac{-\pi'_t x_c - \pi'_t \Sigma_{12}\Sigma_{22}^{-1} x_c}{\sqrt{\sigma^2_\varepsilon + \pi'_t \Omega \pi_t}} \middle| x_c\right) = \text{Prob}\left(\varepsilon^+_{it} \leq \frac{\pi'_t (I_{Tk} + \Sigma_{12}\Sigma_{22}^{-1}) x_c}{\sqrt{\sigma^2_\varepsilon + \pi'_t \Omega \pi_t}} \middle| x_c\right)$$

We can estimate π_t (up to scale σ_ε) by pseudo-maximum likelihood relying on estimates of Σ and Σ_x. Once we have estimated π_t, $t = 1, 2, \ldots, T$, we can estimate β and λ by minimum distance. Let $\theta = (\beta', \lambda')'$, $\lambda = (\lambda'_1, \ldots, \lambda'_T)'$, $\Pi = (\pi_1, \ldots, \pi_T)$, $\hat{\Pi} = (\hat{\pi}_1, \ldots, \hat{\pi}_T), \pi = \text{vec}(\Pi)$ and $\hat{\pi} = \text{vec}(\hat{\Pi})$, the minimum distance estimator of θ is given by

$$\hat{\theta}_{MD} = \arg\min_\theta \, (\hat{\pi} - \pi(\theta))'\hat{W}^{-1}(\hat{\pi} - \pi(\theta))$$

[3]Notice that ς_{cs} is the average of the measurement errors for those individuals belonging to the same cohort than i in period s.

where the optimal choice for $\hat{W}$ is any consistent estimator of the asymptotic variance of $\hat{\pi}$. We obtain the asymptotic distribution of $\hat{\pi}$ in the Appendix. The asymptotic distribution of $\hat{\theta}_{MD}$ is given by

$$\sqrt{C}(\hat{\theta}_{MD} - \theta) \rightarrow_d N(0, (DW^{-1}D')^{-1})$$

where $D = \partial\pi(\theta)/\partial\theta$ and W is the asymptotic variance of $\hat{\pi}$. The asymptotic variance of $\hat{\theta}_{MD}$ can be estimated as

$$\widehat{\text{avar}}(\hat{\theta}_{MD}) = (\hat{D}\hat{W}^{-1}\hat{D}')^{-1}$$

where $\hat{D} = \partial\pi(\hat{\theta}_{MD})/\partial\theta$.

3 A Within-Groups Estimator for the Binary Choice Model

We are going to consider an alternative estimator of β for the binary choice model. In the context of panel data, Bover and Arellano (1997) proposed a within-groups estimator of β based on a consistent estimator of π. The advantage of this estimator, compared to the minimum distance estimator, is that it is very easy to calculate and it does not involve the estimation of the nuisance parameters λ. The disadvantage is that it is less efficient than the optimal minimum distance estimator. Nevertheless, Bover and Arellano (1997) show that it is possible to obtain a linear GMM estimator which is asymptotically equivalent to the optimal minimum distance. Following this approach, we are going to consider a within-groups estimator of β when the data available are time series of independent cross-sections.

Let us come back to the model for the latent variable (1). We can consider the random variable g defined above, which determines the cohort membership for each individual. Taking expectations conditional on g_i in model (1) we get

$$y^*_{ct} = x^{*\prime}_{ct}\beta + \eta^*_c + v^*_{ct}, \qquad (t = 1, \ldots, T, \ c = 1, \ldots, C)$$

where $x^*_{ct} = E(x_{it} \mid g_i \in I_c)$, $y^*_{ct} = E(y^*_{it} \mid g_i \in I_c)$, $\eta^*_c = E(\eta_i \mid g_i \in I_c)$ and $v^*_{ct} = E(v_{it} \mid g_i \in I_c)$. We can write the model for the cohort population means as a system of equations

$$y_c^* = X_c^* \beta + \eta_c^* e + v_c^* \tag{10}$$

where $y_c^* = (y_{c1}^*, \ldots, y_{cT}^*)'$, $X_c^* = (x_{c1}^*, \ldots, x_{cT}^*)'$, $v_c^* = (v_{c1}^*, \ldots, v_{cT}^*)'$, and e is a $T \times 1$ vector of ones. We can transform the variables in (10) into deviations from time means to eliminate the cohort effects. Let $\tilde{y}_c^* = Qy_c^*$, $\tilde{X}_c^* = QX_c^*$, and $\tilde{v}_c^* = Qv_c^*$, where $Q = I_T - ee'/T$

$$\tilde{y}_c^* = \tilde{X}_c^* \beta + \tilde{v}_c^* \tag{11}$$

Let us now consider model (3). Taking expectations conditional on g_i, we can write the system of T equations as

$$y_c^* = \Pi' x_c^* + \varepsilon_c^*$$

We can now apply the deviations from time means operator Q to obtain

$$\tilde{y}_c^* = Q\Pi' x_c^* + \tilde{v}_c^* \tag{12}$$

From (11) and (12) we get the following set of restrictions

$$\tilde{X}_c^* \beta = Q\Pi' x_c^* \tag{13}$$

Notice that, contrary to what happens in the true panel case, the cohort population means in (13) are not observed. However, we can write the set of restrictions in terms of the cohort sample means. We can define $\Psi_c = (\varsigma_{c1}, \ldots, \varsigma_{cT})'$ and $\varsigma_c = \text{vec}(\Psi_c')$. The Π matrix can be written as $\Pi' = (I_T \otimes \beta' + i \otimes \lambda')$ and therefore

$$\Pi' \varsigma_c = \Psi_c \beta + i \otimes \lambda' \varsigma_c$$

Multiplying by Q we get

$$Q\Pi' \varsigma_c = Q\Psi_c \beta = \tilde{\Psi}_c \beta \tag{14}$$

Adding (13) and (14) we get the following set of restrictions between Π and β:

$$\tilde{X}_c \beta = Q\Pi' x_c \tag{15}$$

and multiplying by $\tilde{X}_c'$ we obtain

$$\tilde{X}_c' \tilde{X}_c \beta = \tilde{X}_c' \Pi' x_c \tag{16}$$

Therefore, using a consistent estimator of Π, we can obtain a within-groups estimator of β which is consistent. This estimator is given by

$$\hat{\beta}_{WG} = \left(\sum_{c=1}^{C} \tilde{X}_c'\tilde{X}_c\right)^{-1} \left(\sum_{c=1}^{C} \tilde{X}_c'\hat{\Pi}'\ x_c\right) \tag{17}$$

Substracting β from (17) we get

$$\sum_{c=1}^{C} \tilde{X}_c'\tilde{X}_c(\hat{\beta}_{WG} - \beta) = \sum_{c=1}^{C} \tilde{X}_c'(\hat{\Pi}'x_c - \tilde{X}_c\beta)$$

and using (15) we obtain

$$\sum_{c=1}^{C} \tilde{X}_c'\tilde{X}_c(\hat{\beta}_{WG} - \beta) = \sum_{c=1}^{C} \tilde{X}_c'(\hat{\Pi} - \Pi)'x_c$$

We can rewrite the expression above as

$$\hat{\beta}_{WG} - \beta = \left(\sum_{c=1}^{C} \tilde{X}_c'\tilde{X}_c\right)^{-1} \sum_{c=1}^{C} (\tilde{X}_c \otimes x_c)'\mathrm{vec}(\hat{\Pi} - \Pi)$$

and the asymptotic distribution of $\hat{\beta}_{WG}$ is given by

$$\sqrt{C}(\hat{\beta}_{WG} - \beta) \rightarrow_d N(0, E(\tilde{X}_c'\tilde{X}_c)^{-1}E(\tilde{X}_c \otimes x_c)'WE(\tilde{X}_c \otimes x_c)E(\tilde{X}_c'\tilde{X}_c)^{-1})$$

where W is the asymptotic variance of $\hat{\pi}$. The asymptotic variance of $\hat{\beta}_{WG}$ can be estimated by

$$\widehat{\mathrm{avar}}(\hat{\beta}_{WG}) = \left(\sum_{c=1}^{C} \tilde{X}_c'\tilde{X}_c\right)^{-1} \sum_{c=1}^{C}(\tilde{X}_c \otimes x_c)'\hat{W}\sum_{c=1}^{C}(\tilde{X}_c \otimes x_c) \left(\sum_{c=1}^{C} \tilde{X}_c'\tilde{X}_c\right)^{-1}$$

where $\hat{W}$ is a consistent estimator of W.

4 Monte Carlo Simulations

In the previous sections we have derived the asymptotic distribution of the minimum distance estimator and the within-groups estimator for the binary choice model, using time series of cross-section data. However, it also interesting to analyze the small sample performance of the estimators that we have proposed. For this purpose, we have carried out some Monte Carlo simulations for different values of the parameters of the model.

We consider a binary choice model with just one explanatory variable, where the model for the latent variable is

$$y_{it}^{*} = \beta x_{it} + \eta_i + v_{it} \tag{18}$$

The data have been generated as follows. First, we have generated the cohort population means for the explanatory variable using an AR(1) model.

$$x_{ct}^{*} = \rho x_{ct-1}^{*} + w_{it}$$

The initial values for the x_{c0}^{*}'s have been generated as $iidN(0,\sigma_{x0}^2)$, the w_{it} are also $iidN(0,\sigma_w^2)$ and the first ten cross-sections were discarded. Notice that the variance of the cohort population means is given by

$$\sigma_x^2 = \frac{\sigma_w^2}{1-\rho^2}$$

Then, for different values of ρ, we will change σ_w^2 to keep constant the variance of the cohort population means. For each cohort, we generate the individual observations for the explanatory variable as

$$x_{it} = x_{ct}^{*} + \varsigma_{it}, \qquad \varsigma_{it} \backsim iidN(0,\sigma_\varsigma^2)$$

We generate $T \times nc$ observations per cohort, and for each period we keep in the sample nc observations corresponding to different individuals. The individual effects are generated as

$$\eta_i = \lambda_1 x_{i1} + \ldots + \lambda_T x_{iT} + \theta_i, \qquad \theta_i \backsim iidN(0,\sigma_\theta^2)$$

and we generate the latent variable y_{it}^{*} using (18). The binary variable is $y_{it} = 1(y_{it}^{*} > 0)$.

We carried out experiments for different values of the variances and different values of the correlation parameter ρ. The results for a thousand replications are summarized in table 1. The time dimension is $T = 5$, the number of cohorts $C = 100$ and the number of individuals per cohort is $nc = 25$. The structural parameter $\beta = 1$ and the nuisance parameters (the λ's) are all equal to one.

We considered the optimal minimum distance (MD) estimator and the within-groups (WG) estimator and we have calculated the standard errors based on the asymptotic distribution of the estimators. As we can see in Table 1, the finite sample bias of both MD and WG estimators is quite small for any value of the parameters, and it is even smaller for the WG. The standard deviation of the MD estimator is smaller than the standard deviation of WG. This reflects the fact that the WG estimator is less efficient than the optimal MD. Another important issue is that the mean of the asymptotic standard errors is very similar to the standard deviation for all the values of the parameters that we have considered. If we now compare the performance of the estimators for the different values of ρ, we can see that the bias is quite similar when the autocorrelation of the explanatory variables is not very large. However, when ρ is large ($\rho = 0.8$) the bias is larger. The standard deviation of the estimators increases with ρ. Finally, we are going to analyze the behaviour of the estimators for different values of the variance of the measurement errors (σ^2_ς) and different values of the variance of the cohort population means (σ^2_x). As expected, both the bias and the standard deviation of the estimators are smaller, the smaller the variance of the measurement errors, while the performance of the estimators is better the larger the variance of the cohort population means.

Table 1
Means, standard deviations and standard errors of the estimators
C = 100, nc=25, T=5

			MD	WG	MD	WG
			$\sigma_x^2 = 1$		$\sigma_x^2 = 0.5$	
$\sigma_\varsigma^2 = 1$	$\rho = 0$	Mean	0.9691	1.0236	0.9557	1.0338
		St. Dev.	0.0624	0.0704	0.0785	0.0924
		Mean SE	0.0613	0.0678	0.0794	0.0893
	$\rho = 0.5$	Mean	0.9641	1.0267	0.9501	1.0366
		st. Dev.	0.0764	0.0854	0.0955	0.1104
		Mean SE	0.0739	0.0836	0.0940	0.1060
	$\rho = 0.8$	Mean	0.9328	1.0494	0.9095	1.1188
		St. Dev.	0.1068	0.1360	0.1270	0.3333
		Mean SE	0.1065	0.1336	0.1375	0.4181
$\sigma_\varsigma^2 = 0.5$	$\rho = 0$	Mean	0.9880	1.0089	0.9867	1.0138
		St. Dev	0.0497	0.0511	0.0644	0.0646
		Mean SE	0.0478	0.0508	0.0617	0.0655
	$\rho = 0.5$	Mean	0.9873	1.0137	0.9794	1.0126
		St. Dev	0.0645	0.0664	0.0810	0.0830
		Mean SE	0.0602	0.0656	0.0759	0.0816
	$\rho = 0.8$	Mean	0.9710	1.0189	0.9640	1.0310
		St. Dev.	0.0975	0.1037	0.1136	0.1252
		Mean SE	0.0882	0.0999	0.1112	0.1237

5 Conclusions

We analyzed in this paper the estimation of binary choice models with individual effects, when the data are time series of independent cross-sections. We first obtained a measurement error corrected estimator of the reduced form parameters using the cohort sample means and we derived its asymptotic distribution. Based on the reduced form estimates, we proposed a minimum distance estimator and a within-groups estimator of the structural parameters. We also obtained the asymptotic distribution of those estimators.

Finally, we carried out some Monte Carlo simulations to study the small sample behaviour of our estimators. The main conclusions are that both estimators perform quite well in relatively small samples, but while the minimum distance has a smaller standard deviation, the within-groups is less biased in small samples. Another important result is that the asymptotic standard errors are very similar to the standard deviations.

Appendix. Asymptotic Distribution of the Probit Estimator

The maximum likelihood estimator of π_t, $t = 1, \ldots, T$, is asymptotically equivalent to a Generalized Methods of Moment estimator (GMM) that uses the optimal set of instruments (see Chamberlain (1987)). Thus, we will derive the asymptotic distribution using the GMM estimator. The conditional expectation of y_{it} given x_c is given by

$$E(y_{it} \mid x_c) = \text{Prob}\left(\varepsilon_{it}^{+} \leq \frac{\pi_t'(I_{TK} + \Sigma_{12}\Sigma_{22}^{-1})x_c}{\sqrt{\sigma_\varepsilon^2 + \pi_t'\Omega\pi_t}} \;\middle|\; x_c\right) \tag{19}$$

The expression above can be written as

$$E(y_{it} \mid x_c) = F(x_c, \pi_t, \Sigma, \Sigma_x)$$

and therefore

$$E(y_{it} - F(x_c, \pi_t, \Sigma, \Sigma_x) \mid x_c) = 0$$

Let $u_{it} = y_{it} - F(x_c, \pi_t, \Sigma, \Sigma_x)$, the conditional variance of u_{it} is given by

$$E(u_{it}^2 \mid x_c) = F(x_c, \pi_t, \Sigma, \Sigma_x) \times (1 - F(x_c, \pi_t, \Sigma, \Sigma_x)) = \sigma^2(x_c)$$

The optimal set of instruments is therefore

$$\frac{1}{\sigma^2(x_c)} \frac{\partial F(x_c, \pi_t, \Sigma, \Sigma_x)}{\partial \pi_t}$$

and the moment conditions are given by

$$E\left[(y_{it} - F(x_c, \pi_t, \Sigma, \Sigma_x))\frac{1}{\sigma^2(x_c)} \frac{\partial F(x_c, \pi_t, \Sigma, \Sigma_x)}{\partial \pi_t}\right] = 0$$

The model is just identified, and the method of moments (MM) estimator $\tilde{\pi}_t^*$ solves the following system of equations

$$\frac{1}{N}\sum_{i=1}^{N}(y_{it} - F(x_c, \tilde{\pi}_t^*, \Sigma, \Sigma_x))\frac{1}{\sigma^2(x_c)} \frac{\partial F(x_c, \pi_t, \Sigma, \Sigma_x)}{\partial \pi_t} = 0$$

The optimal instruments are not observable since they depend on π_t, and therefore, the MM estimator $\tilde{\pi}_t^*$ is not feasible. However, we can consider π_t in the instruments as an argument to be estimated and then the MM estimator is given by

$$\frac{1}{N}\sum_{i=1}^{N}(y_{it}-F(x_c,\tilde{\pi}_t,\Sigma,\Sigma_x))\frac{1}{\tilde{\sigma}^2(x_c)}\frac{\partial F(x_c,\tilde{\pi}_t,\Sigma,\Sigma_x)}{\partial \pi_t}=0 \tag{20}$$

where $\tilde{\sigma}^2(x_c) = F(x_c,\tilde{\pi}_t,\Sigma,\Sigma_x)\times(1-F(x_c,\tilde{\pi}_t,\Sigma,\Sigma_x))$. The MM estimator $\tilde{\pi}_t$ coincides with the maximum likelihood estimator. The reason is that (20) are the first order conditions for the maximization of the log-likelihood function.

Expression (20) can also be written as

$$\frac{1}{C}\sum_{c=1}^{C}(y_{ct}-F(x_c,\tilde{\pi}_t,\Sigma,\Sigma_x))\frac{1}{\tilde{\sigma}^2(x_c)}\frac{\partial F(x_c,\tilde{\pi}_t,\Sigma,\Sigma_x)}{\partial \pi_t}=0 \tag{21}$$

where

$$y_{ct}=\frac{1}{nc}\sum_{g_i\ \epsilon I_c} y_{it}$$

Unfortunately, Σ and Σ_x are not observed but they can be estimated using the moment conditions:

$$E\left[(x_{it}-x_{ct})(x_{it}-x_{ct})'\right]=\frac{nc-1}{nc}\Sigma$$

$$E\left[(x_c-E(x_c))(x_c-E(x_c))'\right]=\Sigma_x$$

The expectations can be replaced by their sample counterparts

$$\frac{1}{C}\sum_{c=1}^{C}\hat{\Sigma}_c=\hat{\Sigma}$$

$$\frac{1}{C}\sum_{c=1}^{C}(x_c-\bar{x})(x_c-\bar{x})'=\hat{\Sigma}_x$$

where

$$\hat{\Sigma}_c=\frac{1}{T}\sum_{t=1}^{T}\frac{1}{nc-1}\sum_{g_i\ \epsilon I_c}(x_{it}-x_{ct})(x_{it}-x_{ct})'$$

and

$$\bar{x}=\frac{1}{C}\sum_{c=1}^{C}x_c$$

Substituting Σ and Σ_x by $\hat{\Sigma}$ and $\hat{\Sigma}_x$ in (21), our MM estimator solves the system

$$\frac{1}{C}\sum_{c=1}^{C}(y_{ct} - F(x_c, \hat{\pi}_t, \hat{\Sigma}, \hat{\Sigma}_x))\frac{1}{\hat{\sigma}^2(x_c)}\frac{\partial F(x_c, \hat{\pi}_t, \hat{\Sigma}, \hat{\Sigma}_x)}{\partial \pi_t} = 0$$

where $\hat{\sigma}^2(x_c) = F(x_c, \hat{\pi}_t, \hat{\Sigma}, \hat{\Sigma}_x) \times (1 - F(x_c, \hat{\pi}_t, \hat{\Sigma}, \hat{\Sigma}_x))$.

The expression above can be written as

$$\frac{1}{C}\sum_{c=1}^{C}\psi_t(w_{ct}, \hat{\pi}_t, \hat{\gamma}) = 0 \qquad t = 1, \ldots, T \tag{22}$$

where $w_{ct} = (y_{ct}, x_c')'$ and $\hat{\gamma} = (\text{vech}(\hat{\Sigma})', \text{vech}(\hat{\Sigma}_x)')'$[4]. We can write the systems of T equations in (22) as

$$\frac{1}{C}\sum_{c=1}^{C}\psi(w_c, \hat{\pi}, \hat{\gamma}) = 0 \tag{23}$$

where $w_c = (w_{c1}', \ldots, w_{cT}')'$, $\hat{\Pi} = (\hat{\pi}_1, \ldots, \hat{\pi}_T)$ and $\hat{\pi} = \text{vec}(\hat{\Pi})$. Using a Taylor expansion we can write

$$\frac{1}{\sqrt{C}}\sum_{c=1}^{C}\psi(w_c, \hat{\pi}, \hat{\gamma}) = \frac{1}{\sqrt{C}}\sum_{c=1}^{C}\psi(w_c, \hat{\pi}, \gamma) + \frac{1}{\sqrt{C}}\sum_{c=1}^{C}\frac{\partial\psi(w_c, \hat{\pi}, \gamma)}{\partial\gamma'}(\hat{\gamma} - \gamma) + o_p(1)$$

$$= \frac{1}{\sqrt{C}}\sum_{c=1}^{C}\psi(w_c, \pi, \gamma) + \frac{1}{\sqrt{C}}\sum_{c=1}^{C}\frac{\partial\psi(w_c, \pi, \gamma)}{\partial\pi'}(\hat{\pi} - \pi) + \frac{1}{\sqrt{C}}\sum_{c=1}^{C}\frac{\partial\psi(w_{ct}, \hat{\pi}, \gamma)}{\partial\gamma'}(\hat{\gamma} - \gamma) + o_p(1)$$

Let

$$D_\pi = E\left(\frac{\partial\psi(w_c, \pi, \gamma)}{\partial\pi}\right)$$

$$D_\gamma = E\left(\frac{\partial\psi(w_c, \pi, \gamma)}{\partial\gamma}\right)$$

The D_π matrix is block-diagonal with T blocks. The t-th block is

$$D_{\pi_t} = E\left(\frac{\partial\psi_t(w_{ct}, \pi_t, \gamma)}{\partial\pi_t}\right)$$

Then, we have

$$\frac{1}{\sqrt{C}}\sum_{c=1}^{C}\psi(w_c, \pi, \gamma) + D_\pi'\sqrt{C}(\hat{\pi} - \pi) + D_\gamma'\frac{1}{\sqrt{C}}\sum_{c=1}^{C}(\hat{\gamma}_c - \gamma) + o_p(1) = 0 \tag{24}$$

[4]The vech operator vectorizes a $p \times p$ symmetric matrix by selecting the $p \times (p+1)/2$ elements in the lower triangular part of the matrix.

where $\hat{\gamma}_c = (\text{vech}(\hat{\Sigma}_c)', \text{vech}((x_c - \bar{x})(x_c - \bar{x})')')'$. We can rewrite (24) as

$$\sqrt{C}(\hat{\pi} - \pi) = -(D_\pi')^{-1}[I_{T^2K} : D_\gamma'] \frac{1}{\sqrt{C}} \sum_{c=1}^{C} \varphi(w_c, \pi, \gamma) + o_p(1)$$

where

$$\varphi(w_c, \pi, \gamma) = \begin{pmatrix} \psi(w_c, \pi, \gamma) \\ \hat{\gamma}_c - \gamma \end{pmatrix}$$

and the asymptotic distribution of $\hat{\pi}$ is

$$\sqrt{C}(\hat{\pi} - \pi) \rightarrow_d N(0, (D_\pi')^{-1}(I_{T^2K} : D_\gamma')V_0 \begin{pmatrix} I_{T^2K} \\ D_\gamma \end{pmatrix} D_\pi^{-1})$$

where $V_0 = E(\varphi(w_c, \pi, \gamma)\varphi(w_c, \pi, \gamma)')$. The asymptotic variance of $\hat{\pi}$ can be estimated by

$$\hat{W} = \widehat{\text{avar}}(\hat{\pi}) = (\hat{D}_\pi')^{-1}(I_{T^2K} : \hat{D}_\gamma')\hat{V} \begin{pmatrix} I_{T^2K} \\ \hat{D}_\gamma \end{pmatrix} \hat{D}_\pi^{-1}$$

where

$$\hat{D}_\pi = \frac{1}{C} \sum_{c=1}^{C} \frac{\partial \psi(w_c, \hat{\pi}, \hat{\gamma})}{\partial \pi}$$

$$\hat{D}_\gamma = \frac{1}{C} \sum_{c=1}^{C} \frac{\partial \psi(w_c, \hat{\pi}, \hat{\gamma})}{\partial \gamma}$$

and

$$\hat{V} = \frac{1}{C} \sum_{c=1}^{C} \varphi(w_c, \hat{\pi}, \hat{\gamma})\varphi(w_c, \hat{\pi}, \hat{\gamma})'$$

References

[1] Arellano, M. and O. Bover (1990), "La Econometría de Datos de Panel", *Investigaciones Económicas*, Vol 14, pp 3-45.

[2] Attanasio, O.P. and G. Weber (1993), "Consumption Growth, the Interest Rate and Aggregation", *Review of Economic Studies*, Vol. 60, pp. 631-649.

[3] Blundell, R., M. Browning, and C. Meghir, "Consumer Demand and the Life-Cycle Allocation of Household Expenditures", *Review of Economic Studies*, Vol 61, pp. 57-80.

[4] Bover, O. and M. Arellano (1997), "Estimating Dynamic Limited Dependent Variable Models from Panel Data", *Investigaciones Económicas*, Vol. 21.

[5] Browning, M., A. Deaton and M. Irish (1985), "A Profitable Approach to Labour Supply and Commodity Demand over the Life-Cycle", *Econometrica*, Vol. 53, pp.503-543.

[6] Chamberlain, G. (1984), "Panel Data", in Z. Griliches and M.D. Intrilligator (eds.), *Handbook of Econometrics*, vol II, Elsevier Science.

[7] Chamberlain, G. (1987), "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions", *Journal of Econometrics*, Vol. 34, pp. 305-334.

[8] Collado, M.D. (1997), "Estimating Dynamic Models from Time Series of Independent Cross-Sections", *Journal of Econometrics*, Vol. 82, pp. 37-62.

[9] Deaton, A. "Panel Data from Time Series of Cross-Sections", *Journal of Econometrics*, Vol. 30, pp.109-126.

[10] Hsiao, C. (1986), *Analysis of Panel Data,* Cambridge University Press.

[11] Verbeek, M.and T. Nijman (1992), " Can Cohort Data be Treated as Genuine Panel Data?, *Empirical Economics*, Vol. 17, pp. 9-23.